UNIVERSAL
LIBRARY
OU_152209
UNIVERSAL
LIBRARY

OUP—1695—8-11-77—5,000. FOR STUDENTS ONLY

OSMANIA UNIVERSITY LIBRARY

Call No. 150.13
B65I

Accession No. 63563

Author Blum, M.L. & Naylor, J.C

Title Industrial Psychology

This book should be returned on or before the date last marked below

INDUSTRIAL PSYCHOLOGY

In the series under the editorship of
WAYNE HOLTZMAN *and* GARDNER MURPHY

Harper International Edition

HARPER & ROW, PUBLISHERS

New York, Evanston, and London

INDUSTRIAL PSYCHOLOGY

Its Theoretical and Social Foundations

A REVISION OF

INDUSTRIAL PSYCHOLOGY AND ITS SOCIAL FOUNDATIONS

REVISED EDITION

Milton L. Blum & James C. Naylor

This book is a revision of
Industrial Psychology and Its Social Foundations, revised edition
Copyright 1949, 1956, by Harper & Row, Publishers Incorporated

INDUSTRIAL PSYCHOLOGY: ITS THEORETICAL AND SOCIAL FOUNDATIONS
Copyright 1968 by Milton L. Blum and James C. Naylor

All rights reserved No part of this book may be used or reproduced in any manner whatsoever without written permission except in the case of brief quotations embodied in critical articles and reviews. For information address Harper & Row, Publishers. Incorporated, 49 East 33rd Street, New York, N.Y. 10016., U.S.A.

Library of Congress Catalog Card Number 68 16210

HARPER INTERNATIONAL EDITION

INT 35-02069 AINT 35-30789 EINT 35-60786

3rd Printing

Printed in Singapore by Times Printers, Sdn. Bhd.

To Naomi and Georgia

CONTENTS

Preface xi

1. *Introduction* 1

THE PSYCHOLOGIST. INDUSTRIAL PSYCHOLOGY. THE INDUSTRIAL PSYCHOLOGIST. SCOPE OF THE INDUSTRIAL PSYCHOLOGIST'S WORK. CONSULTING ORGANIZATIONS. RESEARCH CENTERS AT UNIVERSITIES. OTHER RESEARCH ORGANIZATIONS. HISTORY AND DEVELOPMENT OF THE FIELD. MAJOR PROBLEMS OF INDUSTRIAL PSYCHOLOGY. CHAPTERS TO FOLLOW.

2. *A Statistical Rationale for Selection and Placement* 22

MANPOWER PLANNING. BASIC SELECTION MODEL. CORRELATION. REGRESSION. STATISTICAL SIGNIFICANCE. CHARACTERISTICS OF PREDICTORS. DETERMINING THE UTILITY OF A SELECTION INSTRUMENT. A DECISION THEORY APPROACH. SOME CONCLUSIONS.

3. *Problems and Procedures in Selection and Classification* 59

MULTIPLE PREDICTION. PREDICTION SYSTEMS. MODERATORS. DUNNETTE'S SELECTION MODEL. SUPPRESSOR VARIABLES. CROSS-VALIDATION.

4. *Testing in Industry* 88

CHARACTERISTICS OF PSYCHOLOGICAL TESTS. TEST NORMS. TYPES OF TESTS. TESTS VERSUS OTHER SELECTION DEVICES. PROS AND CONS OF TESTING. CURRENT TESTING PRACTICES. SPECIFIC EXAMPLES OF TESTS. A STUDY IN CONTROVERSY. OVERVIEW OF PERSONALITY TESTING IN INDUSTRY. EXAMPLES OF INDUSTRIAL TESTING. UNITED STATES EMPLOYMENT SERVICE TESTING. TESTING OF EXECUTIVES. TESTING OF SCIENTISTS. CREATIVITY. A POINT OF VIEW. RESPONSE SETS OR FAKING. TESTING AND CIVIL RIGHTS.

5. *The Interview and Other Selection Devices* 143

INTERVIEWS. THE SELECTION INTERVIEW. OTHER INTERVIEW CONSIDERATIONS. APPLICATION BLANKS. REFERENCES AND RECOMMENDATIONS. OTHER MISCELLANEOUS PREDICTORS.

6. *The Criterion* 174

DEFINITION OF THE CRITERION. CLASSICAL VIEW OF THE CRITERION. REQUIREMENTS OF CRITERIA. CRITERION DEVELOPMENT. COMPOSITE VERSUS MULTIPLE CRITERIA. THEORY AND CRITERIA IN INDUSTRIAL PSYCHOLOGY.

7. *Performance Appraisal* 195

TYPICAL INDUSTRIAL CRITERIA. SOME EXAMPLES. JUDGMENTAL CRITERIA. ERRORS AND BIAS IN PERFORMANCE APPRAISAL. JUDGMENTAL METHODS OF APPRAISING PERFORMANCE. RESEARCH ON FORCED-CHOICE METHOD. RATER CHARACTERISTICS. DIMENSIONS OF CRITERIA. EXECUTIVE PERFORMANCE APPRAISAL. PERFORMANCE APPRAISAL INTERVIEW.

8. *Training and Learning* 237

THE LEARNING PROCESS. PRINCIPLES OF LEARNING. SOME TRAINING GUIDELINES. EXPERIENCE VERSUS TRAINING. KINDS OF TRAINING. SYSTEMS APPROACH TO TRAINING. SPECIAL TRAINING AREAS. PROGRAMED INSTRUCTION. EVALUATION OF TRAINING EFFECTIVENESS. SPECIALIZED TRAINING. TYPICAL TRAINING DEPARTMENTS IN ACTION. THE TRAINING STAFF.

9. *Attitude Measurement* 274

ATTITUDES AND SOCIAL PHENOMENA. METHODS OF MEASURING ATTITUDES. ATTITUDE SURVEYS AND THEIR APPLICATION. THE ATTITUDES OF EMPLOYEES VERSUS THE ATTITUDES OF EMPLOYERS.

10. *The Hawthorne Studies* 306

STUDY 1: EXPERIMENTS ON ILLUMINATION. STUDY 2: RELAY ASSEMBLY TEST ROOM. STUDY 3: MASS INTERVIEWING PROGRAM. STUDY 4: BANK WIRING OBSERVATION ROOM. STUDY 5: PERSONNEL COUNSELING. IMPLICATIONS OF THE HAWTHORNE STUDIES.

11. *Motivation and Work* 328

THE COMPLEXITY OF MOTIVATION. MASLOW'S THEORY OF HUMAN MOTIVATION. VROOM'S THEORY OF HUMAN MOTIVATION. MOTIVATION AND FRUSTRATION. FINANCIAL INCENTIVES AS A MOTIVATING FORCE. KINDS OF INCENTIVES. RESEARCH ON INCENTIVES. MASLOW'S THEORY AND FINANCIAL INCENTIVES. WHY PEOPLE WORK. AN OVERALL VIEW OF MOTIVATION AND WORK.

12. *Job Satisfaction* 364

CONFUSION OF TERMINOLOGY. REASONS FOR JOB SATISFACTION RESEARCH. FACTORS MEASURED IN JOB SATISFACTION. RELATIVE IMPORTANCE OF DIFFERENT ASPECTS OF JOB SATISFACTION. JOB SATISFACTION AND JOB BEHAVIOR. THEORIES OF JOB SATISFACTION. SOME GENERAL COMMENTS ON JOB SATISFACTION. PREDICTING JOB SATISFACTION.

13. *Industrial Morale* 391

DEFINING MORALE. DETERMINERS OF MORALE. MEASUREMENT OF MORALE. METHODS OF INCREASING INDUSTRIAL MORALE. GROUPS AND GROUP DYNAMICS. A COMPARISON OF METHODS FOR INCREASING MORALE.

14. *Leadership and Supervision* 414

LEADERSHIP AND POLITICAL ATMOSPHERE. LEADERSHIP CRITERIA. TWO APPROACHES TO THE STUDY OF LEADERSHIP. THEORIES OF LEADERSHIP. COMMENTS ON BLAKE'S MANAGERIAL GRID. COMMUNICATION. DO'S AND DON'TS. SELECTION AND TRAINING OF LEADERS. ROLE OF EXECUTIVE WIVES.

15. *Decision Making* 450

THE DECISION PROCESS. NORMATIVE MODELS OF INDIVIDUAL DECISION MAKING. SOME BASIC DECISION VARIABLES. A COGNITIVE MODEL. ORGANIZATIONAL DECISION MAKING.

16. *Organizations* 470

WHAT IS AN ORGANIZATION? ORGANIZATIONAL STRUCTURE. LINKAGE SYSTEMS IN ORGANIZATIONS. THEORIES OF ORGANIZATIONS. RECENT RESEARCH ON ORGANIZATIONAL BEHAVIOR. COMMUNICATION IN ORGANIZATIONS. DECENTRALIZATION OF ORGANIZATIONS.

17. *Job Analysis and Evaluation* 491

USES AND VALUE OF JOB ANALYSIS. JOB ANALYSIS METHODS. SOME EXAMPLES. SOME RESEARCH ON JOB ANALYSIS. JOB EVALUATION. THE EVALUATION OF JOB EVALUATION. PSYCHOLOGICAL CONTRIBUTIONS TO JOB ANALYSIS AND EVALUATION. OCCUPATIONAL INFORMATION. CLUSTERING OR GROUPING OCCUPATIONS.

18. *Accidents, Safety, and Fatigue* 516

ACCIDENT PRONENESS PRINCIPLE. ACCIDENT REDUCTION. PREDICTING ACCIDENTS. FATIGUE. HOURS OF WORK AND PRODUCTION. THE ULTIMATE WORKWEEK. UNPRODUCTIVE WORKING TIME. REST PAUSES. ABSENTEEISM.

19. *Work Environment* 553

MUSIC IN INDUSTRY. THE "AROUSAL" HYPOTHESIS. NOISE. ILLUMINATION. COLOR. VIBRATION. MISCELLANEOUS FACTORS.

20. *Human Performance* 571

TIME AND MOTION STUDY METHODOLOGY. POTENTIAL FAILURE OF TIME AND MOTION STUDIES. RESISTANCE TO TIME AND MOTION STUDIES. MEN IN MOTION. SOME SIMPLE PRINCIPLES. NEED FOR ALLOWANCES IN TIME AND MOTION STUDY. TIME AND MOTION STUDY IN THE HOME. EXAMPLES OF CONTINUED INEFFICIENCIES. BREAKING RESISTANCE. HUMAN ENGINEERING. CODING DISPLAYS AND CONTROLS. PRINCIPLES OF HUMAN PERFORMANCE. PERCEPTUAL-MOTOR SKILLS. INFORMATION PROCESSING BEHAVIOR. MONITORING BEHAVIOR.

Appendix: NAYLOR-SHINE TABLE FOR DETERMINING THE INCREASE IN MEAN CRITERION SCORE OBTAINED BY USING A SELECTION DEVICE. 609

Indices 619

INDEX OF NAMES 621

INDEX OF SUBJECTS 629

PREFACE

Industrial psychology continues to be growing and in a state of transition, and this revision is an expression of the kinds of change that are taking place. The motives for this third edition are similar in many ways to those of the earlier books. They represent the authors' views of industrial psychology at present and their thoughts about where it is going in the future. In 1956, when this book was previously revised, industrial psychology was in the process of truly entering what might be called the "Human Relations" phase of studying man and his place in the world of work. Although the foundation for this movement had occurred years earlier in the classical Western Electric research, it was not until the 1950s that industrial psychology became bold enough to join hands with social psychology and to begin a systematic look at the many interpersonal aspects of human interrelationships existing in the work situation. Indeed, some industrial psychologists did not welcome this move. For years, industrial psychology had remained a secure, scientifically respectable discipline under the guise of a strong empirical emphasis best represented by the traditional areas of employee selection and placement, performance appraisal, and job evaluation. The social aspects of the working man were generally avoided, as they did not easily lend themselves to empirical investigation and so did not fit into the scheme of things.

However, the change in emphasis did occur, and we like to think that the earlier editions of this book contributed to this change by stressing the social foundations of industrial psychology and their growing importance.

Industrial psychology in the mid 1960s seems to be experiencing a transition of emphasis just as substantial as that which occurred in the early 1950s. However, its nature is different. The change which occurred in the early 1950s was a change in the *kinds* of problems and variables considered important to advance our knowledge. The present change of emphasis is one dealing with the *way* in which industrial psychology approaches the study of man at work. It is a change in philosophy rather than a change in content.

What is this philosophical change? Quite simply, it is a growing tendency for modern industrial psychologists to become increasingly involved in the development of theory. The emphasis is changing from a philosophy of rigid empiricism applied to a variety of seemingly unrelated situations and phenomena to a philosophy concerned with the establishment of underlying behavioral principles which will permit a fuller understanding of man's behavior in realistic task environments. From a scientific viewpoint, it is a more mature and exceedingly richer philosophy than its predecessor. As authors, we believe that if industrial psychology is to evolve as a science, its theoretical basis must be developed and emphasized.

Actually, such a transition had to occur—it is inevitable in any maturing scientific area. However, it has been long overdue in industrial psychology, and this continuing lack of attention to theory has been a major source of difficulty in advancing the scope of knowledge within the field. The difficulty was most aptly stated by Dr. Brent Baxter in the closing paragraph of his presidential address to the Division of Industrial Psychology in September, 1965:

> In summary, my message urges you, each one of you, to turn some part of your attention to the development of your own broad theory of human behavior. At least

Notice from the above relationship that only when r_{pp} and r_{cc} are unity (perfect reliability) will the obtained validity be equal to the true validity. As the reliability of the two measures decreases, so will the obtained validity. For example, suppose $r_{pc_{(true)}} = 0.60$, $r_{pp} = r_{cc} = 0.80$, then $r_{pc_{(obtained)}} = 0.60\ \sqrt{0.80 \times 0.80} = 0.60\ (0.80) = 0.48$. Also notice that if the reliability of either the predictor or criterion is zero, then the obtained validity will be zero as well.

CRITERION RELEVANCE

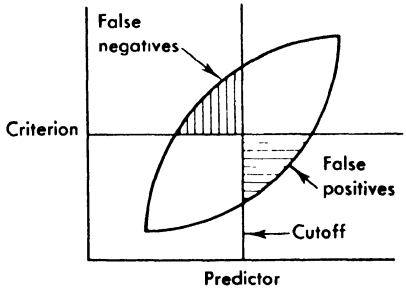
The relevance of a criterion has little to do with the actual empirical utility of a prediction instrument, although it has a great deal to do with its logical utility. This topic will be discussed further in Chapter 6.

A DECISION-THEORY APPROACH

The problem of selection can be viewed from a somewhat different perspective than the one used in the earlier portions of this chapter. This second approach proves interesting in that we shall find that predictor *validity* may not be as important a variable in selection as the traditional viewpoint makes it out to be. Our new perspective is one based upon a decision theory model. We should begin by restating the objective in a typical selection situation. *In many selection situations we wish to establish a cutting score on our predictor which will result in minimizing our decision errors.*

Implicit in this type of situation is the assumption that the selection ratio can be manipulated at will; that is, it is not "fixed" at some value. Also implicit is the notion that our criterion variable can be meaningfully separated into two or more distinct groupings such as "successful" and "unsuccessful." Our objective is to manipulate the cutting score (which is the same as manipulating the selection ratio) in order to *minimize the number of errors made in our process of deciding whether a person should be hired or rejected.*

Earlier we pointed out that there were two distinct types of decision errors in the selection paradigm, false positives and false negatives, as shown below:



Our objective then is to find the cutoff point which will result in the smallest number of total errors. For purposes of convenience we shall start by assuming that both types of error are considered *equally costly*. That is, we have no reason to prefer making a false positive error over a false negative error, or vice versa. By

through 20 cover work variables such as job analysis, accidents and fatigue, the work environment, principles of human performance, and general industrial problems.

While the chapter order in a text such as this is always an arbitrary matter, the present organization of the material seems to the authors to form a logical sequence. However, the book has been written so that any chapter can be read separately or out of the present sequence.

ACKNOWLEDGMENTS

Students and clients highly deserve recognition as contributors to the thoughts expressed in this book. The challenging questions on the parts of students have resulted in the development of both hypothesis and theory.

The many authors and publishers quoted have been most kind in granting their generous permissions.

A strong vote of appreciation should go to those colleagues who were willing to read and comment critically on various portions of the text. In particular, recognition should go to Dr. Norman L. Vincent, Dr. Ralph Stogdill, Dr. Robert J. Wherry, Jr., Dr. Irwin L. Goldstein and Dr. Arthur L. Dudycha. Also to be commended are those students who served as "test subjects" for the manuscript as it underwent revision. Their feedback and comments were invaluable.

Finally, a very special debt of gratitude is owed to Mrs. Geraldine Hickman for her outstanding secretarial assistance over the entire period of the preparation of this manuscript.

MILTON L. BLUM
JAMES C. NAYLOR

INDUSTRIAL PSYCHOLOGY

INTRODUCTION 1

Psychology is the study of man with the aid of scientific methodology. It uses the experimental method—observation under controlled conditions—to gather data. It also uses other methods for data collection, such as the case history method and the longitudinal study of development as it takes place. Psychology accepts introspection (a subject's report that describes mental processes) as a source of data. Regardless of the data collection method, the psychologist is only interested in those findings which can be readily verified and duplicated under similar conditions. The psychologist would refer to this requirement as "replication" of data.

It is extremely important that the reader understand from the onset several points about the field of psychology. First, most people who have not been exposed to the field in any formal fashion associate the psychologist almost exclusively with the study and treatment of mental illness and abnormal behavior. While it is true that within the profession the specialty of clinical psychology and to some extent that of counseling psychology do emphasize the study and treatment of the abnormal behavior of man, the remaining specialties are more concerned with the study of "normal" behavior.[1] After all, most of our society probably would be classified under the heading of normal (if we stretch a point here and there); thus, the psychologist who wants to understand human behavior and develop theories and principles about such behavior will usually confine himself to these "normal" people for purposes of data collection.

A second point is that the term *psychology* is not synonymous with "common sense." Often when "common sense" is applied, the conclusions are incorrect because of insufficient information or a confusion of cause and effect. Another term used incorrectly in connection with psychology is *human nature*. People often attempt to explain behavior by attributing it to "human nature." This is merely a verbalism; it does not explain the behavior. Sometimes the term implies that people act in a certain way because of inherited predispositions. The assumption of an inherited predisposition without valid proof does not stand up in the science of psychology.

THE PSYCHOLOGIST

The profession of psychology is rather complex; it consists of many specialties and interests which are in some way similar and in other ways quite dissimilar. In 1966 there were approximately 30,000 psychologists in the United States. These psychologists differed widely in their fields of interest, place and location of employment, etc. Figure 1.1 presents some descriptive information about United States psychologists.

There are more men than women psychologists, although the ratio varies somewhat with particular subfields of psychology. Thus, from Figure 1.1a we see that there are five men for every two women in the areas of clinical and educational-school psychology, while the ratio changes to nine men to every woman in industrial and personnel psychology. Of the total number of psychologists, as shown in Figure 1.1b, by far the largest number are employed by universities and colleges (about

[1] The reader should not be misled into thinking that there exists a sharp and clear dichotomy between the abnormal and the normal in terms of behavior.

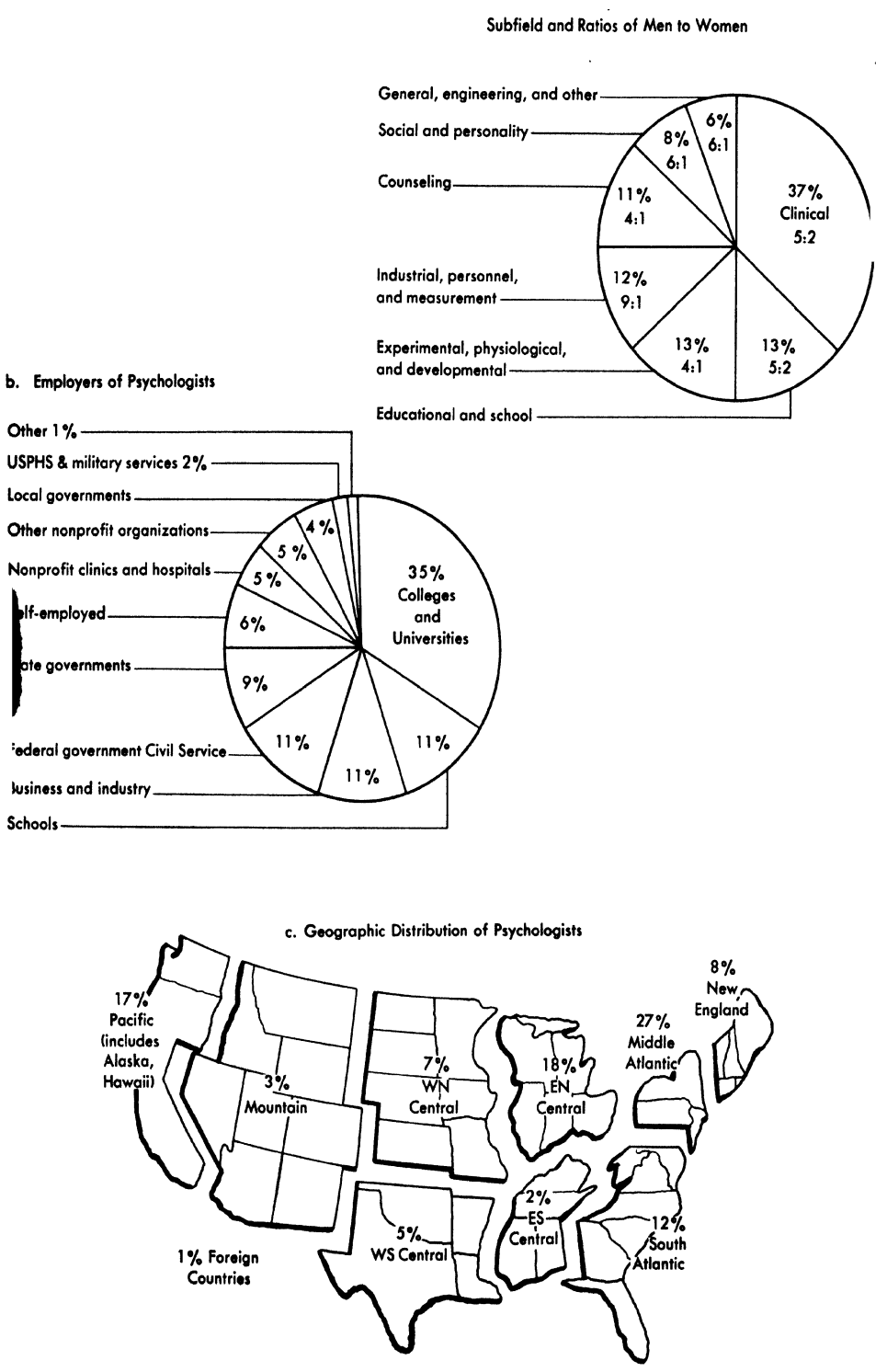


Figure 1.1. Various characteristics of psychologists in the United States. (Adapted from A *career in psychology*. American Psychological Association, Washington, D.C., 1963, 8, 9, 12.)

35 percent), with federal, state, and local agencies of all kinds (excluding schools) employing the second largest number (about 25 percent). Other substantial categories of employment for psychologists are business and industry (about 11 percent) and nonprofit organizations (about 10 percent). Only a relatively small percentage (6 percent) of psychologists are self-employed. Geographically, over one-fourth of all United States psychologists are located in the Middle Atlantic states; the East North Central states (18 percent) and the Pacific states (17 percent) are the other two major geographical locations with substantial numbers of psychologists.

Although the majority of psychologists appear to be located in the coastal states, the majority of the larger training programs are located in the Midwest. Table 1.1 lists the 13 departments with the largest number of Ph.D. graduates in psychology during the academic year 1962–1963. Nine of the thirteen are located in the Midwest, three are Eastern schools, and one is located on the West Coast.

TABLE 1.1 *Doctorates Awarded by the Largest Departments in the Academic Year 1962–1963*

Department	*Doctorates Awarded*
1. Columbia University	45
2. University of Michigan	43
3. Purdue University	40
4. Ohio State University	22
5. University of Minnesota	22
6. New York University	22
7. State University of Iowa	21
8. University of California (Berkeley)	20
9. Indiana University	20
10. Michigan State University	18
11. University of Massachusetts	16
12. Western Reserve University	15
13. University of Illinois	15

SOURCE: Adapted from R. F. Lockman and S. Ross. Survey of graduate education in psychology: Some trends of the last decade. *American Psychologist*, 1964, *19*, 623–628.

Table 1.2 also shows some interesting figures on the relative numbers of psychologists with different specialties who were given degrees the same year (1962–1963). By far the largest percentage of those receiving the Ph.D. degree were those specializing in clinical psychology. Note that industrial Ph.D.'s accounted for only 5.2 percent of the total. With the master's degree quite a different story appears: Those specializing in industrial psychology are the largest single group (almost 25 percent), followed by the specialty areas of educational, experimental, and clinical psychology in that order.

TABLE 1.2 *Types of Graduate Degrees Granted in Psychology in the Academic Year 1962–1963*

	Doctorates		*Masters*	
Area of Specialization	*Number of Degrees*	*Percent of Field*	*Number of Degrees*	*Percent of Field*
Clinical	320	36.4	290	16.1
Experimental	257	29.2	301	16.8
Social and personality	84	9.5	57	3.2
Industrial	46	5.2	446	24.8
Counseling and guidance	43	4.9	157	8.7
Educational	36	4.1	370	20.6
General	28	3.2	23	1.3
Developmental	20	2.3	19	1.1
Psychometrics	19	2.2	35	1.9
School	15	1.7	60	3.3
Engineering and human factors	2	0.2	13	0.7
Unspecified and others	10	1.1	25	1.4
Total:	880		1796	

SOURCE: Adapted from R. F. Lockman and S. Ross. Survey of graduate education in psychology: Some trends of the last decade. *American Psychologist, 19,* 623–628.

INDUSTRIAL PSYCHOLOGY

Industrial psychology draws upon the facts, generalizations, and principles of psychology. It uses the method prescribed in the parent body. Because it applies the techniques of psychology to the industrial scene and the problems confronting it, industrial psychology formulates and modifies procedures to meet the conditions found in business rather than in the laboratory.

Industrial psychology is simply the application or extension of psychological facts and principles to the problems concerning human beings operating within the context of business and industry.

The most important aspect of industrial psychology is its discipline. It clearly recognizes that scientific conclusions must be objective and based upon facts gathered as a result of a defined procedure. It does not treat assumptions or hypotheses as if they were conclusions. Its findings sometimes confirm the obvious and very often are not romantic. Whereas chemists are no longer confused with alchemists, and psychologists are differentiated from physiognomists, the same clear-cut differentiation does not apply to industrial psychologists in relation to their "gold-brick" salesmen breathren who sell the impossible with all the authority of ignorance.

The misinformed often believe that all that is necessary to have a knowledge of psychology is to be a human being with common sense and to be a student of human nature. But a psychologist is quite a different person. He has professional training, is aware of the limitations of his knowledge, and above all, has been indoctrinated in a methodology to obtain conclusions based upon data acquired according to certain prescribed scientific methods.

THE INDUSTRIAL PSYCHOLOGIST

The industrial psychologist may be employed by a company. He may be a consultant to a trade association, to retailers or manufacturers. He may also be a university professor who does research in the field either privately as a consultant or as an associate in the university institute. In the future he may be on the staff of a labor union. Regardless of the field in which he works, the industrial psychologist must report research findings accurately. Hence, there is a great need for an unbiased approach if he is to make his full contribution to society.

Although industrial psychologists have been employed predominantly by large companies, this does not mean that small business organizations cannot benefit from the correct application of the principles and generalizations of industrial psychology. As more students of psychology become interested in the industrial field, and as more small businessmen are made aware of the benefits of industrial psychology, greater advantages will accrue to our society. Because small and large business organizations have similar problems, the findings are likely to be applicable to both.

Since many states do not license industrial psychologists—or any psychologists, for that matter—the question arises as to how one becomes an accredited industrial psychologist. Legitimate recognition occurs primarily along educational and professional lines. The American Psychological Association (APA) is the professional body of psychologists; similar associations exist among dentists, physicians, sociologists, etc. The APA has three classes of membership. The *Associate* is a person who meets such minimum standards as either two years of graduate work in psychology at a recognized graduate school or a master's degree in psychology from a recognized graduate school. At the time of application the candidate must be devoting full time to either graduate or professional work in psychology. The minimum standard for election to APA's *Member* status is the receipt of a doctoral degree based in part upon a psychological dissertation and conferred by a graduate school of recognized standing. *Fellows* represent a higher status of membership; they have the Ph.D. degree, five years of acceptable professional experience subsequent to the granting of the degree, and are elected to Fellow status for outstanding scientific achievement in their chosen area of specialization. Life membership, not likely to be of immediate concern to students, is reserved for members of twenty years' standing who have reached sixty-five years of age.

One of the divisions of the American Psychological Association is Division 14, known as the Division of Industrial Psychology. It has three classes of membership, Associate, Member, and Fellow, which require similar standards as membership in the APA except that the experience requirement must be related to industrial experience as a psychologist. In 1966 this division had approximately 875 members; the entire APA had about 24,000 members. Not all psychologists belong to the APA, of course. In 1966 there were an estimated 6000 psychologists who met the requirements but did not join.

A further degree of attainment outside of APA is the diploma, which is a certification of specialty. This diploma is awarded by the American Board of Examiners in Professional Psychology. Certification procedures have been set up in clinical, counseling and guidance, and industrial psychology. Standards for diplomate status are the Ph.D., five years of qualifying experience, and the passing of written examinations in the specialized field plus an oral examination after the written. In 1966 there were

about 1900 diplomates; about 180 of these held their diploma in industrial psychology.

Sawyer (1960)[2] reports some interesting facts about the membership of Division 14 taken from a survey of 708 Division members in 1959 (there was a total membership of 716 that year). Of the 81 percent with the Ph.D., over half had received their degrees since 1948, with the breakdown by decades being:

1900–1909	1
1910–1919	6
1920–1929	32
1930–1939	120
1940–1949	160
1950–1958	252

TABLE 1.3 *Schools Granting Degrees in Industrial Psychology 1906–1958*

School	*Number of Graduates*
Purdue University	83
Columbia University	63
Ohio State University	63
New York University	41
University of Chicago	30
University of Minnesota	29
State University of Iowa	26
Western Reserve University	26
University of Michigan	22
University of Pennsylvania	21
Harvard University	18
Northwestern University	17
Pennsylvania State University	16
University of Pittsburgh	16
Stanford University	15
University of Southern California	15
Cornell University	11
Yale University	11
University of California (Los Angeles)	10
University of Wisconsin	10
University of California (Berkeley)	9
66 other schools	156
Total	708

SOURCE: Adapted from J. Sawyer. The industrial psychologist: education and employment. *American Psychologist*, 1960, *19*, 670–673.

[2] Full sources for authors or works mentioned but not footnoted are given in the section References" which appears at the end of each chapter.

Those universities which Sawyer found to be most productive in graduating Ph.D.'s in industrial psychology are shown in Table 1.3. Sawyer's survey also indicated that the general membership in Division 14 could be categorized into four general employer groups:

Industry	35%
University	26%
Consulting firms	28%
Government	11%

Percentages extremely close to these were found in a later study by Dunnette (1961).

Of course, not all industrial psychologists belong to the APA or to Division 14. Although it is exceedingly difficult to obtain complete figures, it has been estimated that there were about 2500 psychologists employed on a full-time basis in industry in 1963 (Ross and Lockman, 1963) over three times the number belonging to Division 14. This is considerable growth from earlier estimates of 1000 in 1958 (McCollum, 1959) and of 600 estimated by the APA in a 1954 article on the profession (APA, 1954). It is worth noting that these numbers probably represent lower bound estimates of the number of psychologists involved in the activities of business and industry at each of these time periods. Traditionally, many of the psychologists who hold academic positions at various universities spend a portion of their time in a consulting capacity with various business organizations.

SCOPE OF THE INDUSTRIAL PSYCHOLOGIST'S WORK

Perhaps one of the best ways to begin to understand the complex field of industrial psychology is by enumerating the various types of activities which can be considered as part of the general domain of the profession. Actually, even this is not an easy task. Many authors have attempted to set up classifications for describing the activities of the industrial psychologist, with various degrees of success (e.g., see McCormick, 1955; Stagner, 1957; and Taft, 1946). Several have taken the approach that industrial psychology must be "what industrial psychologists do" and have tried to empirically determine exactly what it is that they do. For example, Canter (1948) analyzed the results of a questionnaire returned by 56 respondents in business and industry, 37 consultants, and 10 psychologists in the field of advertising. The results indicated that employers were generally large corporations, but that consultants' clients were both large and small companies. The psychologists in advertising were generally employed by advertising agencies. Table 1.4 presents an analysis of most important work functions of the respondents.

McCollom (1959) actually interviewed 75 psychologists in 20 different cities, having each one describe his own activities and the activities of other psychologists working in industry with whom they were familiar. He found that the activities of these people could be grouped into the following general categories:

1. *Personnel selection:* Selection and assessment of employees and executives, criterion research
2. *Personnel development:* Performance appraisal, attitude measurement, management development, employee counseling

3. *Human engineering:* Equipment and product design
4. *Productivity study:* Activities concerned with worker fatigue, lighting, general job environment
5. *Management:* Activities involving administrative skills
6. *Other:* Accidents and safety, labor relations

TABLE 1.4 *Analysis of Most Important Work Functions, by Groups*

	Percentage of Time
Business and industry group (N 56)	
Duties involving:	
Personnel functions	33
Scientific, research, and developmental functions	25
Policy and management functions	20
Labor relations functions	12
Education and training functions	5
No response	5
	100
Consulting group (N 37)	
Duties involving:	
Personnel functions	25
Policy and management functions	24
General consulting functions	21
Research and development functions	11
Clinical functions	10
Labor relations functions	3
Market research functions	3
No response	3
	100
Advertising group (N 10)	
Duties involving:	
Market research functions	70
Policy and management functions	30
	100

SOURCE: R. R. Canter. Psychologists in industry. *Personnel Psychology,* 1948, *1,* 145–161.

Perhaps the most authoritative source on the functions and activities which comprise industrial psychology is Division 14 itself, since it exists as the official organization of all psychologists concerned with the application of psychological knowledge to the world of industry. In a 1959 report entitled *The Psychologist in Industry,* the Division cited seven major areas which comprised the content area of industrial psychology. These were: (1) selection and testing, (2) management development, (3) counseling, (4) employee motivation, (5) human engineering, (6) marketing

research, and (7) public relations research. These were very similar to the functions outlined by Taft (1946) who listed job analysis; motion studies, salaries and wages, selection of new employees; transfers, promotions, and terminations; training; problem employees; employee rating; industrial hygiene; morale and research. In summary, it would appear that the psychologist employed by an industrial organization is likely to find himself required to tackle nearly any problem involving people. This is made quite apparent by the Division 14 report, which lists over 150 more specific kinds of problems with which industrial psychologists concern themselves.

An example of the varied nature of the work is given in an article by Bills (1934) describing a typical workday of an industrial psychologist. The day started with a conference at 9 A.M. which resulted from an interview held the preceding afternoon. The conference was to decide on the placement of an employee for a two-month

TABLE 1.5 *Representative Firms Employing Diplomates in Industrial Psychology and Diplomates' Titles*

Firm Name	*Title*
American Home Products Corp.	Director of Personnel
Atlantic Refining Co.	Research Assistant
B F. Goodrich Co.	Coordinator of Training
Carbide & Carbon Chemical Co.	Asst. Director, General Industrial Selection Department
Chrysler Corp.	Educational Supervisor
Columbia Broadcasting System, Inc.	President
Commonwealth Life Insurance Co.	Director of Research
Continental Oil Co.	Director of Advertising
D E. McNicol Pottery Co.	Vice-President and General Manager
Detroit Edison Co.	Industrial Psychologist
E. I. du Pont de Nemours & Co.	Manager, Personnel Research Section, Pres. Div
Fieldcrest Mills	Department of Personnel Research & Training
General Motors Corp.	Chairman, Personnel Evaluation Services
General Shoe Corp.	Director Supervisory Training
Gulf Oil Co.	Head, Training Unit
Harwood Manufacturing Co	President
International Business Machine Corp.	Coordinator, Educational Research
Knox Reeves Advertising, Inc.	Vice-President, Marketing Director
Life Insurance Agency Management Association	Research Associate
Marshall Field Co.	Vice-President
Needham, Louis and Brorby	Vice-President
Procter & Gamble	Head, Research Department
Prudential Insurance Co.	Asst. General Manager, Director of Personnel Research
S. C. Johnson & Son, Inc.	Personnel Director
Standard Oil Co.	Advisor, Employee Relations Research
United States Steel Corp.	Manager, Personnel Department
Wm. Esty Co.	Executive Vice-President

period. This employee had been diagnosed as mentally ill, and the company was attempting therapy by a psychiatrist rather than resorting to immediate dismissal. The day's second problem for the psychologist was an attempt to predict whether a person's ability to punch Hollerith cards could be ascertained in three weeks. The next problem concerned the transfer of two employees to fill two vacancies. Problems concerning salary levels and rating scales occupied the remaining portion of the morning. Immediately after lunch the psychologist had to meet with the Employee's Loan Fund Committee; the particular problem facing them concerned a misapplication of some funds. The psychologist next talked with a cleaning woman who felt that the supervisor was not giving her a fair deal. Another employee who had been married for six months asked the psychologist for advice on how to inform her disapproving parents.

Actual test development and research thus played only a small role in this particular day's work. The psychologist who is on the staff of an organization is likely to find that the specific minor problems which arise each day may interfere with his major work. He must therefore be flexible and prepared to handle a number of tasks and projects concurrently.

Not only are the duties of the industrial psychologist varied, but his title ranges from *President* to *Industrial Psychologist* to *Educational Supervisor*. Table 1.5 lists business and industrial firms that employ diplomates in industrial psychology and their titles. The firms selected for inclusion in this table are merely representative. The table is intended to give an idea of the range of titles and not to be a complete listing of companies employing industrial psychologists.

Table 1.5 reveals that few industrial psychologists are assigned the title of the profession. The clear picture is that the psychologist is assigned a title related to his job function or duty.

In addition to the diplomates employed in business and industry, psychologists work for many other companies, including the following. Again, no attempt is made to furnish a complete list but, rather, to give an idea of the wide range of companies employing psychologists.

Aetna Life Insurance Co.
American Can Co.
American Viscose Corp.
Armco Steel Corp.
Caterpillar Tractor Co.
Continental Can Co.
Corning Glass Works
Creole Petroleum Corp.
Eli Lilly & Co.
Esso Standard Oil
Ford Motor Co.
General Motors Corp.
Hughes Aircraft
International Harvester Co.
Kimberly-Clark Corp.
McCann-Erickson, Inc.
Metropolitan Life Insurance Co.
Midland Cooperative Wholesale
Minneapolis-Honeywell Regulator Co.
Minnesota Mining & Manufacturing Co.
Pittsburgh Plate Glass Co.
Rand Corp.
RCA: Victor Division
State Farm Insurance Co.
The Chesapeake & Ohio Railway Co.
The Dayton Co.
The Prudential Insurance Co. of America
Washington Gas Light Co.
Young & Rubicam

CONSULTING ORGANIZATIONS

An industrial psychologist, in addition to being directly employed by a business organization, is also found in many consulting organizations. Today consulting on problems of a psychological nature is "big business." Most psychological consulting firms may be classified into one of two possible categories: Those who emphasize the area of psychological testing and those who emphasize the nontesting aspects of industrial psychology. Admittedly the separation is generally not a distinct one, particularly with the very large consulting firms who tend to handle both categories. However, the emphasis usually still remains visible. Following are some brief descriptions of several of the most prominent consulting firms typical of each of the above categories.

CONSULTING FIRMS: TESTING

The oldest and largest organization of this type is the Psychological Corporation which was organized in 1921 by a group of psychologists. The Psychological Corporation is organized into the following divisions: Industrial, Test, Personnel and Market Research, Professional Examinations, and the Experimental Laboratory.

The Industrial Division installs testing programs for companies, improves training programs for management personnel, and trains supervisors in various techniques and principles in human relations. The Test Division publishes and sells psychological tests to schools, government agencies, and industrial firms. The Personnel and Market Research Division does research on consumer motivation and behavior, public relations, and media; it also carries on activity in personnel selection attitudes and job evaluation. The Professional Examinations Division has evaluated several hundred thousand nursing school applicants and currently also conducts entrance examination programs for schools of veterinary medicine and pharmacy. The final division, the Experimental Laboratory, undertakes research projects in applied psychology, such as the development of test scoring machines, research in logical abilities, etc. The Psychological Corporation employs about 25 Ph.D.'s in its various divisions.

A second major consulting firm which emphasizes psychological testing is Science Research Associates (SRA). This corporation specializes in developing material and services for education, industry, and government based upon research in the behavioral sciences. It offers a variety of publications and services for instruction and training, educational and vocational guidance, personnel selection and placement, evaluation of knowledge and performance, and the processing of specialized data.

The third of the "big three" in testing is the Educational Testing Service (ETS), a nonprofit organization founded in 1947 through the combined efforts of the American Council on Education, the Carnegie Foundation, and the College Entrance Examination Board. Its major purposes are to develop tests for schools, colleges, and the psychological profession; to assist the users of tests in methods and procedures; and to do research in psychological testing. In 1966 over 70 psychologists were listed by the APA as being affiliated with ETS (APA Directory, 1966). In the fiscal year ending June 30, 1966, ETS had a gross income of over $23 million —ample testimony to the importance of testing in our current culture.

CONSULTING FIRMS: GENERAL

The firm of Richardson, Bellows, Henry and Company, Inc. was organized shortly after World War II. Its major office is in New York City, but it has branch offices in other major cities. This company primarily deals with employee attitude measurement, executive appraisal, the development and improvement of training programs, manuals for client companies, and research in test development and employee evaluation.

Dunlap and Associates, Inc. is a consulting firm organized in 1948 to provide research and consulting services in the fields of executive evaluation, marketing research, human engineering, and training. Its staff now numbers approximately 200, with about one-fourth of the professional staff possessing a Ph.D. In its 1964 Annual Report, it showed a gross income of nearly $4 million (see Figure 1.2).

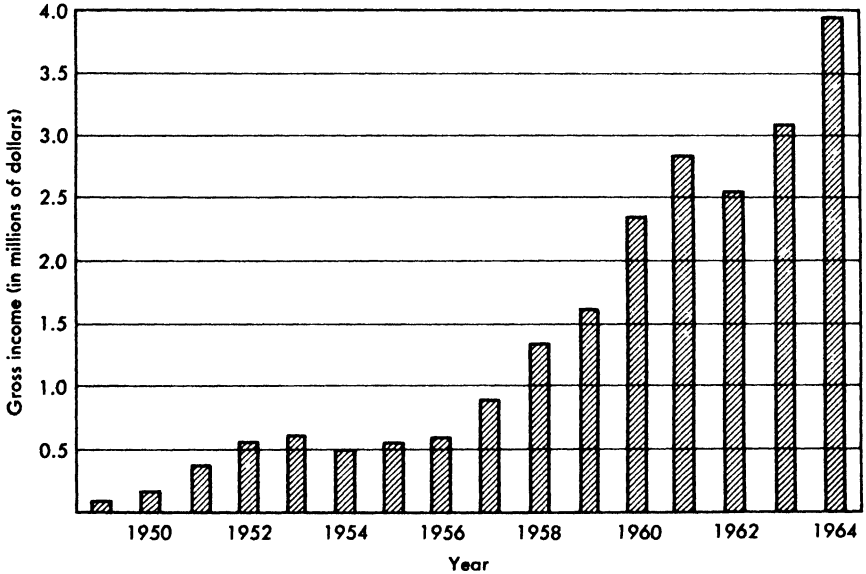


Figure 1.2. Gross income of Dunlap and Associates, Inc. from 1949 to 1964. (Adapted from the 1964 Annual Report, Dunlap and Associates, Inc., Stamford, Conn.)

Rohrer, Hibler, and Replogle have offices in six cities and primarily provide psychological services to business leaders. Their approach is first to obtain a psychological evaluation of the "problem" executive and then to conduct a series of individual conferences with him designed to change his behavior into more desirable patterns of action (Flory and Janney, 1946). This company was founded in 1927 and has five senior partners, all of whom have a Ph.D.

The American Institutes for Research in the Behavioral Sciences (AIR) is a nonprofit research organization which carries out a variety of research programs in human engineering, training, organizational behavior, and education. Its major offices are in Pittsburgh, with subsidiary offices in Washington, Los Angeles, and Palo Alto. During the 1962–1964 period AIR's average number of full-time employees was around 230 (AIR Annual Report, 1964).

RESEARCH CENTERS AT UNIVERSITIES

The practical aspect of solving problems by using research techniques was intensely demonstrated during World War II. The armed forces did not have the personnel or the unlimited facilities to do all the work that was needed. Accordingly, many universities established research centers to handle problems related to industrial psychology and other areas of psychology. This situation has continued, and now there are many such centers handling on a contract basis research needed by many government agencies.

Although research in personnel problems and human engineering is of primary interest for our subject matter, additional work is done in the fields of physiological and social psychology, and in such topics of general psychology as learning, retention, and the higher cognitive processes.

Research contracts are awarded by the Department of Defense as well as other branches of the government such as the Atomic Energy Commission, the Department of the Interior, the Veterans' Administration, and the National Aeronautics and Space Administration.

Specific problems related to job analyses and specifications; selection, classification, and criteria research; training; psychometrics; and human engineering are solved by contract research. Examples of nonsecurity contracts are as follows: billet analyses for guided missiles personnel; a study of the relationships between Navy billets and civilian occupations; submarine personnel selection; research on supervisor selection; radar mechanics functional knowledge test battery; effectiveness in technical training; speed factors in tests and in criteria; causation of accidents; the worker as a factor in equipment design; display and signal pattern discrimination; and attention value of warning signals. There are many, many others.

Possibly the largest university research center in psychology is at the University of Michigan. It is known as the Institute for Social Research and has two main divisions, the Survey Research Center and the Research Center for Group Dynamics. Although the staff of highly competent professionals is primarily concerned with social psychology, the work clearly shows the intimate interrelationship of social and industrial psychology.

The Survey Research Center is more concerned with problem-oriented than with discipline-oriented research. This means it promotes interdisciplinary research combining the efforts of psychologists, economists, anthropologists, sociologists, and political scientists. Its research undertakings are sponsored by government agencies, private business, and research foundations. An example of one of its major concerns is its work in the field of psychological economics and consumers' economic attitudes. On the other hand, the Research Center for Group Dynamics is primarily concerned with the phenomena of group behavior and with an attempt to derive the principles of group dynamics which lead to group formation, change, or dissolution. It has worked in the areas of group productivity, communication, and intergroup relations.

These two centers together with the parent body employ over 500 full- and part-time personnel, including a permanent part-time force of about 250 field interviewers residing in communities across the nation. The organization chart of the Institute (Figure 1.3) depicts its wide range activities. The Institute has published

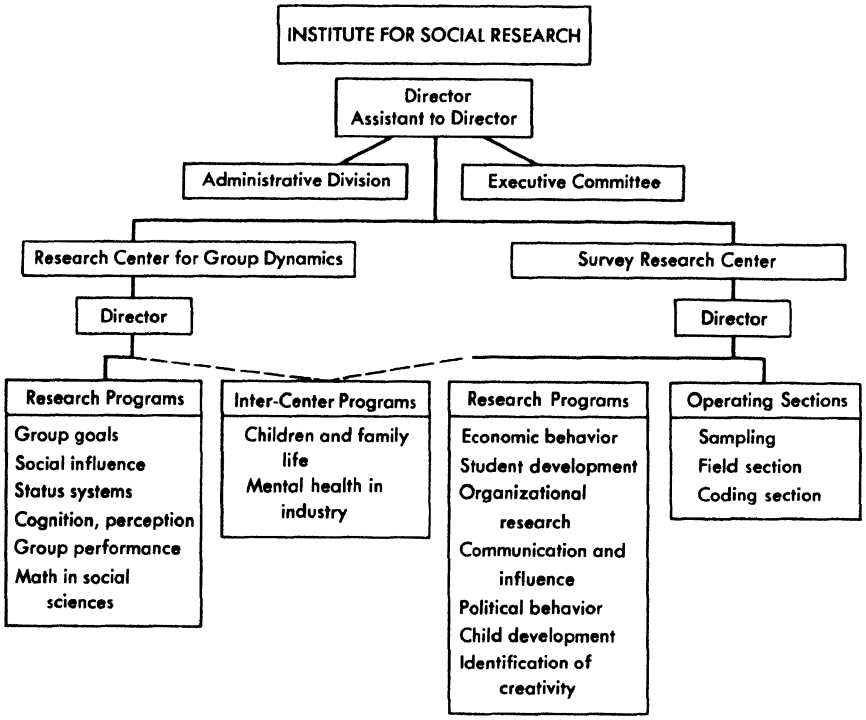


Figure 1.3. Organizational chart describing the structure of the University of Michigan's Institute for Social Research in 1963.

well over 2000 articles, reports, and books as a result of its research activities (Seashore, 1962).

Another notable but less clearly structured interdisciplinary research center is the joint research composite at Ohio State University. This composite is made up of three distinct suborganizations, the Bureau of Business Research, the Behavioral Sciences Laboratory, and the Administrative Services Division.

CENTERS OF LABOR AND INDUSTRIAL RELATIONS

A number of universities have well-established centers or institutes related to the problem of labor and industrial relations. The Industrial Relations Center of the University of Minnesota is concerned with a program of training and research to improve labor-management relations and to facilitate cooperation between the two. The Center recognizes that such work cannot be solely the province of any one department; thus it actually draws from eight different departments, one of which is psychology.

The Institute of Industrial Relations of the University of California has two divisions, one at Berkeley and the other at Los Angeles. It pursues three major lines of endeavor—research, instruction, and community relations—and runs many conferences and lecture series as well as resident institutes for labor and for management. It also accepts the importance of the interdisciplinary approach.

Other outstanding programs of this nature are located at Cornell University (The New York State School of Industrial and Labor Relations) and at the University of Illinois (Institute of Labor and Industrial Relations).

HUMAN ENGINEERING LABORATORIES

Several university psychology departments have research laboratories in which the emphasis has been directed toward research on human engineering topics. Probably the best known of these is the Laboratory of Aviation Psychology (recently renamed the Human Performance Center) at Ohio State University. In addition, the University of Illinois's Aviation Psychology Laboratory and the University of Michigan's Engineering Psychology Laboratory are equally well-established centers of human engineering research.

OTHER RESEARCH ORGANIZATIONS

There are two other important types of organizations in which industrial psychologists are employed: nonprofit organizations sponsored by industry, and various governmental agencies which specialize in research problems relevant to industrial psychology.

NONPROFIT ORGANIZATIONS SPONSORED BY INDUSTRY

A number of nonprofit research and consulting organizations have been sponsored and funded by companies in a common industry to do research for the entire group of member companies. A typical example is the Life Insurance Agency Management Association (LIAMA) formed in 1922. It now has over 360 full-member participating companies and 118 associated members scattered all over the world. It is organized into such divisions as company relations, research, actuarial, institutional relations, and administrative. The three major parts of the research division are Human Resources Research, Financial Management and Functional Cost Research, and Market and Survey Research. Typical projects include studies of agent and managerial selection procedures, training of agents, and the measurement of public attitudes toward life insurance.

The National Industrial Conference Board was founded in 1916 as an independent and nonprofit institution for business fact-finding through research. It is supported by more than 3000 subscribing associates from business organizations, trade associations, labor unions, libraries, and colleges. The research program is carried on by four major divisions. The Division of Personnel Administration is of greatest interest to the industrial psychologist. Its reports serve as useful guides to those who want to learn from case studies as well as have a manual type of reference.

Other examples of such organizations are the Committee on Highway Safety Research and the W. E. UpJohn Institute.

GOVERNMENT RESEARCH AGENCIES

As was pointed out earlier in the chapter, many industrial psychologists are employed by the United States Government. Examples of government agencies which have traditionally been involved in psychological research are listed below:

U.S. Army Personnel Research Office
Washington, D.C.

U.S. Air Force Personnel Research Laboratory
San Antonio, Texas

U.S. Navy Aviation Psychology Laboratory
Pensacola, Florida

U.S. Navy Electronics Laboratory
San Diego, California

U.S. Air Force Behavioral Sciences Laboratory
Dayton, Ohio

U.S. Navy Submarine Base
New London, Conn.

Two of the largest single agency employers of psychologists are nongovernmental research and development agencies. However, both of these agencies work primarily for governmental agencies. HumRRO, the Human Resources Research Office of George Washington University, was established in 1951 to apply scientific methods to the solution of army problems in training, motivation, and leadership. The System Development Corporation (SDC) was later established to perform a somewhat similar function for the Air Force. As of 1964 HumRRO employed approximately 75 research psychologists and SDC employed nearly 150.

The organizations mentioned do not exhaust the list of industries, consulting organizations, centers at universities, cooperative research centers supported by industry, or government agencies. They have been selected merely as examples, but of course all pass rigid test of authenticity and compliance with stated purpose. Others too numerous to mention also meet such standards.

Examples have been cited to impress the reader with the realness of research. Each organization has been described all too briefly, but the bibliographic references can furnish the necessary additional facts.

HISTORY AND DEVELOPMENT OF THE FIELD

It is exceedingly difficult to apply a date to the founding of any discipline. However, industrial psychology may have gotten its start on December 20, 1901. That was the evening that Dr. Walter Dill Scott, a psychologist at Northwestern University, gave an address discussing the potential application of psychological principles to the field of advertising (Ferguson, 1962). This was followed by a series

of 12 magazine articles which were subsequently combined in 1903 into a book entitled *The Theory of Advertising* (Scott, 1903)—undoubtedly the first book which involved the application of psychology to the world of business. Scott, who later became president of Northwestern University, followed up this initial book with several others in the next few years on the same general topic (Scott, 1908a, 1908b). He also published several books in which he tried to bring psychology to bear upon the broader domain of business per se, rather than just advertising (Scott, 1911a, 1911b). In fact, Dr. Scott's accomplishments as the first industrial psychologist are far too numerous to detail here; the reader is referred to more appropriate sources (Ferguson, 1962).

In spite of Scott's prominent activity and many published works during the years 1901 to 1913, it is interesting that he has often been ignored in historical accounts of the field of industrial psychology. This is particularly so in past years. Today he seems to be gradually regaining his rightful place as the man who really "started it all."

Where Scott has not been credited with being the first industrial psychologist the honor has usually been awarded to Hugo Münsterberg. In 1913 he published his text *Psychology and Industrial Efficiency* (Münsterberg, 1913). Münsterberg as an author tended to be somewhat less aggressive and flamboyant than Scott in his approach to applying psychology to business. Thus in his book the differences of approach between the pure and the applied science are carefully reflected in the first few chapters. Münsterberg writes cautiously and defensively of his attempts to establish an applied field of psychology as a necessary counterpart of the pure field. Münsterberg's book, which has served as a model for the development of industrial psychology, includes such topics as learning, adjustment to physical conditions, economy of movement, monotony, fatigue, and buying and selling.

During World War I psychologists were quite active in the war effort, developing group tests for army recruits and aiding in the development of procedures for the selection of officer personnel. In fact, many of the major postwar developmental areas of industrial psychology such as group testing, trade testing, rating scales, and the personality inventory had their roots in the activities of psychologists in the World War I war effort.

In 1917 the *Journal of Applied Psychology* made its appearance; the importance of this periodical in the expansion of the field will be apparent from the number of references throughout this book to articles which have appeared in it. At about the same time, colleges began to introduce courses in applied psychology; and as the subject has developed, the trend has been to offer courses in specific fields of applied psychology such as industrial psychology, personnel psychology, vocational psychology, and advertising psychology.

In 1919 Scott and several others founded the Scott Company of Philadelphia. This was the first psychological consulting firm ever organized and was oriented toward personnel problems in industry. This was followed two years later by the previously mentioned Psychological Corporation—the oldest firm of this type still active in the United States.

During the post-World War I years industry first began to show an interest in the discipline of industrial psychology. Certain firms such as Procter & Gamble, the Philadelphia Company, and the Hawthorne Plant of Western Electric all formed their own personnel research programs. In fact, it was at the Hawthorne Western Electric plant that the famous "Hawthorne" studies were begun in 1927 (Roethlis-

berger and Dickson, 1939). These studies, which lasted into the middle 1930s and which are described in Chapter 10, are considered by many, the present authors included, to be probably the most significant series of research studies ever carried out in industry in terms of the impact they had on the growth and development of industrial psychology. They provided the foundation and impetus for the expansion of industrial psychology beyond the realm of selection, placement, and working conditions into the study of motivation, morale, and human relations.

The Depression itself had considerable effect on the development of industrial psychology. While it may have slowed growth in some directions, it nevertheless opened many additional areas for survey. After the Depression the importance of employee attitudes began to be recognized; consequently much development since that time has been in this area.

Without much question World War II was a major factor in the growth of psychology in industry. Although the American Association for Applied Psychology was formed in 1937 as the official organization of industrial psychology (it later became Division 14 of the APA), it was the huge psychological contribution to the war effort which proved to so many people that applied psychology had important and practical contributions to offer. These contributions were far too numerous to document in any detail here. As in World War I, great emphasis was placed upon the development of tests for selecting and classifying recruits. Also developed were selection programs for officers, various training programs of specialized types, and job analysis and performance evaluation techniques.

Personnel Psychology, another major journal of applied research, first appeared in 1948. It publishes the results of factual psychological studies in such fields as training, job analysis, selection, evaluation, motivation and morale, work conditions, and equipment design. Its articles are intended for interested and informed management and yet at the same time they are designed to meet the technician's requirement of accurate and complete reporting.

Another notable development in applied psychology since World War II was the establishment of other separate divisions of the APA devoted to various aspects of the field: the Division of Military Psychology (Division 19); the Society of Engineering Psychologists, a division of the APA (Division 21); and the Division of Consumer Behavior (Division 23). While the percentage of joint memberships in these divisions and Division 14 is high (i.e., many psychologists belong to two or more divisions), the fact that they have been formed is ample testimony to the need of the psychologist to recognize the ever-increasing complexity and specialization of interest in the web of activities which define applied psychology.

Several other major organizations have been created since World War II to represent various interest subgroups in applied psychology. The two most notable are the Human Factors Society and the Ergonomics Society. The former is an American group for applied psychologists with interest in human engineering problems. It publishes its own journal, *Human Factors*. The latter group is the British counterpart of the Human Factors Society, although it was the first to be organized. Its journal is entitled *Ergonomics*. *Ergonomics* and *Occupational Psychology* are the two major British publications devoted to applied psychological research. Recently (in 1966) another American journal, *Organizational Behavior and Human Performance*, started publication. The purpose of this journal is to publish articles devoted to the development of theory in applied psychology.

MAJOR PROBLEMS OF INDUSTRIAL PSYCHOLOGY

Before proceeding to the methods and content of industrial psychology, it might be best to mention certain major problems which the profession has to face in its future growth and development.

THE CONSULTANT AND THE STAFF PSYCHOLOGIST

As was pointed out earlier, the industrial psychologist is likely to obtain his livelihood through one of three major sources of employment. He is either a consultant, an employee of a company or the government, or a university teacher. Very often he combines two of the three roles, but whether he does or not depends upon his interests, opportunities, degree of identification, and tempo.

A psychologist directly employed full time by a company or by a government agency is often referred to as a "staff" psychologist. Generally speaking, the duties and tasks of the consultant and the staff psychologist overlap. There is no clear-cut difference insofar as type of assignment is concerned. The major difference is that the consultant may be concurrently working for a number of clients or employers, whereas the staff psychologist fills a more specific role in the organization chart for a single employer.

Although a schism between the staff psychologist and the consultant is undesirable if the profession is to be advanced in industry, the answers given in Canter's study (1948) to the question "What do you think of consulting firms as the best solution to industrial psychological problems?" pose a serious future problem. One-half of the staff psychologist group was unfavorable toward such firms; the consulting group was generally favorable. This situation demands attention and should be cleared up.

A note of optimism is reflected toward the field in general since 80 percent of the respondents reported that executives were becoming more "psychological minded." A further indication of the increased acceptance of the psychologist by leaders in industry comes from a 1962 survey conducted by Feinberg and Lefkowitz (1962). They administered a questionnaire to 89 executives who were attending a seminar sponsored by the American Management Association. When asked whether they would be interested in hiring an industrial psychologist, over two-thirds replied favorably. These "yes" respondents felt the industrial psychologist could be of greatest benefit in the areas of employee motivation, employee selection and training, executive selection and training, human engineering, consumer research, production efficiency, and accident control.

COMMUNICATION

One of the difficulties of any profession is that its language and techniques sometimes become so involved that the outsider is really left out. If industrial psychology is to gain an important place in industry, psychologists must learn to talk and write in a fashion that is clearly understandable to others who are equally interested in the mutual problems and who sometimes have an even greater stake in a solution. Not only must the industrial psychologist learn to communicate adequately with the

nonpsychologist, but even the problem of communication within the field itself is becoming a problem. The ever-increasing complexity of industrial psychology and the specialization of interest of the psychologists working on different problems in different settings has created many barriers to the flow and dissemination of knowledge among researchers and practitioners. While such problems may be the inevitable corollary of a dynamic discipline, the authors feel that the communication problem is one of the most critical in industrial psychology today.

RESISTANCE TO CHANGE

Research findings as well as research itself can ordinarily be expected to meet with resistance on the part of employees and, in many instances, employers. The successful practitioner of industrial psychology must be immediately and forever aware of this phenomenon. It would be purely academic if one anticipated that industry is waiting with open arms to apply the knowledge of industrial psychology.

Attempts at change, no matter how well-intentioned, produce threats and will be resisted. This resistance may take the form of hostility and aggression against the change itself or against the administrator of the projected change. Often the employee imagines the nature of the change well in advance of the possibility of a change. The unreality of the imagination only makes the resistance stronger. When changes are associated with speedups or layoffs, the resistance to any contemplated change is even more intense. It is not enough to state that no detrimental action to the employee's welfare is contemplated. The claim must be proved. Anything that is not clearly understood can be an insecurity-producing factor. Change often upsets established pattern. People are not easily corrected, nor are they able to give up habits freely. Research often intends to change behavior that has become routine and thus can be expected to be resisted.

Resistance comes not only from the employee but from all levels of management and the employer. The naive employer often wants research to prove his point or position. Such a guarantee is not possible: The conclusions of research depend upon the data and cannot be established by manipulation of data to conform to a preestablished outcome.

All, however, is not hopeless provided at least four fundamentals are recognized. First, the reasons for the contemplated change should be clearly explained. Second, those who will be involved in the change should have ample opportunity for participation in the implementation of that change. Third, change should be a two-way affair rather than an attempt to force all to agree to a one-sided decision. Fourth, the administrator of the change should recognize at all times that change is a real, imagined, or potential threat, and that he must do all he can to eliminate or reduce the possible threat regardless of the form it assumes.

CHAPTERS TO FOLLOW

Each of the chapters in this book will discuss one segment of the various duties of the industrial psychologist. There are, of course, many sequences in which the chapter could be presented and many topics which could be excluded or included.

The material is presented, however, according to the way the authors perceive their subject matter. It will be well to remember that each chapter is related closely to all the other chapters, and that chapter divisions are merely educationally desirable tools—they are not the realities. The individual in a work situation, be he employer or employee, has attitudes, satisfactions, motives, and abilities, and many varieties of feelings and knowledge are always in force at the same time.

REFERENCES

American Institutes for Research in the Behavioral Sciences. *Annual report.* Pittsburgh, 1964.

American Psychological Association. *Directory.* Washington, D C , 1966.

American Psychological Association. Psychology and its relation with other professions. Washington, D.C , 1954.

American Psychological Association, Division of Industrial Psychology. *The psychologist in industry* Washington, D.C , 1959.

Bills, M. A. A day in the life of an industrial psychologist. *Personnel Service Bulletin,* 1934, no. 5.

Canter, R. R. Psychologists in industry. *Personnel Psychology,* 1948, *1,* 145–161.

Dunlap and Associates, Inc. *Annual report.* Stamford, Conn., 1964.

Dunnette, M. D. Incomes of industrial psychologists. *American Psychologist,* 1961, *16,* 533–536.

Educational Testing Service. *Annual report.* Princeton, N.J., 1962–1963, 1965–1966.

Feinberg, M. R , and J. Lefkowitz. Image of industrial psychology among corporation executives. *American Psychologist,* 1962, *17,* 109–111.

Ferguson, L W *The heritage of industrial psychology.* Finlay Press, Hartford Conn , 1962.

Flory, C. D , and E T. Janney. Psychological services to business leaders *Journal of Consulting Psychology,* 1946, *10,* 115–119.

McCollum, I. N Psychologists in industry in the United States. *American Psychologist,* 1959, *14,* 704–708.

McCormick, E. J. Role of the psychologist in industry. *Journal of Personnel and Administrative Industrial Relations,* 1955, *2,* 23–31.

Munsterberg, H. *Psychology and industrial efficiency.* Houghton Mifflin, Boston, 1913.

Roethlisberger, F. J., and W. J. Dickson. *Management and the worker.* Harvard University Press, Cambridge, Mass., 1939.

Ross, S., and R. F. Lockman. A career in psychology. American Psychological Association, Washington, D.C., 1963.

Sawyer, J. The industrial psychologist education and employment. *American Psychologist,* 1960, *15,* 670–673.

Scott, W D. *The theory of advertising* Small, Maynard and Company, Boston, 1903.

Scott, W. D. *The psychology of advertising.* Small, Maynard and Company, Boston, 1908(a).

Scott, W. D. *The psychology of advertising in theory and practice.* Small, Maynard and Company, Boston, 1908(b).

Scott, W. D. *Increasing human efficiency in business.* Macmillan, New York, 1911(a).

Scott, W. D. *Influencing men in business.* Ronald, New York, 1911(b).

Seashore, S. E. A description and analysis of institutional context for social research. Paper presented at the meeting of the American Sociological Association, Washington, D.C., August, 1962.

Stagner, R. Some problems in contemporary industrial psychology. *Bulletin, Menninger Clinic,* 1957, *21,* 238–247.

Taft, R. The staff psychologist in industry. *American Psychologist,* 1964, *1,* 55–61.

2 A STATISTICAL RATIONALE FOR SELECTION AND PLACEMENT

The most traditional activity of the psychologist in industry has been employee selection and placement. The challenging task of trying to distinguish "good" workers from "poor" workers with the aid of tests and other selection devices may be said to coincide with the early formulations of industrial psychology. These problems attracted such pioneer industrial psychologists as Hugo Münsterberg, Walter Dill Scott, and Walter V. Bingham into the world of industry in the early 1900s. Münsterberg used psychological tests prior to 1910 in connection with various selection problems for the Boston Elevated Railway Company. Scott and Bingham worked on the problem of selecting salesmen in the period 1915–1917 while they were colleagues at the Carnegie Institute of Technology.

MANPOWER PLANNING

People represent a most important resource to a business or industrial firm. Thus the terms *selection* and *placement* designate separate phases in the ever-important area of manpower planning. The trends toward automation and computerization are making the selection and placement of people more important rather than less important. Despite any or all technological advances, profits resulting from a company's efficient operation require the expeditious use of manpower via correct selection and placement.

Selection, as the name implies, involves picking for hire a subset of workers from the total set (population) of workers available for hire at any given moment in time. Efficient selection is therefore a nonrandom process, insofar as those selected have been chosen on the assumption that they are more apt to make "better" employees than those who have been rejected. The task of the industrial psychologist is to make certain that the assumption is indeed a valid one as a result of using objective and scientific procedures and instruments rather than subjective and biased judgments. An equally critical problem is that of assigning the new hires to the jobs available. Placement, in its simplest form, asks, "given the hiring of N men to fill N different job vacancies, which man should be put on which job?" As might

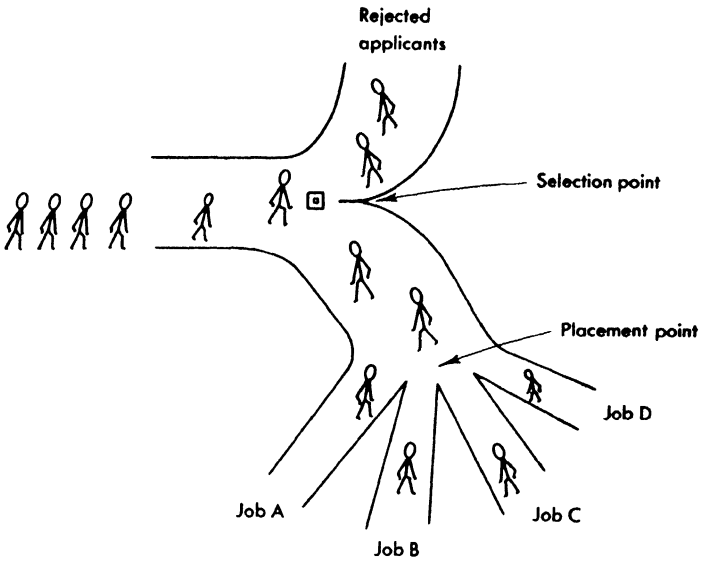


Figure 2.1. Highly simplified diagram of the manpower allocation process.

be suspected, in actual practice the processes of selecting and placing employees often merge into a single process. Figure 2.1 shows in highly simplified form the general process of manpower allocation.

INDIVIDUAL DIFFERENCES

That people differ is self-evident. Measuring any characteristic of people with a measuring instrument will result in a *distribution* of scores—some people will have high scores, others low scores, etc. As an area of psychology, the study of individual differences has provided the impetus for the entire psychological testing movement. Psychological tests measure the extent of differences among people and have been used to obtain estimates for a wide variety of human characteristics.

The existence of individual differences provides the basis for the selection and placement process. Any group of workers, irrespective of occupation, will vary considerably in terms of their relative work efficiency and performance. For example, many years ago C. L. Hull reported some ratios of efficiency between the most efficient and the least efficient workers in a variety of different occupations (Table 2.1). Note that in the case of one occupation (spoon polishers) the best worker was over five times as efficient as the poorest worker.

It is not surprising that employers show considerable interest in wanting to develop ways of selecting those workers who will be more efficient and thus reducing the possibility of hiring those who may be relatively inefficient. This is particularly true for those employers who go to considerable expense training new employees before determining whether they will be successful; and it is very important when union contracts prevent terminating employees for any reason after a brief trial period (often no longer than six weeks).

Another example of how widely workers at the same plant and even at approximately the same job level can vary on a single trait is shown in Table 2.2. The

TABLE 2.1 *Ratio of Least Efficient to Most Efficient Individual in a Variety of Gainful Occupations*

Occupation	*Criterion*	*Ratio of Poorest to Best Worker*
Heel trimmers (shoes)	Number of hours per day	1:1.4
Loom operators (silk)	Percent of time loom is kept in operation	1:1.5
Hosiery maters	Hourly piecework earnings	1:1.9
Loom operators (fancy cotton)	Earnings	1:2
Knitting machine operators	Pounds of woman's hose produced per hour	1:2.2
Elementary school teachers	Ratings of superiors	1:2.5
Spoon polishers	Time per 36 spoons	1:5.1

SOURCE: C. L. Hull. *Aptitude testing*. Harcourt, Brace & World, New York, 1928.

table presents the scores of nearly 2000 workers in single plant on the *Wonderlic Personnel Test*, a short paper-and-pencil test of general mental alertness or intelligence developed specifically for use in industry. For a person unfamiliar with the diversity of talent likely to be found at any job in any industry the range of scores may seem hard to believe—the ratio of ability in terms of IQ equivalent is nearly 2:1 between the top and bottom worker.

The same data are shown plotted graphically in the form of a frequency polygon in Figure 2.2. This illustrates that the distribution of many characteristics of people,

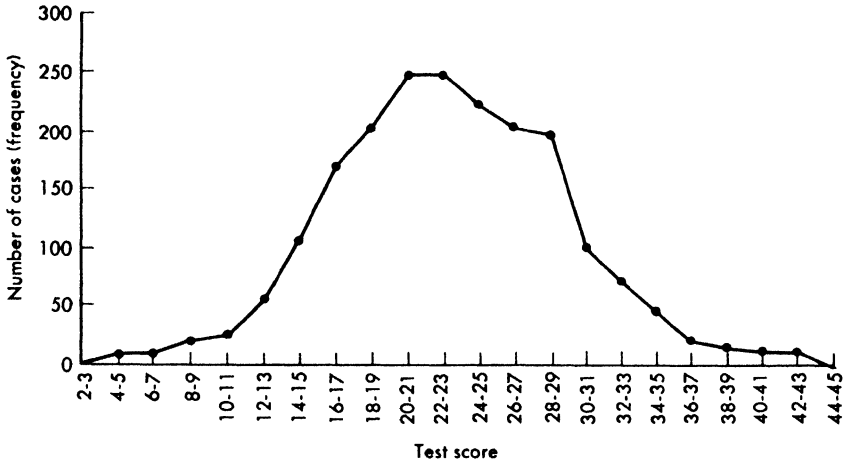


Figure 2.2. Frequency distribution showing the number of employees obtaining various scores on the Wonderlic test in a manufacturing plant. (Adapted from W. J. McNamara. Retraining of industrial personnel. *Personnel Psychology*, 1963, *16*, 233–247.)

TABLE 2.2 *Distribution of Scores on Wonderlic Test for a Random Sample of 1940 Employees in a Manufacturing Plant*

Raw Scores	*IQ Equivalent*	*Percentile*	*Number of Cases*
42–43	131	100.0	4
40–41	129	99.8	4
38–39	127	99.6	8
36–37	124	99.2	15
34–35	123	98.4	43
32–33	119	96.2	69
30–31	116	92.6	100
28–29	113	87.5	197
26–27	110	77.3	203
24–25	107	66.8	227
22–23	104	55.2	249
20–21	101	42.3	246
18–19	97	29.6	194
16–17	93	19.6	164
14–15	89	11.2	105
12–13	85	5.8	57
10–11	80	2.8	24
8–9	76	1.6	21
6–7	72	0.5	7
4–5	68	0.2	3

SOURCE: W. J. McNamara. Retraining of industrial personnel. *Personnel Psychology*, 1963, *16*, 233–247.

whether the characteristics be height, weight, or scores on a test, assumes a shape which approximates quite closely a bell-shaped curve or distribution. This is known as the *normal curve*. Because of this isomorphic relationship between real-world data and a mathematical equation that describes a theoretical curve, the theoretical curve can be used as a model for these data. Thus one might speak of test performance for factory workers as being, in general, *normally distributed*, implying that the general shape of the distribution of scores approximates the normal curve. Many statistical techniques so important to the industrial psychologist depend upon the principle of normally distributed scores.

Of course, not all frequency distributions are as symmetric as in Figure 2.2. Sometimes, conditions affecting job performance can cause distributions to change shape radically. For example, consider the curves shown in Figure 2.3. The first distribution represents the "true" frequency distribution one might find on a job if everyone performed at or near his normal capacity. The other two distributions illustrate how the shape can be influenced by such factors as work restriction or unlimited incentive. The middle curve has a shape that is *negatively skewed* while the bottom curve is *positively skewed*. Skewed distributions of test results do occur. For example, very difficult tests usually result in positive skewness and very easy tests result in negative skewness.

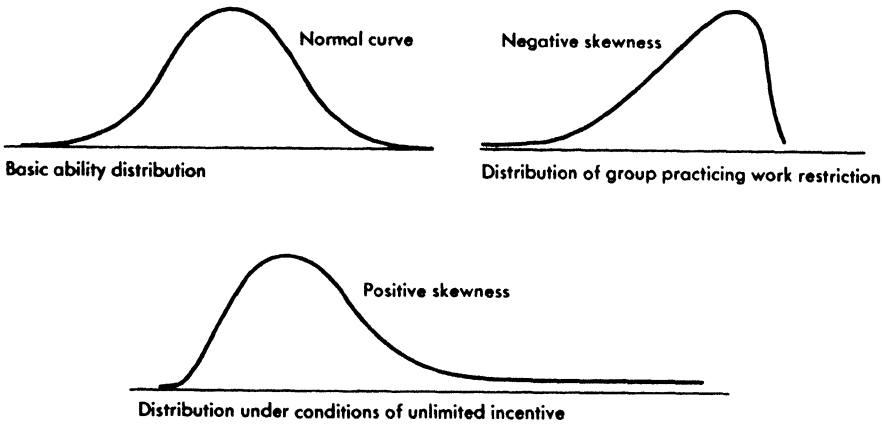


Figure 2.3. Distribution of scores of actual performance on a job may differ as a function of specific job conditions independent of the basic ability distribution.

BASIC SELECTION MODEL

The theory of testing in industry is based upon the individual differences among workers. Indeed, the purpose of the selection process is to take full advantage of such differences in order to select primarily those persons who possess the greatest amount of attributes important for success on the job.

Figure 2.4 diagrams the basic selection model. Each of the steps shown is important in developing a proper selection procedure.

Step 1: Examination of the job(s) having vacancies. This step consists of job analysis and will be discussed in greater detail in Chapter 17. Job analysis is a most critical and basic problem area in industrial psychology and is, or should be, the foundation of any industrial testing. A thorough knowledge and understanding of a job is of paramount importance and must precede the use of any test in the selection and placement of workers.

Step 2: Selection of criterion and predictor. The second step involves two parts—choosing an indicator which measures the extent of how "good" or successful a worker is (typically referred to as the *criterion*) and choosing a particular measure that can be used to predict how successful a worker will be on the job (typically referred to as the *selection device* or *predictor*). Criterion selection is an involved but basic problem; the solution decides how job success is defined and/or measured. Chapters 6 and 7 discuss some of these problems.

On the predictor side, the psychologist finds a wide variety of potentially useful devices which can be successful in discriminating between "good" and "poor" workers. Often used are such predictors as tests, interviews, application blanks, and letters of recommendation, among others. Chapters 4 and 5 present a more thorough look at various kinds of predictors and their relative merits.

Step 3: Measurement of performance. Once the criterion and the predictor have been selected it is necessary to obtain measures on both from a sample of workers on the job. This can be done either by giving the predictor to present

employees and simultaneously obtaining criterion measures, or by giving the predictor to new hires and waiting a specified time before obtaining the criterion measures (thus allowing enough time for the new hires to establish themselves as successful or unsuccessful). Both methods are used, and each has its advantages and disadvantages. These will be covered in more detail when we discuss kinds of validity.

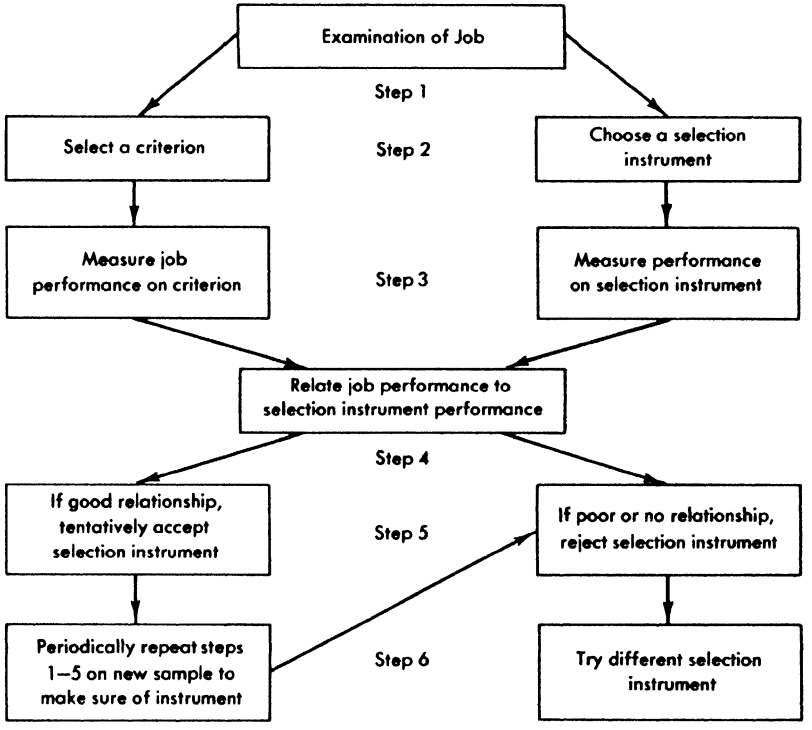


Figure 2.4. Basic selection model.

Step 4: Relating predictor to criterion. The fourth step in the selection process involves determining whether a true and meaningful relationship exists between the employee scores on the predictor and the criterion. Only if such a relationship exists can the selection process be considered successful. Establishing the existence of such a relationship is called *assessing the validity* of a predictor. This is usually a statistical process which involves the use and understanding of correlational methods and significance tests.

Step 5: Deciding upon the utility of the selection device. Making the final decision as to whether to use the predictor to select new job hires depends not only upon the size of the relationship found (in step 4) and its significance, but also upon many other conditions: the number of applicants, number of job openings, proportion of present employees considered successful (the *base rate*), and respective variances of the successful and unsuccessful worker groups. All of these additional aspects concerning the use of a predictor are discussed in later portions of this chapter.

Step 6: Reevaluation. The fact that the predictive situation is a dynamic, ever-changing one should never be forgotten. What makes for good selection today may not be at all appropriate tomorrow; applicants change, jobs change, and employment conditions change. Thus any good selection program should be reevaluated periodically to make certain it is doing the job for which it has been designed.

CORRELATION

Prediction of job success involves determining the extent to which the predictor is related to the criterion. For example, suppose one were interested in setting up a selection program to hire new file clerks. Suppose further that it had been decided to use a paper-and-pencil test of clerical aptitude as a potential predictor of file clerk efficiency, and that efficiency was to be determined by ratings of supervisors. Table 2.3 shows some hypothetical data for this assumed situation, giving scores for twelve file clerks on both the clerical test and the efficiency criterion measure. Figure 2.5 shows a graph of the data in Table 2.3.

TABLE 2.3 *Scores on Clerical Test and Job Proficiency Criterion for 12 File Clerks*

Clerk	*Clerical Test*	*Job Proficiency*
1	10	3
2	60	6
3	20	2
4	50	7
5	80	9
6	10	2
7	20	4
8	50	5
9	60	10
10	40	5
11	70	8
12	30	4

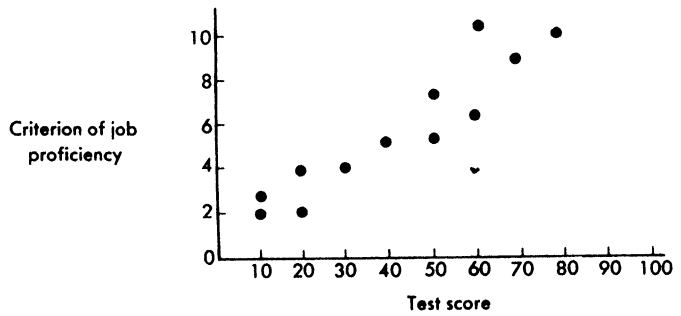


Figure 2.5. Plot of data in Table 2.3.

Notice that there appears to be a systematic trend. In general, the higher a person scored on the clerical test, the higher she scored on the measure of job proficiency. We can therefore deduce that there is a definite *relationship* between test performance (the predictor) and job proficiency (the criterion). We can also deduce that if we select those people who score higher on the test, we are more apt to hire people who will be more proficient than if we hire people independently of test score.

ESTABLISHING THE DEGREE OF RELATIONSHIP

The degree of relationship between any two variables may be defined as the extent to which these two variables vary together in a systematic fashion. The more technical term for this is the degree of *covariance* existing between variables. A formal measure of the degree of covariance between any two sets of scores is provided by a statistic known as the coefficient of correlation. When two sets of scores are highly related, we say they are highly *correlated*. The most common measure of correlation is the Pearson Product Moment Correlation Coefficient which is designated by the symbol r.

As a measure of relationship, r varies between $+1.00$ and -1.00. When r is $+1.00$, the two sets of scores are *positively* and perfectly related to each other. When r is -1.00, the two sets of scores are *negatively* and perfectly related to each other. When $r = 0.00$, the two sets of scores have no relationship to each other at all. Figure 2.6 shows graphs of different magnitudes of r.

In predicting job success the *sign* of the correlation coefficient is not important, but the magnitude is. The larger the absolute size of r, the better the prediction of criterion scores on the basis of information obtained from the predictor.

To understand the rationale of correlation it may be helpful to consider a pictorial representation of covariance and its relationship to r. Any set of scores will possess some amount of variation—in fact, as we have already seen, the scores of people on many traits follow a normal distribution with a small number of very



high scores, a small number of very low scores, and most of the scores occurring in the middle of the distribution. Suppose we represent this variance in a set of criterion scores as shown above, where the total area is defined as being 1.00. We can do this since it is possible to transform any set of raw scores so that their variance becomes equal to 1.00 using what is known as a z score transformation.

Similarly, suppose we have a set of predictor scores which also vary and are normally distributed, and again the area is defined as being equal to the quantity 1.00.

We can now represent r geometrically as being related to the amount of overlap (covariance) of the two sets of scores. A more precise definition of r as a statistic

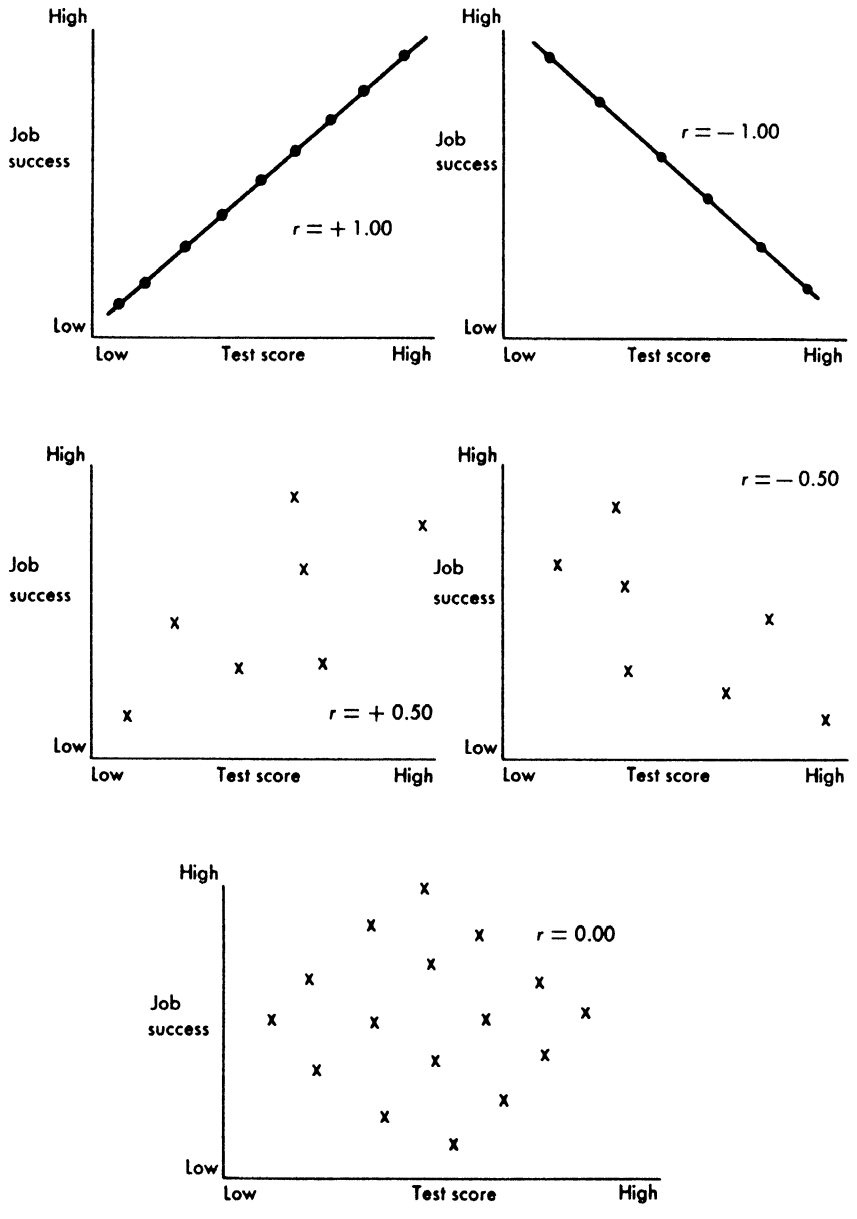
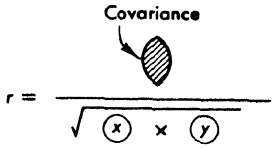
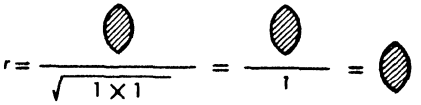


Figure 2.6. Correlation plots for various degrees of relationship between test score and job success.

is that it is the ratio of the amount of covariance between two variables to the square root of the product of the respective variances (sometimes called a geometric mean) which can be diagramed as shown below:



But when both (x) and (y) are equal to 1.00, we have r being equal simply to the amount of overlap between the two variables, as



Returning to the data given in Table 2.3, it is possible to compute the correlation between these two sets of scores using the formula

$$r = \frac{N\Sigma xy - (\Sigma x)\,(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2]\,[N\Sigma y^2 - (\Sigma y)^2]}}$$

The reader is advised that r cannot be interpreted as a percentage. If $r = 0.50$, this does not imply that 50 percent of the variance in the criterion is predictable from the selection variable. The square of r, however, can be so interpreted. A correlation of 0.50, when squared, gives $r^2 = 0.25$, which may be interpreted as the percent of variance in the criterion predicted by the selection variable.

The statistic r^2 is sometimes called the *coefficient of determination* because it

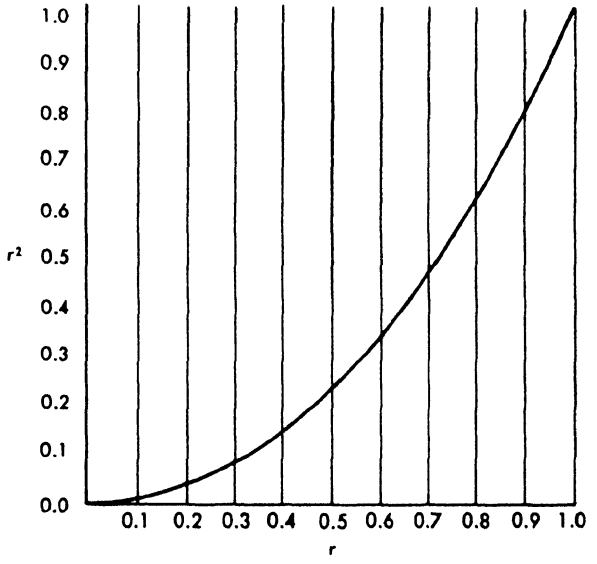


Figure 2.7. Relationship between r and r^2.

represents the amount of variance in one variable which can be "determined" by knowing the scores on a second variable. Figure 2.7 shows the relationship between r (the measure of relationship) and r^2. Note that it is possible to obtain r's of rather substantial size and still account for only a small proportion of the criterion variance.

REGRESSION

As we have seen, the correlation coefficient r measures the degree of relationship between two variables. By itself, however, it does not provide us with a procedure by which we can *predict* one set of scores from another set. The technique by which this is done is called *regression analysis*. Regression may be thought of as being related to correlation as follows: *Correlation* measures the magnitude or degree of relationship between two variables, while *regression* gives a description of the type of relationship between variables which in turn can be used to make predictions.

To illustrate regression, consider the scores plotted in Figure 2.8a. Obviously there is a substantial positive relationship existing between the predictor and the criterion in this case. Unfortunately, Figure 2.8a does not provide us with any information about the exact relationship other than the fact that it is a linear one (r always measures only the degree of linear, as opposed to curvilinear relationship, between two variables). If we want to predict criterion scores from some selection device, it is clear that we need to describe the observed relationship between predictor and criterion more specifically. This is accomplished by finding the line or function that best describes the data points. This is called fitting a "line of best fit" to the data. Since we are assuming the relationship to be linear (we used r to measure its magnitude), the type of line we use must be straight, that is, no curved lines are permitted. This best-fitting straight line is called the *regression line* and can be used to predict the criterion from the predictor.

Figure 2.8b shows two different lines of best fit which might be obtained if we asked two different persons to examine the data and then draw a line through the points which in their opinion seems best to describe the trend or relationship between the variables. While the general trend is similar, we find that the two people do not completely agree in their estimate of the relationship. This disagreement would in turn result in disagreement in the *predicted* criterion score depending upon which estimated regression line was used. Given a job applicant with a score x on the selection instrument, we would predict a criterion score of y_1 for this applicant if we were to use the first person's regression line; if we used the second person's regression line we would predict y_2 as the most likely criterion score. Which regression is correct?

This is a difficult question to answer unless there is some basis for deciding what a "best fit" really is. Fortunately, statisticians have generally agreed that a best-fitting line is one which goes through the points so that it minimizes the sum of the squared distances (in the y dimension) of the points from the line as shown in Figure 2.9.

A line which accomplishes minimizing Σd_i^2 is called a "least-squares" regression line. Such regression lines are mathematically directly related to r. Using the least-

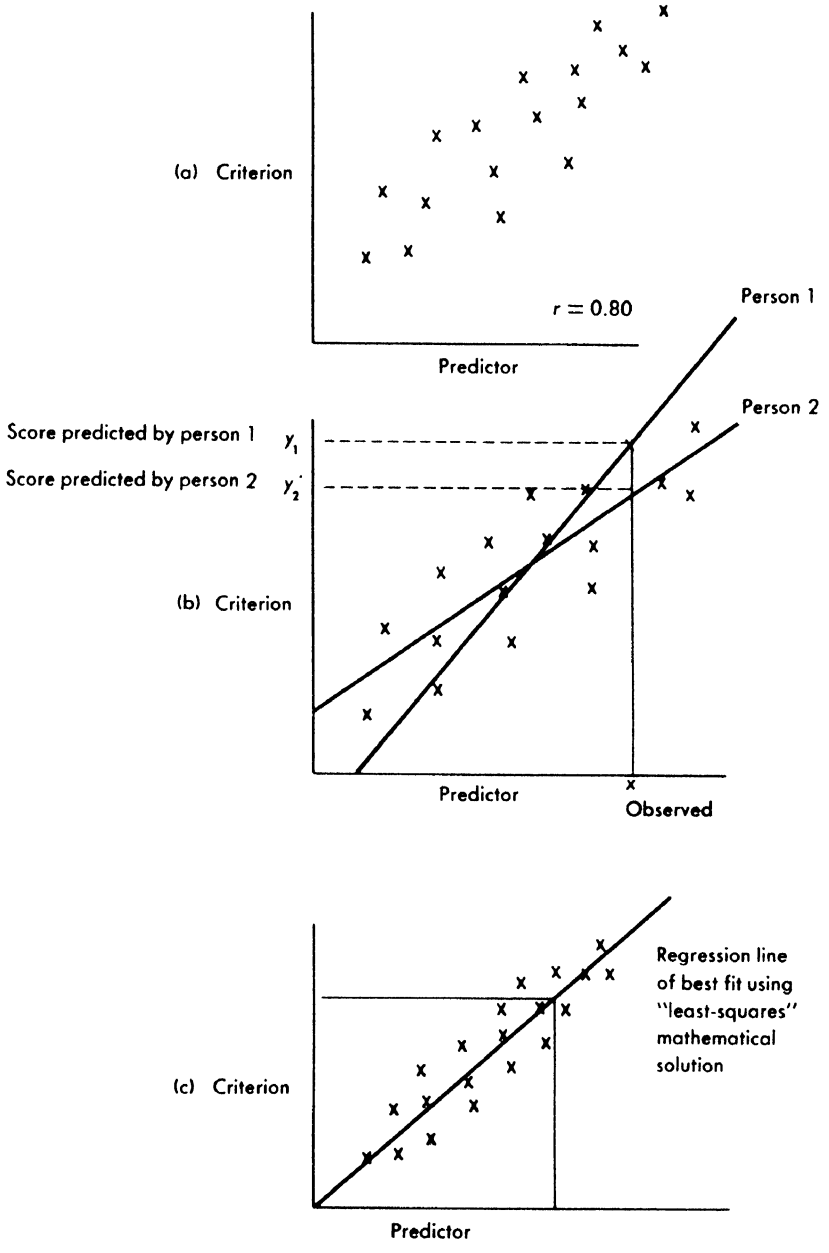


Figure 2.8. Finding a "line of best fit" to a set of points.

squares method for obtaining our prediction line will assure that different people will end up with the same line (assuming they make no errors in calculation). Similarly, the predicted criterion score for any particular x value will not vary depending upon who fits the prediction line (see Figure 2.8c).

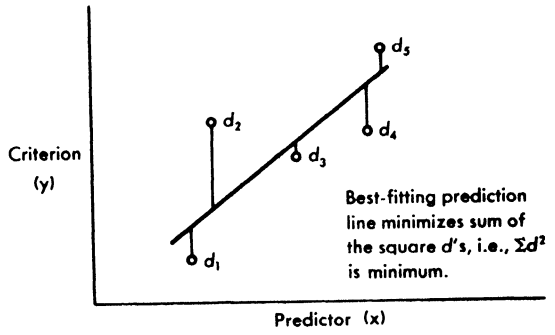


Figure 2.9. Example of "errors" in prediction for a line of best fit.

At this point the reader may ask, "Why do we need to predict criterion scores when we already have them?" The answer is quite simple. The initial measurement of the extent of the relationship between the predictor and criterion obviously requires both sets of scores or else the relationship could not have been established. Should the selection device prove useful, it can then be used with all new applicants for whom there can be a predictor score but for whom a criterion score does not exist. Our objective is to predict the criterion performance of future applicants. If a new applicant scores high on a test that was found to have a high positive relationship to the criterion, then we should expect him to have a high probability of turning out to be a successful hire.

STATISTICAL SIGNIFICANCE

Selection research always requires a knowledge and application of the basic concept of *statistical significance*. This basic concept allows us to state, with a specified likelihood of being correct, whether our findings represent the "true" state of affairs or whether they are simply an artifact of the general laws of chance. Expressed another way, tests of significance provide the researcher with a means of determining if the correlation coefficient is simply a chance phenomenon or if it adequately reflects the "true" correlation coefficient.

If a test of statistical significance indicates the obtained correlation is, in all likelihood, not due to chance sampling factors, then one may proceed to use the selection device with a certain degree of confidence that it will indeed perform successfully the task of discriminating good from poor performers on the criterion measure. On the other hand, if the test of significance indicates that the relationship between predictor and criterion is probably due to chance factors, the psychologist certainly has no justification for using the selection device to decide the fate of future job applicants—a fact that is sometimes tragically ignored by either

naive or unscrupulous individuals working in the area of selection and placement in industry.

It is important to understand that tests of significance are essentially indicators of chance probability. For example, *how sure must one be that results are not due to chance?* Generally speaking, most researchers accept with confidence results that could only be due to chance 5 times or less out of 100. They assume the risk of being wrong in their decision 1 time out of 20 by this process. This is known as adopting a "5 percent level of significance" in one's research. More cautious researchers will adopt even more stringent significant levels such as the 1 percent level. This means that they will only accept as nonchance those outcomes which have a chance likelihood of 1 time or less out of 100 cases. It is standard practice for all researchers to specify the level of significance which they have adopted for their research, as we will see in later chapters.

CHARACTERISTICS OF PREDICTORS

The two paramount requirements for any predictor are *validity* and *reliability.* In the industrial setting different kinds or types of validity exist, although the one that is most preferred is called *predictive validity.*

There are also different types of reliability measures. The concern with reliability and validity is not limited to predictors but applies to criteria as well—however, the discussion of the criterion is deferred to Chapter 6.

VALIDITY

The validity of a predictor can generally be defined as *the extent to which the predictor achieves certain aims of the user* by measuring what is supposed to be measured. The specific kind of validity involved thus depends upon the particular aim of the user in any situation.

PREDICTIVE VALIDITY The aim of the user is to employ his measuring instrument to predict the future performance of employees on some other variable (criterion). Predictive validity is established statistically through correlation and regression. The important distinction of predictive validity is one of a time element. Predictor scores are obtained on individuals at one point in time (e.g., time of hire) and criterion scores are obtained at a later date (e.g., at the end of six months). The resulting relationship thus truly represents the "predictive" power of the instrument. Predictive validity is the most important type of validity in selection since it is the only type which truly duplicates the selection situation. Another name which is sometimes used for predictive validity is *follow-up validity.*

CONCURRENT VALIDITY The aim here, at least in theory, should be to estimate the *present* performance of employees on some criterion measure from scores on the predictor. Concurrent validity is also established using correlation and regression techniques, but with no time lag between the obtaining of predictor and criterion scores. A sample of current employees is used to ascertain the predictor-criterion relationship, and then the resulting regression may be applied by obtaining predictor scores on the remaining job holders. In other words, we are interested

in predicting the *present* status of people, not their status at some future time. It is extremely important to point out that *high concurrent validity does not assure high predictive validity*. Unfortunately, concurrent validity is too often used in industry as a substitute for predictive validity. Management is sometimes unwilling to wait for the time required by the predictive method, and may not realize that present employees can represent a basically different population of workers from job applicants. Workers who are presently employed have survived screening in both hiring and continuity, and the poorer workers who were hired may have left either voluntarily or by request. This makes it very difficult to justify generalizing concurrent validities to a predictive validity situation.

CONTENT VALIDITY When the validator assumes that his predictor is representative of a given class of situations, he is involved in content validity. He has a specific notion about the kind of knowledge, skill, attitude, or performance that should be tapped by the measuring instrument, and he considers the instrument valid to the degree that its content is representative of what he wants to tap. Content validity is generally not measurable in any statistical or quantitative sense. One finds the greatest use of content validity among users of achievement tests, such as final exams in a college course. A final exam could only be considered to have content validity if it adequately represented (sampled), in terms of its items, the content of the course. If it did not represent a coverage of course material, it certainly could not be considered an appropriate test to use for a final exam—i.e., it would not have content validity.

CONSTRUCT VALIDITY With this type of validity the user wishes to infer the degree to which the persons being evaluated possess some trait or quality (construct) presumed to be reflected in test performance. The general procedure involves administering several test instruments which logically appear to measure the same construct, and then observing the relationships between these measures. Construct validity has not been used to any considerable degree by the industrial psychologist; it tends to be more often used in theoretical rather than pragmatic situations.

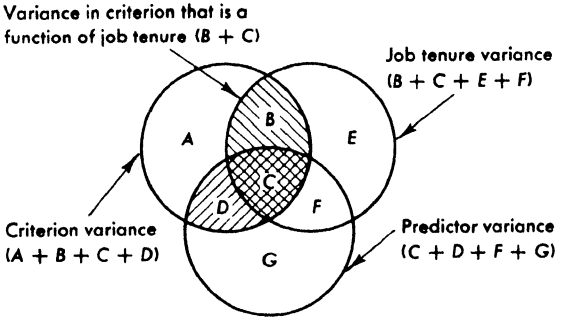
SYNTHETIC VALIDITY One might consider synthetic validity to be "assumed" predictive validity. Suppose we have a test that in a number of situations has demonstrated a high predictive relationship to various performance criteria of industrial foremen. Suppose further that a small manufacturing plant wants to use a test in selecting foremen, but too few foremen work in the plant to carry out even a concurrent validity study. This plant might decide to use the test without any formal statistical evaluation on the assumption that it had been demonstrated as being successful in other, larger plants. This procedure could only be considered valid if (1) the foreman's job in this plant is similar to the foremen jobs involved in the statistical evaluation of the test, and (2) the foreman applicants at this plant are typical (come from the same population) as the applicants for the foremen jobs in the larger plants. Synthetic validity should only be substituted for predictive validity with the full awareness of its possible limitations.

FACE VALIDITY Another kind of validity often used to describe a test involves the degree to which a user is interested in having his test "look right" to the test taker. Job applicants often become upset if the prediction instruments they are

required to take appear to have little or no relationship to the job for which they are applying. For example, if one is selecting people for a machinist position and a test of arithmetic ability is used as a predictor, the test items should deal with numbers applied to mechanical problems rather than being worded in more general terms such as the buying of apples or oranges. If the applicant fails to see the relevance of the predictor to the job for which he is applying, as often happens on personality tests, he may undergo a serious loss in motivation in the test situation, become derisive, or, on the other hand, feel insecure. This not only damages the selection program but also may harm the image of the company and damage the image of tests in any industrial setting. The authors would hazard a guess that some of the bad publicity received by users of selection devices in industry can be due to the user overlooking the need for his tests to have face validity.

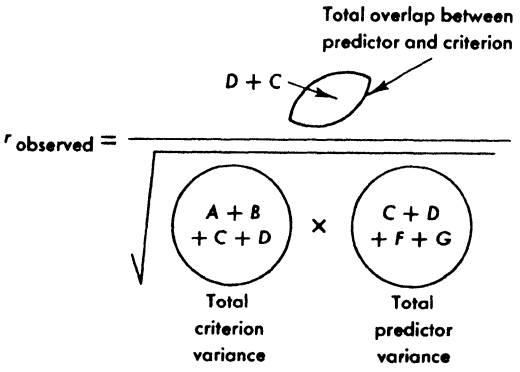
AGE, TENURE, AND JOB EXPERIENCE AND THEIR EFFECT ON VALIDITY Examination of the performance of workers on any particular job often shows a definite relationship between such variables as age and experience and the criterion. The more complex the job, the more these kinds of relationships are likely to exist. For many jobs a substantial amount of experience is necessary before employees become proficient at their work. The correlation between these kinds of variables and criteria of job success presents a serious problem in selection. Caution is necessary, particularly if one uses the concurrent validity procedure as a means of establishing the utility of any prediction device.

If, for example, there exists a high correlation between the criterion and length of time on the job, how should a high concurrent validity coefficient be interpreted? Does this mean that the predictor is truly reflecting ability differences among workers as measured by the criterion, or are worker differences primarily due to experience on the job? If it is the latter, then all the predictor is going to accomplish is to differentiate those workers with long tenure from those who have been hired more recently. The observed validity is generally an overestimate of the predictive efficiency of the selection instrument. In fact, unless one can clearly demonstrate that the predictor is *not* correlated with traits such as age and tenure which may themselves be determiners of job performance, all concurrent validities obtained with that predictor must be highly suspect. To illustrate the point, consider the situation where one has a criterion, a predictor, and a criterion-related variable such as job tenure which is largely responsible for the differences in skill shown on the criterion by employees, as follows:

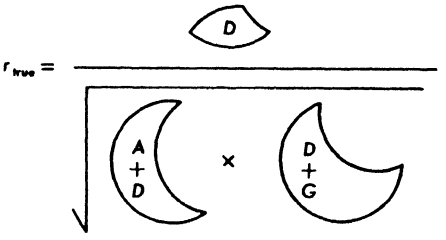


C + D = Observed concurrent validity of predictor
D = Amount of "tenure-free" criterion variance accounted for by predictor
C = Amount of "tenure-determined" criterion variance accounted for by predictor

The observed validity is generally but not always an overestimate of the true validity, since:



The true or unbiased concurrent validity, which represents the correlation between predictor and criterion that is completely free of the influence of job tenure, is given by the equation—



The correlation (r_{true}) as shown in the diagram actually represents, in pictorial fashion, what is known in statistics as a "partial" correlation coefficient. It reports the correlation between predictor and criterion after the effects of job tenure have been removed from both the predictor scores and the criterion scores of present employees. It is important that tenure effects be removed from both criterion and predictor in the concurrent situation. If these effects are not statistically removed from the *criterion*, we will end up predicting the influence of tenure rather than job performance, with little or no relevance to predictive validity. If job tenure effects are not removed from the *predictor*, we may also obtain a validity coefficient which cannot be considered relevant to any truly predictive validity situation.

Certainly the problems of criterion and predictor correlated variables in the concurrent setting illustrate some of the serious constraints involved in this method of validation. It can safely be stated that there is absolutely no equal substitute for the type of validity known as predictive validity when constructing and using a selection instrument.

RELIABILITY

In general terms, the validity concept deals with what is being measured by a measuring device. A second and perhaps equally important characteristic of predictors is the need to know the consistency of the measure, regardless of what is being measured. Stated another way, we need to establish the degree of stability of any measuring device; the measurement obtained from a predictor must be consistent. The degree to which any measuring instrument is consistent or stable and would yield the same scores again and again if necessary is defined as the *reliability* of that test instrument.

Like validity, reliability is usually measured by means of the correlation coefficient. Since reliable measurement implies stability from one situation to another, reliable instrument should produce either the same scores or at least similar rankings of individuals in two situations. By computing the correlation we obtain a mathematical expression of the extent to which that occurs. Thus a reliable measuring instrument is one on which individuals receive the same score (or nearly the same) in repeated measurements. When the correlation coefficient is used to measure the similarity of scores for a group of people on two applications of the same measure, it is called a *reliability* coefficient.

The actual process by which one can assess the reliability of a measure depends upon numerous factors. There are three major alternative "kinds" of reliability, each of which has its separate advantages and disadvantages. They are sufficiently different in their underlying logic to justify examining each in some detail. The three techniques for obtaining instrument reliability are: (1) repeated measures on the same people with the same test or instrument, (2) measurement on the same people with two "equivalent" forms of the measuring instrument, and (3) separation of the measuring device into two or more equivalent parts and intercorrelating these "part" scores.

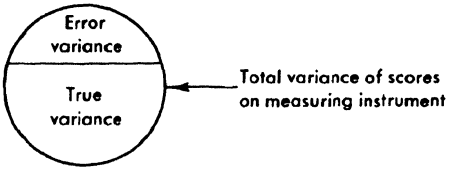
Before considering each method, we should examine in a more specific manner certain kinds of reliability or stability of measurement that we might conceivably be interested in under different circumstances. Let us assume that any time we use a measuring instrument to obtain a person's score, the score received is a function of several factors, as follows:

$$x_i = x_{\text{true}} + x_{\text{error}}$$

where x_i = observed score for person i on test
x_{true} = true score for person i on test—this is the actual amount of quality measured by the test that person i really possesses
x_{error} = error score for person i on test—this is the amount that person i's score was affected by the operating of various chance or time factors.

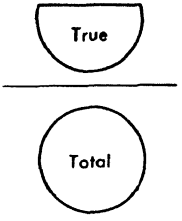
If all measuring instruments and measuring methods were "error-free," then we would always obtain the true scores of people, and the correlation between two measurements on the same group of people would always be + 1.00 or perfect reliability (assuming no change in the true scores is to be expected). Unfortunately, such error-free measurement is never completely available, since a wide variety of things contribute to performance at any particular moment in time. Thus, x_i may

either be greater than or less than x_{true} for any particular measurement, and correlations computed between measurements are always less than unity. In terms of our pictorial representation of the variance of performance among people on any measuring device, whether it be test or interview, predictor or criterion, this total variance may be divided into the two major components of true variance and error variance:

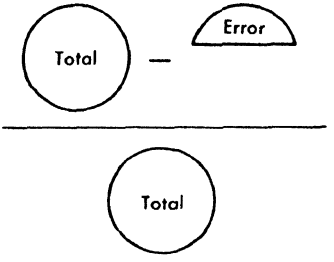


where total variance = total variability of observed test scores
true variance = variability of people in terms of their true amounts of the characteristic being measured
error variance = variability of people's error scores

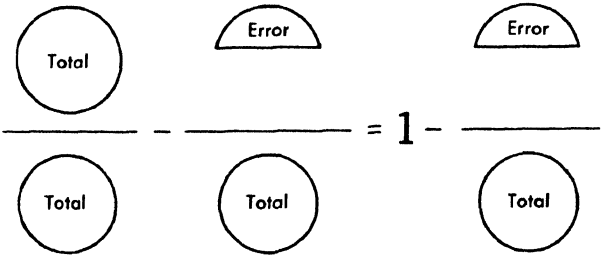
Reliability may be defined as a ratio of true variance to total variance, or



thus,



which can be expressed as



The larger the proportion of true score variance, or conversely, the smaller the amount of error variance present in the measuring process, the greater the reliability of measurement.

The critical factor differentiating the three major procedures for determining reliability is in the process of deciding what is to be considered error variance and what is to be considered true or systematic variance. There is no single reliability for any test. Rather, the reliability will depend upon the needs of the moment. For example, the psychologist might ask any of the following kinds of questions about the measurement process:

1. How accurately can I measure people with this test at any given moment in time?
2. How accurately will measures taken with this test today be representative of these same people at some future point in time?
3. How accurately will the scores on this test represent the true ability of these people on the trait being sampled by the test?

All three are legitimate reliability questions. However, each places a somewhat different emphasis on various sources of error variation in test scores. These sources of error variation have been expressed by Thorndike and Hagen (1963) as:

1. Variation due to the test at a particular moment in time
2. Variation in the individual from time period to time period
3. Variation due to the particular sample of tasks chosen to represent the quality being measured

Let us now proceed to examine each reliability method, keeping in mind the error sources so we can determine how each method treats each source.

TEST-RETEST METHOD One obvious method to evaluate stability consists of measuring the same individual's performance twice with the same measuring instrument. This type of reliability includes variation sources 1 and 2 as error. Thus the resulting reliability is one which measures the stability of the true score over time. There are numerous problems with the test-retest method that are created by having the individuals measured on the *same* test twice. For example, unless the time period is fairly long between administrations, the variable of a memory factor is likely to bias the responses of people on the second administration. Another difficulty is that variation due to the particular sample of tasks or items chosen is treated as systematic variance which adds to the reliability.

Thus any person who by chance happened to know more answers simply because a few of the test items touched, say, upon a hobby of that person, would also be favored in the second administration because the same items, rather than a new sample, is used. He should therefore score high on both testings due to variation source 3 being treated as true variance.

PARALLEL TESTS METHOD One way of avoiding having error source 3 as true variance is to use two completely comparable or "equivalent" forms of the measuring instrument. These two forms should be as identical as possible, except that *specific* items or questions on each form would *not* be the same although they would each represent a *similar sample of items chosen*. One form may be admin-

istered immediately after the other or they may be administered at spaced intervals, depending upon whether one is concerned with having variation source 2 included as error variance.

This kind of reliability, when spaced testing is used, represents the most rigorous evaluation of stability that can be made. However, it is often impossible or at best extremely difficult to construct alternate forms of a measuring instrument.

How does one construct two alternate but equivalent forms of a measure of job performance or two alternate forms of a personal history form? In many cases, not without considerable difficulty. This lack of a truly comparable measuring device has caused psychologists to look for additional methods of assessing reliability besides the test-retest and parallel form procedures.

SUBDIVIDED TEST METHOD The third major reliability method is often referred to as a measure of the *internal consistency* of a measuring device. It provides an indication of the extent to which people score the same, relative to each other, on different subdivisions of the overall instrument. This method is probably the most widely used method of measuring reliability since it requires that only one form be constructed and yet does not require repeated administrations of that form.

Its mechanics are very simple. In its most basic form, the internal consistency method is the parallel forms procedure in which the parallel forms are two halves of the same test. These half-tests are selected so as to be as equivalent as possible, although often the test is simply divided into two halves by putting all the odd-numbered items into one half and all the even-numbered items into the other half. This is called the *odd-even* version of the *split-half* technique.

It is important to remember that the separation of the total test into equivalent halves occurs only when scoring that test—not when administering it. Since the two subtests are each only half as long as the original, each represents a sample of behavior only half as large as the total test. Thus the correlation (reliability) between halves is likely to be an *underestimate* of the reliability of scores based upon the entire test. To obtain an estimate of what the reliability of the complete test is, the Spearman-Brown Prophecy formula may be applied as follows:

$$r_{tt} = \frac{2r_{\frac{1}{2}\,\frac{1}{2}}}{1 + r_{\frac{1}{2}\,\frac{1}{2}}}$$

where r_{tt} = reliability of the total test (estimated)
$r_{\frac{1}{2}\,\frac{1}{2}}$ = observed correlation between the two halves of the test.

For example, if the observed correlation between halves were 0.40, then the Prophecy Formula would estimate the reliability of the complete test as:

$$r_{tt} = \frac{2\,(0.40)}{1 + 0.40} = \frac{0.80}{1.40} = 0.57$$

The split-half method thus provides a method for estimating reliability with a single test and a single administration. There are, however, certain drawbacks to its use. When one has a test that involves primarily speed factors (such as certain simple clerical tests), the alternate-half procedure gives a spuriously high result.

Since speed tests usually involve easy items, it is only a question of whether they were responded to that determines whether they were correct or incorrect. Thus, splitting the test on an odd-even basis, for instance, would result in virtually identical scores for both halves—thus a high positive correlation.

KUDER-RICHARDSON METHOD Another version of the split-half method is frequently used for measuring reliability. Related to a statistical technique known as the analysis of variance, its most frequent form is known as the Kuder-Richardson procedure. The Kuder-Richardson (K-R) method is also an internal consistency reliability which essentially treats *each test item* as a subtest, so that instead of having two halves there are n subtests, where n is the total number of items on the measuring instrument. The K-R technique is equivalent to computing all possible correlations between pairs of tests items (there will be $n[n-1]/2$ such pairs), taking the average of these, and adjusting the result by using the Spearman-Brown prophecy formula:

$$r_{tt} = \frac{K\bar{r}_{ii}}{1 + (K-1)\bar{r}_{ii}}$$

where r_{tt} = estimated reliability of total test
$\bar{r}_{ii}$ = average correlation between items
K = number of pairs of items

Like the split-half forms procedure, the Kuder-Richardson procedure ignores variation source 2 and is not appropriate for speed tests.

A summary comparison is given in Table 2.4. This table shows the various reliability methods and compares them in terms of the kinds of variation they include as error variance.

TABLE 2.4 *Sources of Variation Represented in Different Procedures for Estimating Reliability*

	Experimental Procedure for Estimating Reliability					
Source of Variation	*Immediate Retest (Same Test)*	*Retest After Interval (Same Test)*	*Parallel Test Form Without Time Interval*	*Parallel Test Form With Time Interval*	*Odd-Even Halves of Single Test*	*Kuder-Richardson Analysis of Single Test*
How much the score can be expected to fluctuate owing to:						
Variations arising within the measurement procedure itself	X	X	X	X	X	X
Changes in the individual from day to day		X		X		
Changes in the specific sample of tasks			X	X	X	X

SOURCE: Adapted from R. L. Thorndike and Elizabeth Hagen. *Measurement and evaluation in psychology and education,* 2nd ed. Wiley, New York, 1963, 182.

DETERMINING THE UTILITY OF A SELECTION INSTRUMENT

The *utility* of a prediction device is the degree to which its use improves the quality of the people being selected beyond what would have occurred had that device not been used. So far we have discussed validity and reliability, both of which play an important role in determining the utility of any selection instrument. In addition there are, however, several other factors which are equally critical in determining utility in any situation involving group selection. These additional variables are (1) criterion reliability, (2) criterion relevance, (3) the selection ratio, and (4) the percent of present employees considered successful.

The reader is cautioned to keep in mind that group prediction is the process of selecting systematically a subsample of applicants who are more apt to succeed, on the average, than either the group as a whole or any subgroup selected at random from the whole. This differs from the process of individual prediction where one is concerned with predicting the likelihood of success of a *particular person*, rather than a group of persons. There are, of course, selection situations that involve both group and individual aspects of selection. One example is a selection program used by the United States Navy in their selection of flight cadets. The Navy uses group prediction techniques prior to and during the several different phases of flight training. In addition, it is also necessary for predictions to be made for a specific cadet and his individual likelihood of success in the program. The need for these latter kinds of predictions occurs whenever a cadet's record during training has been poor enough to bring him before a review board. For this Navy program the same basic predictors are used for both the group and the individual predictions.

The reliability of the criterion, a topic that has been deferred to the criterion chapter (Chapter 6), does not differ in method or logic from predictor reliability; it is usually assessed by one of the previously discussed reliability techniques. Criterion relevance, another topic discussed in greater detail in Chapter 6, is the criterion equivalent of validity but is usually *not* a formal statistical process.

PREDICTOR VALIDITY

The primary statistical index influencing the utility of any prediction instrument is its validity. While it will be demonstrated that low and even zero validity situations can still result in quite successful selection under special conditions, the validity coefficient remains the central variable in selection. To illustrate, consider the dia-

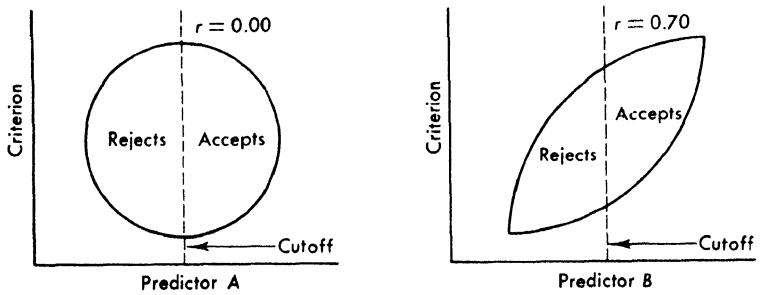


Figure 2.10. Two different predictor-criterion relationships.

grams shown in Figure 2.10 in which two different predictor-criterion relationships are shown, one with a validity of 0.00 and the other with a validity of 0.70. In both cases a cutoff score on the predictor has been established that allows us to take the top 50 percent of the people taking the test. Which predictor will result in the greatest increase in average criterion score of the selected group over what has been achieved by prior methods (that is, random selection)? Looking first at predictor A in terms of how the people are distributed on the *criterion dimension only*, we find that the mean criterion score of the "accepted" group is exactly the same as the "rejected" group. That is, the people who are accepted by taking the top half of the scores on test A do *not* tend to have higher criterion scores than do the lowest 50 percent of the scorers on test A, as shown in Figure 2.11.

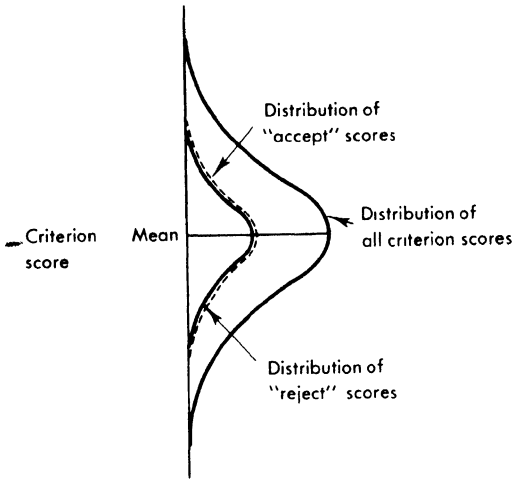


Figure 2.11. Frequency distribution of criterion scores for (1) total group, (2) accepted group, and (3) rejected group using predictor A.

However, when we look at predictor *B* we get quite a different picture. We can see immediately that those people above the cutoff on A seem to do better on the criterion than do those below the cutoff. That is, the people above the cutoff have a higher *average criterion score* than do those below. This is shown in Figure 2.12, which again shows the three distributions of criterion values.

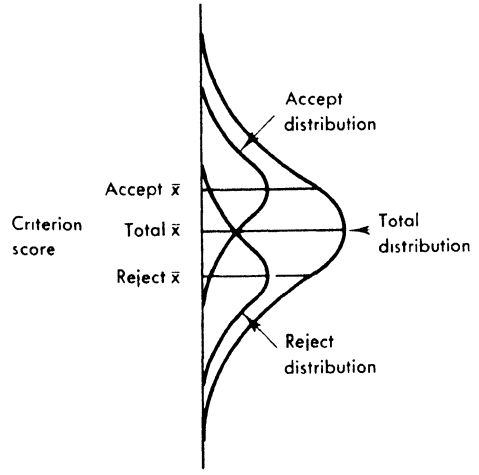


Figure 2.12. Frequency distribution of criterion scores for (1) total group, (2) accepted group, and (3) rejected group using predictor *B*.

Thus we appear to have our first general principle in test utility: given any arbitrarily defined cutoff on a test, the higher the validity, the greater the increase in average criterion score for the selected group over that observed for the total group. In other words, the difference

$$(\bar{x}_{\text{selected group}}) - (\bar{x}_{\text{total group}})$$

will increase in direct proportion to the test validity. Indeed, it can be shown algebraically that this is so (later we will see certain exceptions to this first principle).

Recently Naylor and Shine (1965) have published a set of tables which offer easy computation of the increase in average criterion score which will be achieved with any test given that the test validity and test cutoff point can be specified. This table is given in the Appendix along with explanations and examples of its use.

SELECTION RATIO AND PERCENT OF SUCCESSFUL EMPLOYEES

Two other variables which play an important role in determining the utility of a predictor are the selection ratio and the percentage of present employees considered successful. The reader will recall that the utility of a predictor was defined as the improvement in quality of the hires obtained using a prediction device when compared with present methods of selection. Quality is typically defined in terms of (1) the average criterion score of the group, or (2) in terms of the proportion of people in that group who have criterion scores above some value that is considered to be minimal in order for one to be a successful employee. For any given obtained validity coefficient between criterion and the predictor, a manipulation of either the selection ratio and/or a change in the percent of present employees considered successful will result in marked changes in the resulting quality of the hired (selected) employees.

SELECTION RATIO Simply described, the selection ratio (SR) may be expressed as:

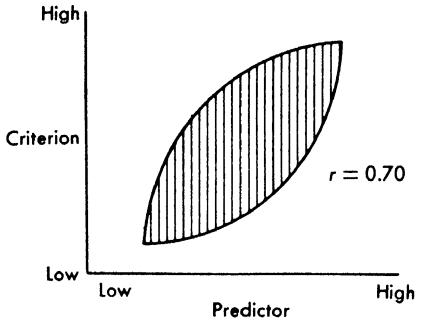
$$n/N = \text{SR}$$

where n = number of job openings
N = number of job applicants available for placement

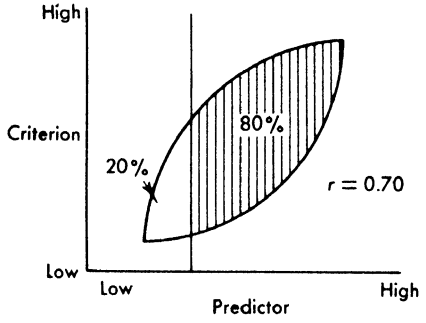
When the SR is equal to or greater than 1.00, the use of any selection device has little meaning. With more job openings than job applicants, the applicant is in a seller's market where the company may need to purchase his services regardless of his quality. If, however, the SR is less than 1.00, then there are more job applicants than positions and the employer is in a position to be selective in terms of whom he hires.

The way in which the SR can influence the selection process can best be demonstrated by referring to Figure 2.13. In Figure 2.13a, a scatterplot of scores is shown that is approximately the shape that could be expected with a large sample of people and a correlation between predictor and criterion of 0.70 (the higher the correlation, the more nearly the scatterplot will approach a straight line; the lower the correlation, the more nearly the scatterplot will approach a circle). The proportion of the oval that is shaded represents the proportion of applicants who are actually

a. Scatterplot showing relationship ($r = 0.70$) between predictor and criterion. SR = 1.00, i.e., all applicants are being hired.



b. Same scatterplot when SR = 0.80, i.e., there are eight job openings for every ten applicants. Thus the lower 20 percent on the predictor can be rejected.



c. Same scatterplot when SR = 0.20, i.e., there are only two job openings for every ten applicants. Thus the lower 80 percent on the predictor can be rejected.

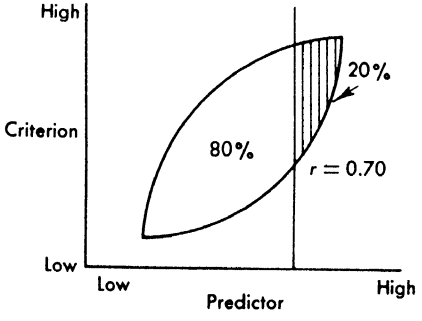


Figure 2.13. Diagrams showing the effect of selection ratio on the average quality of those being hired.

hired, *viz.*, the SR. In Figure 2.13a, an SR of 100 is presented; there is a job opening for every applicant so all will be hired.

In part b of Figure 2.13 we see what happens to the average quality of those being hired when the SR becomes 0.80. Since there are jobs for only 80 percent of the applicants, the employer will logically hire the 80 percent having the highest predictor scores, since the predictor is highly related to subsequent criterion performance. These 80 percent are represented by the shaded area of the oval falling to the right of the cutoff point on the predictor. Since those being eliminated generally possess low criterion scores, it is easy to see how the average criterion score for those hired with an SR of 0.80 is higher than it is if a random group of applicants were placed on jobs as in Figure 2.13a. This increase in average quality is shown even more dramatically in Figure 2.13c which illustrates an SR of 0.20. Faced with a situation where there are ten applicants for every two jobs, the employer is "sitting pretty"—he now can select the top 20 percent of the performers. These individuals are represented by the shaded area of the oval falling to the right of the cutoff in Figure 2.13c. The difference in average quality of criterion score for this select subgroup as opposed to that of the entire group is very large. The benefits to the employer in terms of dollars in this situation should certainly be substantial.

The general principle that a lower selection ratio will always result in better quality employees being hired holds as long as the relationship between the predictor and the criterion is some value greater than zero (negative or positive *r*'s are equally effective if of equal magnitude). In fact, it can be demonstrated that the principle of the selection ratio can be effectively utilized in some cases even if *all* applicants need to be hired. This can occur if there are at least *two* jobs, each with a number of openings and each of which has its own predictor with greater than zero validity.

PERCENT OF PRESENT EMPLOYEES WHO ARE SUCCESSFUL In our discussion concerning validity and the SR we have so far assumed the criterion to be continuous and therefore the higher the criterion score, the more *satisfactory* that worker is considered to be. Let us now suppose a criterion score exists that defines whether a worker is either satisfactory or unsatisfactory—that is, if he performs above a standard he is considered satisfactory and if he performs below this standard he is considered unsatisfactory. The diagrams in Figure 2.14 illustrate this. In part a, a relationship of about 0.70 between the criterion and the predictor is shown. Note that the horizontal line, called the *criterion cutoff*, separates all workers into two groups: those considered successful and those considered unsuccessful. Such a cutoff will, of course, have to be rather arbitrary in its nature. However, in many cases it is not too difficult to arrive at some consensus concerning minimal acceptable performance.

Part b of Figure 2.14 shows the same data with a predictor cutoff based upon a selection ratio of about 0.5. The last part of the figure shows both cutoffs together. When combined in this fashion it becomes possible to distinguish among the various subportions of the data which are formed by the intersection of the two cutoff lines.

Part A. Those applicants who are to the right of the test score cutoff and above the criterion cutoff are called *true positives*. They are those who the test says ought to be successful and who indeed will be successful according to the criterion. They represent correct decisions based upon the test.

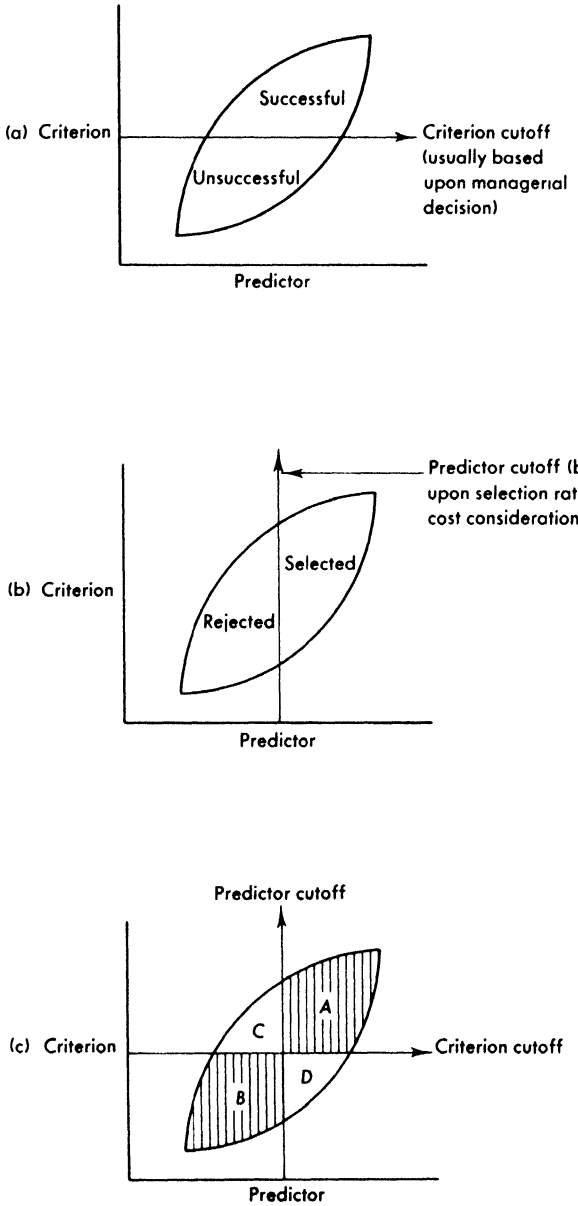


Figure 2.14. Effects of establishing criterion and predictor cutoff points on a bivariate distribution of scores.

Part B. This segment includes those applicants who have scores below the predictor cutoff and below the criterion cutoff. Called the *true negatives*, these applicants, like the true positives, represent correct decisions based upon the predictor.

Part C. These applicants have scores *below* the predictor cutoff but *above* the criterion cutoff. These people would not be hired if hiring decisions were based upon the test, in spite of the fact that their eventual criterion score was high enough to place them in the satisfactory category. This represents one kind of mistake or error that occurs in testing and is referred to as *false negatives.*

Part D. The last segment of the oval consists of job applicants who would be hired but who would subsequently turn out to be unsatisfactory in their work. These persons also represent "mistakes" in the selection process and are known as *false positives.*

Several meaningful ratios may be constructed using the various parts of Figure 2.14c. For example,

$$(1) \qquad \frac{C+D}{A+B}$$

This is a ratio of the number of errors in selection to the number of employees correctly placed. The size of this ratio depends upon all three variables: the location of the criterion cutoff, the location of predictor cutoff, and the validity coefficient. Not only is the size of this ratio affected by these variables, but so is the relative magnitude of the two types of errors, *C* and *D*. Usually the employer is more concerned with minimizing fake positives than he is concerned about the number of false negatives. This often is seized upon by those opposed to testing as being one of the major evils of scientific selection via tests, namely that some people are rejected who would be successful on the job if given an opportunity to prove themselves. The reader will have to deliberate the pros and cons of this problem for himself—the authors merely point out the difficulty. However, the authors hasten to add that industrial psychologists can be as socially minded as their critics. Industrial psychologists generally have the data to tell the whole story, while some critics without any data at all merely "holler" about one error.

Another ratio of importance is given by

$$(2) \qquad \frac{A+C}{A+B+C+D} = \text{percent presently successful}$$

This represents the percentage of the present employees who are satisfactory. It is a base percentage that expresses the degree of success that is obtained with whatever selection methods were used prior to the introduction of the predictor. The third ratio,

$$(3) \qquad \frac{A}{A+D} = \text{percent successful using predictor}$$

is an expression of the proportion of *hired* applicants who will be successful if one uses the predictor as an aid to selection along with the methods currently being employed. To the extent that (3) is greater than (2), the predictor is adding something to the selection process. In comparing the relative magnitude of (2) and (3), some general principles may be stated:

1. For any particular validity and criterion cutoff, a reduction in the SR will cause an increase in the *effective* validity. Thus, one can compensate for low statistical validity if one can be selective in his hires.

2. For any particular statistical validity and selection ratio, the smaller the percentage of present employees considered satisfactory, the larger the percentage *increase* of satisfactory applicants obtained using the predictor. In other words, if we define the difference between ratios (2) and (3) as

$$\text{Utility} = \frac{A}{A+D} - \frac{A+B}{A+B+C+D} = \text{percent increase in effectiveness}$$

where effectiveness is defined as the percent of successes being employed, then the greatest benefit will be observed under those conditions where the poorest job is currently being done—a logical outcome. There are some exceptions, of course. For example, consider Figure 2.15.

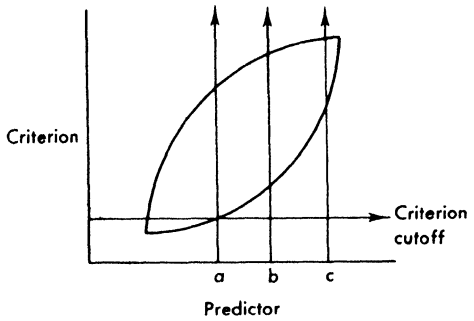


Figure 2.15. Effect of different selection ratios in a situation where the majority of present employees are considered satisfactory.

Notice from Figure 2.15 that no matter which of the three different selection ratios one uses, 100 percent of all hired applicants will eventually be judged satisfactory. Thus here is one situation where large selection ratio changes are of no consequence.

TAYLOR-RUSSELL TABLES

A detailed expression of the exact relationships between the size of the validity coefficient, the selection ratio, and the percent of presently satisfactory employees has been prepared by Taylor and Russell (1939). Under given conditions of validity, selection ratio, and percent satisfactory, their tables allow one to determine the percentage of hires who will be satisfactory using the predictor in conjunction with current methods. However, the Naylor-Shine tables discussed in the section on predictor validity would appear to have several advantages over the Taylor-Russell tables. The Naylor-Shine tables are formulated in terms of differences in average criterion score between the selected group and the original group; Taylor and Russell use differences in the percent successful between the selected group and the original group. Thus, the Naylor-Shine tables would seem to give a more meaningful index of test utility. Also, the use of the Taylor-Russell tables requires that the employees be separated into two groups, "successful" and "unsuccessful," by selecting some arbitrary point on the criterion dimension that represents "minimal satisfactory performance." The Naylor-Shine tables do not require any decision of this sort for their use and are therefore more general in their applicability.

A Note of Caution. Both the Naylor-Shine tables and the Taylor-Russell tables have certain limitations which are *very* important. Both methods for evaluating test utility are based upon the assumptions that (1) the relationship between predictor and criterion is a linear one, and (2) the validity coefficient used is one obtained by concurrent validity procedures.

Smith (1948) and others have pointed out the dangers that exist if one tries to use tables such as Taylor and Russell's under conditions where the relationship is not linear between the predictor and the criterion. Such a relationship is shown in Figure 2.16. When such nonlinear relationships exist, both tables are completely inappropriate for determining test utility.

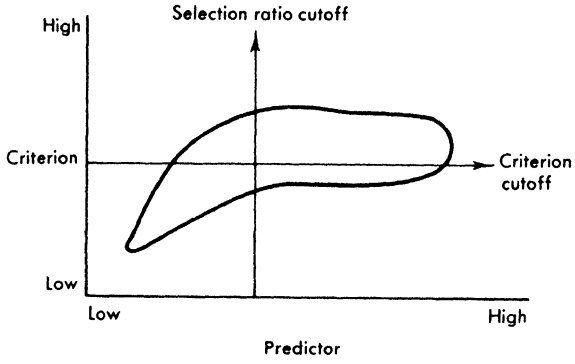


Figure 2.16. Example of nonlinear scatterplot—a situation which makes the Naylor-Shine and Taylor-Russell tables inappropriate.

The fact that both tables assume a validity coefficient based upon concurrent validation procedures may come as a surprise since earlier it was pointed out that concurrent validity was not a particularly good substitute for predictive validity. However, test utility involves determining the increase in either average criterion score (Naylor-Shine tables) or percent of successful employees (Taylor-Russell tables) over that *currently being obtained with present employees.* The basic scatterplot is one based upon present employees hired by the normal selection procedures —the typical concurrent validity paradigm.

RELIABILITY OF PREDICTOR AND CRITERION

Reliability of the criterion and of the predictor are also important, primarily because they influence or put limits upon the size of the validity coefficient that may be obtained. There is a basic algebraic relationship which exists between validity and the reliability of the predictor and the criterion which is

$$r_{pc\,(\text{obtained})} = r_{pc\,(\text{true})} \sqrt{r_{pp} \times r_{cc}}$$

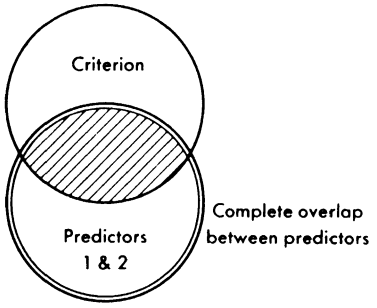
where $r_{pc_{(\text{obtained})}}$ = observed correlation (validity) between the predictor and criterion

$r_{pc\,(\text{true})}$ = "true" correlation (validity) between the predictor and criterion

r_{pp} = reliability of predictor

r_{cc} = reliability of criterion

Because of the intercorrelation between these predictors, the diagram shows that the amount of overlap between predictor 2 and the criterion can be divided into two parts: that area unique to predictor 2 and that area shared with predictor 1. Thus, the use of a second predictor in this situation allows us to account for more criterion variance than could be done using predictor 1 alone, but all of the criterion variance predicted by 2 is not new variance. A general rule can therefore be stated concerning multiple predictors. *All other things being equal, the higher the correlation between predictors, the less the overall prediction will be improved by using both predictors together.* The extreme case, of course, would be the situation where the predictors were perfectly correlated and we would have no additional criterion variance accounted for by the addition of predictor 2 to our selection battery.



In the case of two predictors which are correlated with each other, we can express R^2 as a function of the separate validities and the size of the intercorrelation between predictors with the formula[2]

$$R^2_{c.12} = \frac{r^2_{1c} + r^2_{2c} - 2r_{12}r_{1c}r_{2c}}{1 - r^2_{12}} \tag{3.2}$$

note that if $r_{12} = 0$, then formula 3.2 reduces to

$$R^2_{c.12} = r^2_{1c} + r^2_{2c}$$

which is formula 3.1.

A more explicit illustration of the influence of predictor intercorrelation upon the size of the multiple correlation coefficient may be obtained from Table 3.1, where examples of R and R^2 values are given for pairs of predictors having validities of 0.30, 0.50, and 0.70 under hypothetical conditions of 0.00, 0.30, and 0.60 intercorrelation. Figure 3.3 shows the general trend using the data given in Table 3.1. The moral to the psychologist is quite evident—avoid using predictors which are apt to be highly related to one another.[3]

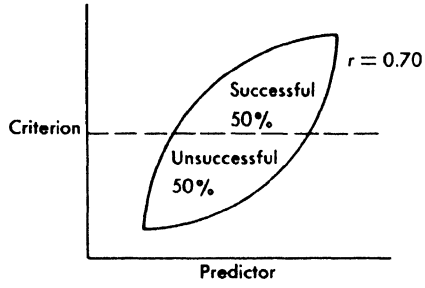
[2] $R^2_{c.12}$ should be read as "the squared multiple correlation" when predicting the criterion from predictors 1 and 2 together.

[3] A procedure which allows selection of the most efficient set of k predictors from a larger group of M predictors is available. The method is known as the Wherry-Doolittle Test Selection Procedure.

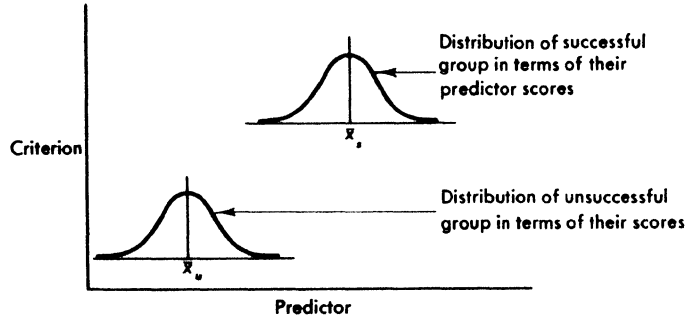
making this assumption it is possible to cast the problem directly in terms of minimizing the total number of both kinds of errors rather than having to weigh the two types of errors by their respective "cost."

LOCATION OF THE CUTOFF POINT

To illustrate how the problem of finding an optimum location for our cutting score can be approached, consider the case where we have a specified validity (e.g., about 0.70) and a specified percent of present employees considered successful (often referred to in this context as the "base rate"). This can be diagramed as follows:

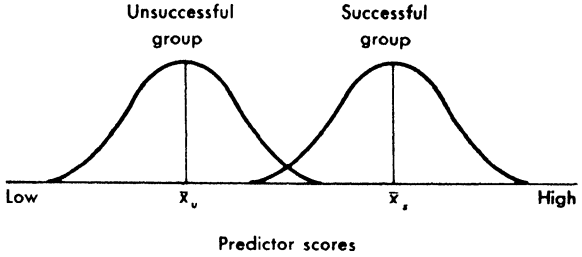


The next step is to present the same data in a slightly different form. First, we know that our total group of employees is assumed to have a normal distribution in terms of their predictor scores. Second, and equally important, *both* subgroups (successful and unsuccessful) are assumed to have normal distributions. By looking at the example above it is easy to deduce that the *mean predictor* score of the successful group is going to be higher than that of the unsuccessful group. We might diagram this as:



Both distributions will be equal in size since they are based upon the same number of people (i.e., 50 percent in each group). There is an algebraic relationship between the difference between the means of the two subgroups as viewed in this fashion and the size of the correlation coefficient. If the group means are significantly different from each other (say at a significance level of 0.05), then the correlation coefficient will also be found to be significant at the same level.

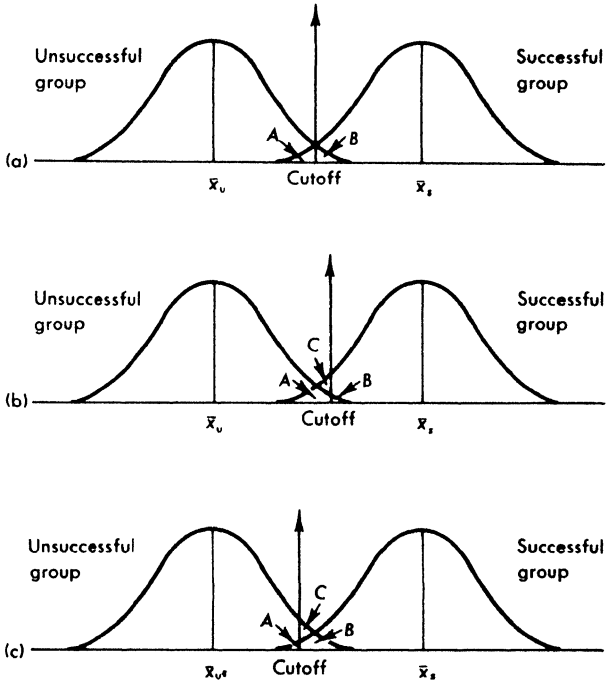
Taking our diagram one step further, we can place the two frequency distributions of the subgroups side by side on the same baseline, as shown below.



After doing this we can now return to our original question—where do we locate a cutoff on the predictor so that the total number of errors will be minimized?

It turns out that the mathematical solution to this problem results in a very simple answer: *The cutoff point which minimizes total error is the point at which the two distributions intersect each other.*

This can be easily demonstrated on a conceptual level by looking at the three cases illustrated below. The same difference between the means (that is, the same correlation) is used in each case—all that has been changed is the location of the cutoff point on the predictor.



In illustration (a), the number of false positives (failures who are above cutoff) is given by the area *B*. The number of false negatives (successes who are below the cutoff) is given by the area A. Thus,

$$\text{Total error} = A + B$$

For illustration (b), the number of false positives is given by B and the number of false negatives is given by $A + C$. Thus,

$$\text{Total error} = A + B + C$$

For illustration (c), the number of false positives is given by $B + C$ and the number of false negatives is given by A. Thus,

$$\text{Total error} = A + B + C$$

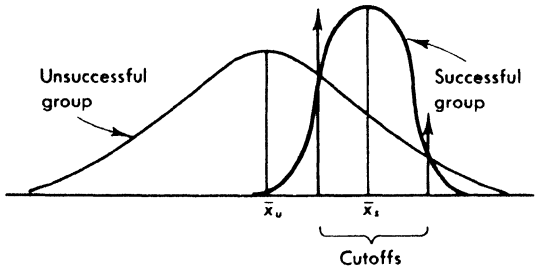
Since inspection of all three illustrations quickly confirms that the area $A + B$ is the same for all three cases, then it is obvious that error is increased by some amount C whenever the cutoff is moved *away* (in either direction) from the point at which the two distributions intersect each other.

SOME UNUSUAL RAMIFICATIONS

We now have a general principle for locating a cutting score that will minimize the total number of errors in a selection decision-making situation—namely, at the point of intersection. It turns out that, as long as both types of errors are considered equally costly, this is a very general rule and is *not* affected by (1) the relative sizes of the two groups (i.e., percent considered successful), or (2) the respective variances or dispersions of the two distributions.

This leads to some interesting and very important aspects of the general prediction problem concerning the relationship of test validity to test utility. Rorer, Hoffman, LaForge, and Hsieh (1966) have pointed out three such interesting cases.[1]

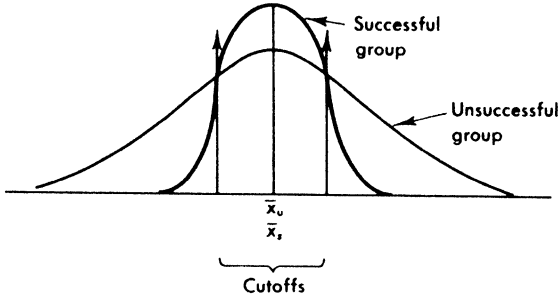
Case 1: Both the means and the variances of the two groups differ from each other. Suppose that our successful group is of equal size to the unsuccessful group and has a significantly higher mean on the predictor, but its variance is much smaller. A diagram of such a situation is as follows:



Our principle of establishing cutoff points says that we should place them wherever the two distributions intersect. Note that this happens *twice* in this particular case. Thus, we have an upper cutoff and a lower cutoff. We should select only those people who fall within the interval between cutoffs in terms of their test score. Any *other* cutoff points will result in greater total error than would be obtained with the ones located at the points of intersection.

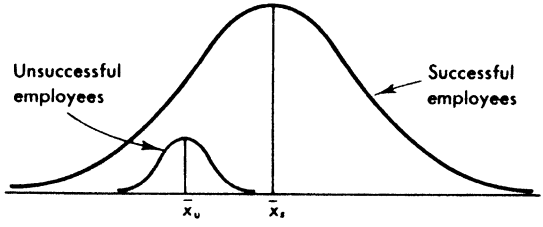
[1] It should be pointed out that the assumptions for the use of the correlation coefficient are violated in this case (unequal variances for subgroups), so that it should not even be used as a measure of relationship.

Case 2: Groups have equal means but different variances. In this very interesting case the two groups do not differ in terms of their mean predictor score—that is, *on the average* the unsuccessful employees do just as well on the test as do the successful employees. This implies that the correlation coefficient is zero between the predictor and the criterion. However, we have further stated that the two groups differ in terms of their variability. If we assume the successful group is the group with the smaller variability for purposes of exposition, we can express this diagrammatically as follows:



Even though the two groups have the same mean criterion score, it is possible to develop cutoff points which will improve prediction over that currently being enjoyed through present methods, since the two distributions intersect at two points due to their unequal variability. Thus, we have the unique situation where there would be no apparent validity (as measured by a correlation coefficient) but where prediction can be much improved by use of appropriate cutoffs.[2]

Case 3: Group means are considerably different but group size is also greatly different. Suppose we are dealing with a situation in which the base rate of unsuccessful employees is very small, that is, about 90 percent of our present employees are considered successful. Such a situation is shown in the following diagram.



Here we have another unique situation. Even though the group means may be substantially different thus giving a substantial correlation between criterion and predictor, it is not going to be possible to establish *any* cutoff which will result in reducing error over what is currently obtained with present methods. Because of the marked difference in size between the two groups, we see that the two dis-

[2] It should be pointed out that again the assumptions for the use of the correlation coefficient are violated in this case (unequal variances for subgroups), so that it should not even be considered as a measure of relationship.

tributions do *not* intersect at any point. Under our present selection system we are only making errors 10 percent of the time. If we move our cutoff from left to right in case 3 (it is located in the extreme left to begin with, since we are currently selecting all these people) we will, of course, start eliminating some of the unsuccessful people currently being employed under the present system. At the same time, however, we are going to start rejecting employees who would turn out to be successful. Looking at the diagram quickly tells us that this increase in false negatives would be greater than the corresponding decrease in false positives no matter where we put our cutoff. Thus, any test-based cutoff will result in more errors than we have without the test, even though the test is highly valid.

SOME CONCLUSIONS

The basic decision-theory approach to selection and placement outlined above has a number of advantages over the more classical approach based upon the correlation model. Recent work by Rorer, Hoffman, LaForge, and Hsieh (1966), Cronbach and Gleser (1965), Darlington and Stauffer (1966), and Mahoney and England (1965) indicates the growing interest in the decision-theory approach as applied to selection problems. There is no question but that it is a more general and better model for handling this kind of decision task, and we predict that in the future problems of selection and placement will be treated in this context more frequently—perhaps to the eventual exclusion of the more stereotyped correlational model.

REFERENCES

Cronbach, L. J., and G. C. Gleser. *Psychological tests and personnel decisions*, 2nd ed. University of Illinois Press, Urbana, Ill., 1965.

Darlington, R. B., and G. F. Stauffer. Use and evaluation of discrete test information in decision making. *Journal of Applied Psychology*, 1966, 50, 125–129.

Mahoney, T. A., and G. W. England. Efficiency and accuracy of employer decision rules. *Personnel Psychology*, 1965, 18, 361–377.

Naylor, J. C., and L. C. Shine. A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 1965, 3, 33–42.

Rorer, L. G., P. J. Hoffman, G. E. LaForge, and K. Hsieh. Optimal cutting scores to discriminate groups of unequal size and variance. *Journal of Applied Psychology*, 1966, 50, 153–164.

Smith, M. Cautions concerning the use of the Taylor-Russell tables in employee selection. *Journal of Applied Psychology*, 1948, 32, 595–600.

Taylor, H. C., and J. T. Russell. The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 1939, 23, 565–578.

Thorndike, R. L., and E. Hagen. *Measurement and evaluation in psychology and education*, 2nd ed. Wiley, New York, 1963.

3 PROBLEMS AND PROCEDURES IN SELECTION AND CLASSIFICATION

Chapter 2 described the prediction process as involving a single predictor of job success and a single criterion measure. However, such simple situations are rather rarely used in industry. More often, the industrial psychologist simultaneously considers several potential predictors (test scores, background data, etc.) as a means of improving the quality of those being hired. Similarly, he considers several criteria measures (amount produced, supervisor rating, job tenure, etc.) as indicators of job performance and success. His objective becomes one of trying to predict several criteria from several predictors, and this involves a process known as a multiple prediction. This chapter examines the multiple-prediction problem in considerable detail by comparing and evaluating some of the major procedures and methods currently available.[1]

A second simplification made in Chapter 2 was the treatment of selection and placement as two separate processes which used the same basic prediction paradigm. It is important that the reader remain aware that in reality the two processes are inseparable. As Horst (1961) has pointed out, one cannot logically select a sample of n people out of a population of N individuals and *then* assign these people to different jobs. If the assignment to different tasks takes place as a separate process after selection, one is left with no basis for selecting the n people in the first place. In other words, unless some job or task is involved from which a criterion of "goodness" can be obtained, it is impossible to have meaningful selection. Similarly, if a job or task *is* involved, then so is the process of placement.

In examining the major selection and placement methods it will become evident how inseparable the two processes really are. However, in some methods, notably the multiple hurdle model, selection and placement are forced to be somewhat distinct and separate stages.

[1] The authors suggest that this chapter be read in one of two ways. The easiest will be merely to understand on the verbal level the statistical concepts used. The second method of reading will be to absorb the statistical techniques. Which method is used will depend upon the purpose and motivation of the reader.

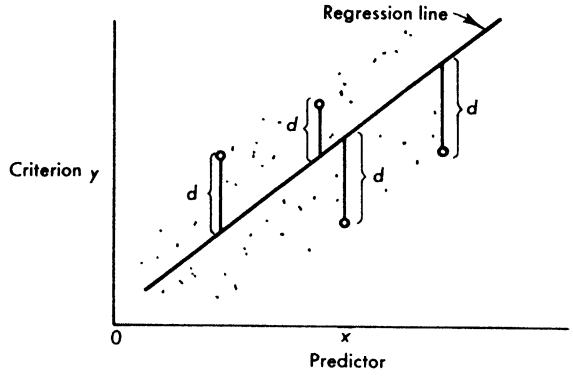


Figure 3.1. Best-fitting line for a set of points.

MULTIPLE PREDICTION

Before proceeding to an examination of the basic selection models which are available to the psychologist, it is necessary to concern ourselves with a brief look at the general multiple prediction model. This model is usually referred to as the *multiple regression model.* In Chapter 2 we assumed a linear world and said that in the general prediction paradigm we develop a regression line to fit the set of data points defined by people's scores on a predictor (the x-axis or abscissa) and on the criterion (the y-axis or ordinate). Figure 3.1 shows such a situation.

The regression line in Figure 3.1 is a straight line and is located so that the sum of the squared distances from each point to the line (running parallel to the y-axis) is as small as possible. We use a best-fitting straight line since we have assumed a linear relationship between x and y. The basic formula for a straight line is

$$\hat{y} = a + bx$$

where $\hat{y}$ = predicted score on criterion

a = a constant indicating the point at which the regression line crosses the y-axis

b = slope of the line, represented by $\triangle y/\triangle x$, or the change in y observed for a corresponding change in x

x = observed score on predictor

Thus, the basic regression line model appears as shown in Figure 3.2.

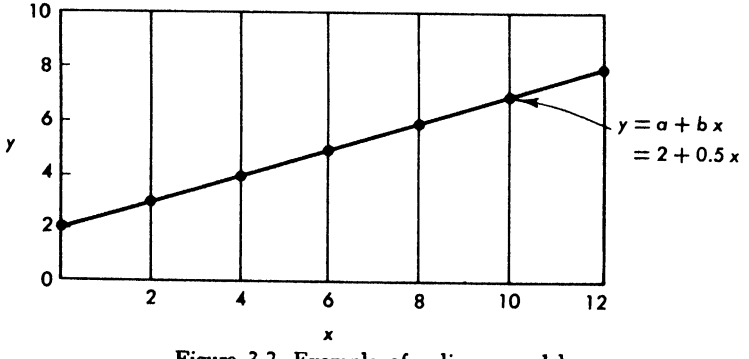


Figure 3.2. Example of a linear model.

Note that in Figure 3.2 the regression line crosses the y-axis at a value of 2. Thus $a = 2$. Also note that for every 2-unit increase in x there is a corresponding 1-unit increase in y. *Thus* $\triangle y/\triangle x = \frac{1}{2} = 0.5 = b$. The regression equation then becomes

$$\hat{y} = 2 + 0.5x$$

Given any x value, we have a regression line that allows us to predict a y score corresponding to it. For example, if x were 8, then

$$\begin{aligned}\hat{y} &= 2 + 0.5\,(8)\\ &= 2 + 4\\ &= 6\end{aligned}$$

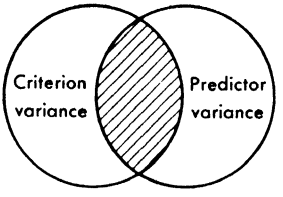
To summarize: In the single predictor case, one computes a best-fitting straight line to the observed points, where the term "best fit" means the sum of squared deviations of the observed values around the line will be a minimum. The formulas necessary to compute the constants a and b which define this best-fitting line are called "least-squares" formulas and are as follows:

$$b = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma x^2 - (\Sigma x)^2}$$

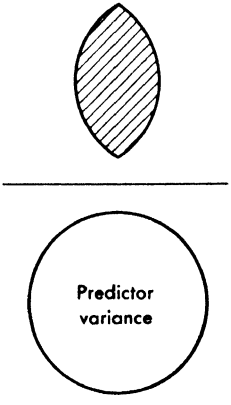
and

$$a = \bar{y} - b\bar{x}$$

The formula for b is a ratio of the covariance between the predictor and criterion and the total variation in the predictor. Thus if we have



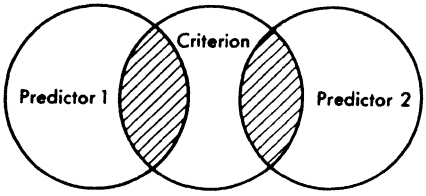
then b equals



and if the reader will refer back to the diagram on page 31 of Chapter 2 he will see that when the criterion variance and predictor variance are equal, then $b = r$, or the slope of the regression line equals the correlation coefficient.

TWO PREDICTORS

It is logical to assume that if predictor x_1 can contribute to the successful prediction of criterion scores, and if predictor x_2 can also contribute to the successful prediction of criterion scores, then using both predictors together should allow for better overall prediction than using either predictor individually. However, the degree to which the two predictors (when combined) will improve the predictability depends upon several factors, most important of which is the correlation between the two predictors themselves. Consider, for example, the situation where two predictors each correlate substantially with a criterion but do *not* correlate with each other, as follows:

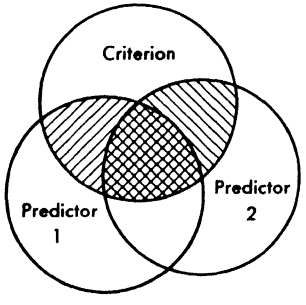


Clearly, a great deal of additional criterion variance can be explained using predictor 2 along with predictor 1. The combined relationship between two or more predictors and a criterion is called a *multiple* correlation and has the symbol R. As was the case with r^2, the value of R^2 represents the total amount of criterion variance which can be explained by using several predictors. When predictors 1 and 2 are not correlated with each other, the squared multiple correlation coefficient can be shown to be an additive function of the individual squared correlation coefficients, or

$$R^2_{c.12} = r^2_{1c} + r^2_{2c} \tag{3.1}$$

Thus, when r_{12} (the intercorrelation of predictors) is zero, then the squared multiple validity is the sum of the squared individual validities.

When two predictors *are* correlated with each other, things become somewhat more complex. Consider a situation (as in the following diagram) where each predictor has substantial individual validity but where r_{12} is also rather large.



concerned with the total efficiency of the system rather than the extent to which each person is placed on the job he can perform best. It avoids putting anyone on a job where his performance will be substandard.

MULTIPLE CUTOFF SYSTEM

It was clearly pointed out in the discussion of the multiple regression system that the model used assumes linear relationships between the predictors and the criterion.[5] Such a system is often objected to on the basis that while for many traits there may be a linear relationship between predictor and criterion over most of the range, there may be some *minimum acceptable amount* of this trait that is necessary in order to be a successful worker. This kind of relationship between job performance and test is shown in Figure 3.7.

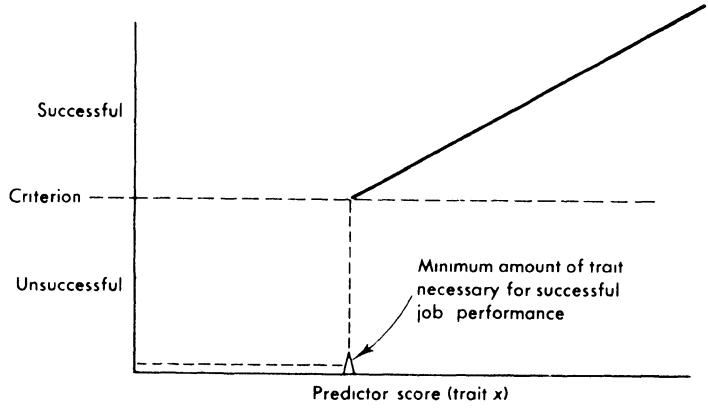


Figure 3.7. An illustration of the distortion of the predictor-criterion relationship that occurs if everyone below a certain criterion value is arbitrarily assumed unsuccessful, regardless of his other abilities.

The predictor-criterion function in Figure 3.7 shows what happens when one assumes that (1) there is some minimum amount of the predictor ability (trait X) necessary for job success, and (2) any lack or deficiency in trait X below this minimum *cannot* be compensated for by having a great deal of some other ability which has also been shown to predict job success. One example of such a situation might be an assembly job requiring both good vision and manual dexterity. Generally speaking, one might find that the better a worker's vision and the better his dexterity, the more successful that worker would tend to be on the job. However, there might be a point along the vision dimension beyond which no amount of dexterity would help.

The selection and placement procedure which takes into account this problem of minimum acceptable values is called the *multiple cutoff method*, meaning that a cutoff point is established separately for each predictor. Unless a person has a score above the cutoff on *all* predictors for a given job, he will not be placed on that job. Thus, no concept of additivity of traits exists with this method. Falling below minimum on any predictor will disqualify the individual.

Figures 3.8 and 3.9 show the regions of acceptance and rejection using the

[5] It also assumes linearity between predictors when expressed as in the equation $\hat{y}=a+b_1x_1+b_2x_2$.

TABLE 3.1 *The Effects of Correlation Between Predictors upon the Squared Multiple R Obtained Using Those Predictors*

r_{1c}	r_{2c}	r_{12}	r_{1c}^2	r_{2c}^2	R	R^2
0.30	0.30	0.00	0.09	0.09	0.42	0.18
0.50	0.50	0.00	0.25	0.25	0.71	0.50
0.70	0.70	0.00	0.49	0.49	0.99	0.98
0.30	0.30	0.30	0.09	0.09	0.37	0.14
0.50	0.50	0.30	0.25	0.25	0.62	0.38
0.70	0.70	0.30	0.49	0.49	0.87	0.76
0.30	0.30	0.60	0.09	0.09	0.34	0.12
0.50	0.50	0.60	0.25	0.25	0.56	0.31
0.70	0.70	0.60	0.49	0.49	0.78	0.61

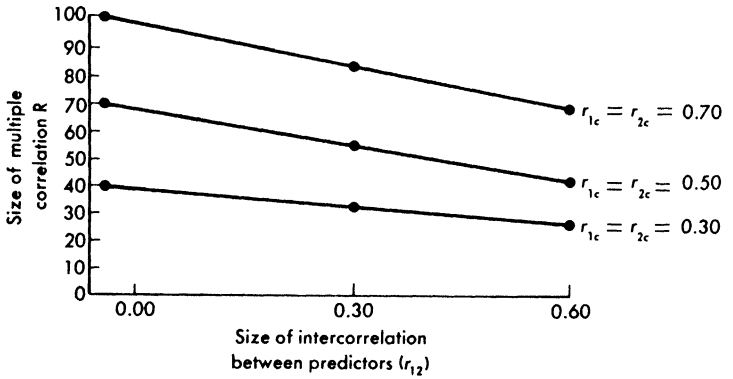


Figure 3.3. An illustration of the effect of predictor intercorrelation upon the multiple correlation coefficient.

PREDICTION EQUATIONS

The prediction equation in a two-predictor situation is an extension of the one-predictor model. The general form of the equation is

$$\hat{y} = a + b_1x_1 + b_2x_2 \tag{3.3}$$

This is the equation for a plane instead of a straight line. For the reader familiar with geometry, Figure 3.4 presents a three-dimensional drawing of the relationships between the variables x_1, x_2, and y corresponding to equation 3.3. Formulas are available which allow for computing the constants a, b, and b_2 which will result in the best-fitting regression plane. Once these constants have been determined, the resulting equation can then be used to make criterion performance predictions of new job applicants, given their scores on the separate predictors.

To illustrate, suppose data are available on 100 men hired for job X during a particular month which includes scores in two tests as well as criterion data after a six-month period. These data can be analyzed to determine the values for a, b_1,

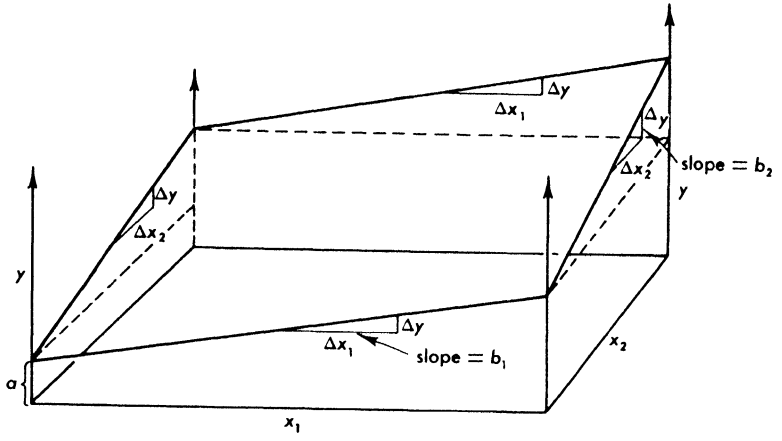


Figure 3.4. The geometry of a two-predictor regression equation. The tilted "top" of the box is the regression "plane."

and b_2 which best described the relationships between the variables. Suppose the following equation was the end result:

$$\hat{y} = 2 + 0.5x_1 + 0.9x_2 \tag{3.4}$$

This equation says that the most likely criterion score for any new hire will be equal to one-half his score on test 1 plus nine-tenths his score on test 2 plus two.[4] Thus if a new applicant scores 20 on test 1 and 30 on test 2, his predicted criterion performance at the end of six months from time of hire would be:

$$\begin{aligned}\hat{y} &= 2 + 0.5\,(20) + 0.9\,(30)\\ &= 2 + 10 + 27\\ &= 39\end{aligned}$$

The extension of the two-predictor model to a k-predictor model, where k is some large number of potential prediction of job success, is not too difficult conceptually. Our model expands to the form

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k \tag{3.5}$$

However, the computational procedures for solving for the least-squares values of all the constants in such an equation becomes rather complex unless one has computer facilities available. The reader is also cautioned to remember that in all the preceding discussion there has been the implicit assumption of a linear world, i.e., all the relationships between pairs of variables are linear ones. It is possible to modify the multiple regression model to avoid this assumption, but that is beyond the scope of this book.

PREDICTION SYSTEMS

By this time it should be clear that the typical prediction problem, be it selection, placement, or both, involves using a number of predictors. These predictors are used in the best manner possible as guides in making a decision about employment.

[4] We of course assume our new hire comes from the same population as the 100 men from whose data we derived our equation.

There are decisions such as "should he be hired for this job?" or "should she be sent to this training program?" There are several strategies that the psychologist can adopt in terms of his approach to the decision-making process. Depending upon the particular prediction system adopted, employment decisions may turn out quite differently. While each system has its own advantages and disadvantages, each provides a method for making decisions about people based upon a group of traits or qualities (the predictors) believed to be relevant to job success. The major systems are: (1) the *multiple regression system*, (2) the *multiple cutoff system*, (3) the *profile matching system*, and (4) the *multiple hurdle system*. Each system will be examined in the following sections in greater detail.

MULTIPLE REGRESSION SYSTEM

As the name implies, this placement system utilizes the multiple regression model for making decisions about individuals. The multiple regression model takes the form

$$\hat{y} = b_1x_1 + b_2x_2 \qquad \text{(assume } a = 0\text{)} \qquad (3.6)$$

The use of such a model in selection assumes that (1) the traits x_1 and x_2 are linearly related to criterion performance, and that (2) possession of a "lot" of one of the traits compensates for having only a "little" of the second trait.

Given a situation, for example, where $b_1 = 2$ and $b_2 = 4$ and $a = 0$, the formula

$$\hat{y} = 2x_1 + 4x_2 \qquad (3.7)$$

would be used for predicting job success. Let us suppose that a criterion score of 50 could be considered satisfactory performance by the employees and anything less resulted in nonsatisfactory performance. Table 3.2 shows some test scores on the two predictors for four theoretical job applicants. The predicted criterion score for each applicant has also been computed using equation 3.7. Note that all four applicants have exactly the same predicted criterion performance even though their test score patterns differ quite markedly. As we proceed from person A through person D, we see that their scores on test 2 systematically diminish. However, this drop is compensated for by a corresponding increase in test 1 performance. In fact, a close inspection will show that a gain of two points on test 1 is necessary to compensate for the loss of each point on test 2. This should not be surprising, since the relative weight given to test 2 is twice that given to test 1 in our regression model (i.e., $b_1 = 2$, $b_2 = 4$).

Figure 3.5 shows even more clearly the dynamics of the selection process created by the data in Table 3.2. The envelope of scores shown in the scatterplot of Figure

TABLE 3.2 *Predictor Scores and Predicted Criterion Scores Using* $\hat{y} = 2x_1 + 4x_2$ *for Four Theoretical Job Applicants*

Applicant	x_1 *Score on Test 1*	x_2 *Score on Test 2*	$\hat{y}$ *Predicted Criterion Score*
A	5	10	50
B	11	7	50
C	15	5	50
D	25	0	50

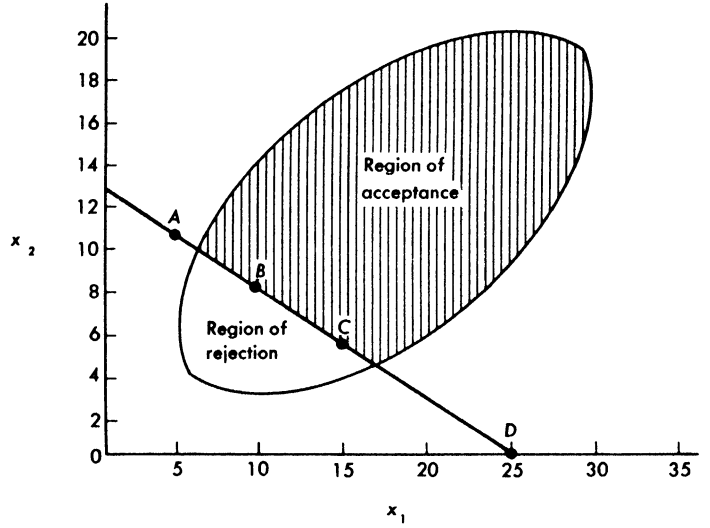


Figure 3.5. Regions of rejection and acceptance for job applicants formed by a situation where the regression equation is $y = 2x_1 + 4x_2$. Example shown illustrates a situation where x_1 and x_2 are moderately correlated with one another.

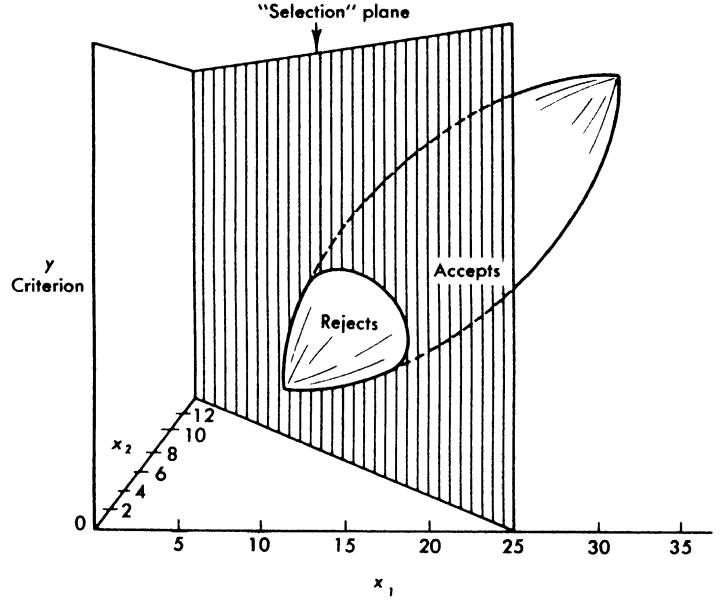


Figure 3.6. Extension of Figure 3.5 to three dimensions so that the observed criterion scores can also be shown. The scatterplot of scores, assuming linearity, will form a shape similar to that of a football. The rejection line is actually a plane that bisects this football into two parts. All people on the far side of the plane will have predicted criterion scores of 50 or better; all those on the near side will have predicted criterion scores of less than 50.

3.5 presents a situation where the two predictors of performance, x_1 and x_2, are positively correlated. If the correlation r_{12} were zero, the scatterplot would of course be a circle. However, the shape of the scatterplot is not critical to the trade-off concept inherent in the multiple regression system. Since we have said that any person with a predicted score of 50 or better was to be considered "satisfactory," we can plot the "50-point line" in Figure 3.5 which shows all possible combinations of test 1 and test 2 scores which will result in a criterion score of exactly 50 points using equation 3.7. As the figure indicates, all four applicants lie along this line.

An interesting aspect of Figure 3.5 is that the line divides the population of job applicants into two groups or regions. All applicants to the right and above the line are going to have criterion scores (using equation 3.7) which will be above 50. All applicants to the left and below the line will have criterion scores of less than 50. Thus, only the former will be accepted for employment since it is predicted that their performance will be satisfactory. The latter applicants, with predicted performance less than satisfactory, will be rejected with this selection system. Figure 3.6 extends Figure 3.5 into three dimensions, showing the observed criterion scores as well as the predictor scores for all individuals.

It is important to note that the plane in Figure 3.6 that divides the employees into those who would be selected using the multiple regression model given by equation 3.7 and those who would be rejected is *not* the regression plane. It is more properly called the *selection plane*. The reader is referred back to Figure 3.4 for an illustration of the regression plane in a two-predictor multiple regression system.

ASSUMPTIONS, ADVANTAGES, AND DISADVANTAGES OF MULTIPLE REGRESSION SYSTEM

The multiple regression prediction system is a powerful selection procedure when used appropriately. Provided the basic assumption that all relationships are linear is true, it has a mathematical elegance that is hard to exceed. One knows, for example, that the model minimizes the errors in prediction. Another advantage of this system is that predictors are combined to obtain the most efficient estimate of subsequent performance.

One of the major points of controversy concerning the multiple regression model involves the trade-off principle so implicit in its use. Whether X units of one variable can be substituted for X units on another variable is always a moot question.

Certainly the method can be an extremely flexible one. It is possible to set up equations for each of a number of jobs using either the same or different predictors. As a result, predicted scores can be computed for each person for each job. People could then be hired and placed on a specific job using one or more of the following procedures:

1. Place each person on that job for which the predicted score is highest. This assumes the organization will profit most if each person is placed where he has the most aptitude, regardless of the absolute amount of that aptitude. If no positions are open in that job, he would be placed on another job for which he received the second best criterion score. One problem with such a procedure is that it is possible that the jobs themselves might have different minimum requirements for success. Thus, it could happen that his best score (predicted performance for job A) might not be adequate for predicted success on job A, while his second best score (predicted performance on job B) might be well above the value needed to predict success on job B.

2. Place each person on that job where his predicted score is farthest above the minimum score necessary to be considered satisfactory. This method is more

but like all good things it may be overdone. When unscrupulous individuals who claim to be psychologists fail to maintain rigidly high standards imposed by professional practice, then the use of testing can be harmful. All claims must be based upon substantial data, and they must be examined. The profession would rather have psychological testing proceed slowly than have it suffer a series of setbacks because of exaggerated claims or invalid uses. Precisely because of the generalized acceptance of tests, considerable caution must be exercised.

COMPETENCE OF THE TESTERS Before a business concern hires a psychologist or a professional organization to introduce a testing program, it should take the precaution of determining professional competence. As suggested in Chapter 1, the attainment of graduate degrees in psychology, membership in professional societies, and relative extent of experience should serve as guideposts. Costs should be related to the professional level of the staff and time spent in performing services. Further, the competent psychologist cannot guarantee absolute or positive results. All that he can do is indicate the possibilities and degree of expected success.

"SECRET" TECHNIQUES AND TESTS In the profession of psychology, and specifically in the use of psychological tests, secret formulas and secret techniques are highly unlikely. People who claim to use them are either unprofessional or unethical and should be so considered until they offer evidence to prove their claim. This evidence should be independently evaluated by professionals.

USING TESTS WITHOUT AMPLE REASON Previous sections have stressed the need for the statistical validity of a test to be established prior to its use. Sometimes all that is necessary for a test to become part of a selection procedure is for it to have a catchy name or to "look like" it should measure whatever skill is wanted (i.e., have face validity to the user).

Some years ago in New York City, finger dexterity testing was the vogue. One of the department stores used such a test to select packers, and within a short time many other stores followed suit. These other stores had no data but assumed that the first one knew what it was doing. Actually, this was not the case. The first store based its approval of the testing program on scanty research and an ambiguous report. Finger dexterity tests can be used with some success for certain purposes, but they have no value in the selection of packers. Blum and Candee (1941a) obtained correlations of about zero between the test results and the production records of a group of department store packers and wrappers. In a follow-up study the same authors found that speed and accuracy in clerical detail bore a closer relationship to employee production than finger dexterity (Blum and Candee, 1941b).

It is necessary to have an understanding of the background of psychological testing before using such tests. Otherwise one is not in a position to know the difference between a psychological test and a mere series of questions.

A joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurements Used in Education has published a set of recommendations in connection with test standards (American Psychological Association, 1954). Were all test developers and publishers to adhere rigidly to these recommendations, the consumer of tests would be on safer ground. The serious student of testing should be completely familiar with this set of standards.

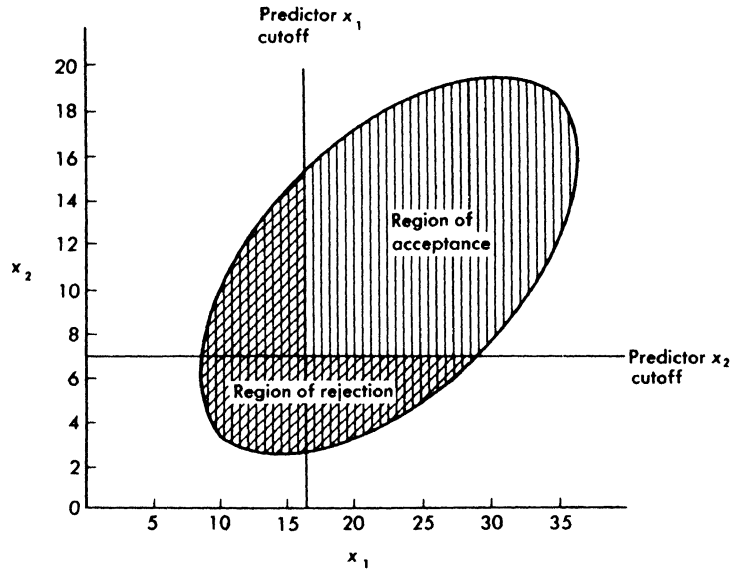


Figure 3.8. Regions of rejection and acceptance for job applicants formed by a multiple cutting score system situation where the cutting scores for x_1 and x_2 are 17 and 7 respectively.

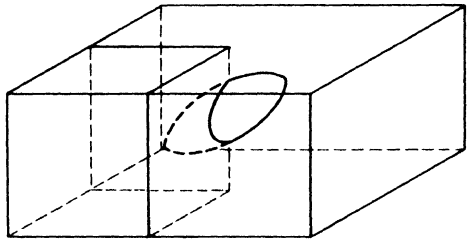


Figure 3.9. Extension of Figure 3.8 to three dimensions so that criterion scores for all individuals are represented. That section of the football scatterplot inside the small box is the rejects; that which protrudes outside is the acceptances.

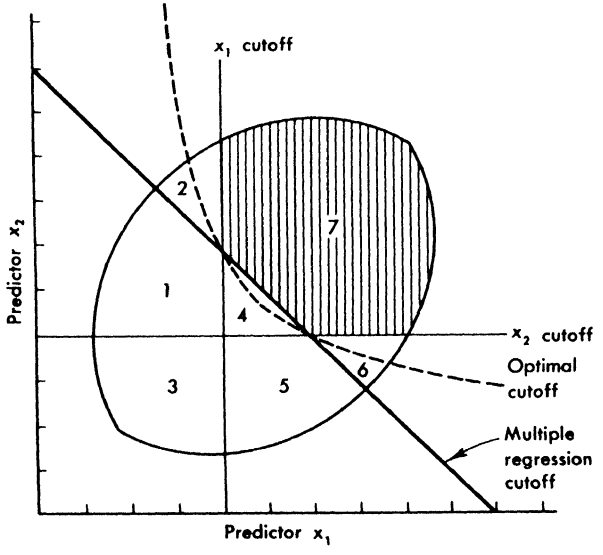


Figure 3.10. Comparison of the regions of acceptance and rejection for the multiple regression and multiple cutoff procedures. (See text for explanation.)

multiple cutoff system for data similar to that used to illustrate the multiple regression system in Figures 3.5 and 3.6.

Perhaps the best way to compare the two methods is to indicate how they differ in terms of who will be selected for the job. Figure 3.10 shows the cutoff lines for both selection methods. Note, first of all, that regardless of the method used, those people in area 7 will *always* be accepted and those people in areas 1, 3, and 5 will *always* be rejected. The people who will be differentially treated as a function of the selection procedure are those in areas 2, 4, and 6.

Using the multiple regression selection system, all people in areas 2 and 6 will be accepted while those in area 4 will be rejected. The reverse will occur using the multiple cutoff procedure; the people in area 4 will be accepted and those in areas 2 and 6 will be rejected. Thus the question resolves to one of the relative desirability of these two groups of individuals.[6] The solution is mathematically complex and has been shown by Lord (1963) to be primarily a function of the reliability of the two predictors. In fact, under most conditions probably *neither* procedure gives exactly the best solution in selecting that group of employees with the highest average criterion score. Instead, the optimal selection strategy appears to be some form of compromise between the two methods (see the dotted line in Figure 3.10).

DETERMINING CUTTING SCORES If one adopts the multiple cutting score technique, it becomes necessary to decide upon the separate minimum acceptable scores for each of the predictors. This is not an easy task since there is no specified "correct" way of setting up a score below which all people will be disqualified. If the reader will refer back to Chapter 2 and reconsider the relationships involved in the selection ratio and the percent of employees considered satisfactory (the cutting score), he will begin to see how complex the problem is when two predictors are involved. Generally, the process of setting up cutting score values becomes one of trial and error in which different values for each predictor are tried. For each pair of cutting scores, the researcher must determine how high the average or composite criterion score of those selected is with respect to other cutting score combinations. He must also take into account the number of job openings with respect to the total number of applicants (the selection ratio measure).

ASSUMPTIONS, ADVANTAGES, AND DISADVANTAGES OF MULTIPLE CUTTING SCORES To summarize the points stated above, the method of cutting scores really assumes a nonlinear relationship between predictors and criterion. Second, it disavows the concept of substitution of test scores, at least in certain parts of the range. The one clear advantage is that it is usually an easy method for the personnel man to implement because no elaborate computational procedures or formulas are required. However, as was mentioned, a certain amount of trial and error is necessary to get cutting scores which will work in the most satisfactory manner. One of its more critical disadvantages is that it does not provide a single score for each individual which can be used to predict how successful he will be on one job relative to his success on another job. Thus, actual job placement via cutting scores can become exceedingly cumbersome.

[6] The reader is cautioned not to interpret equal areas of Figure 3.10 as representing equal numbers of employees. Because the scatterplot is really a *bivariate normal distribution* (i.e., a composite of two normals for a given correlation) the density (number of people) on the periphery of the scatterplot is much less than the number in the center. Therefore, area 4 probably contains many more people than areas 2 and 6 combined, even though the actual size of the areas would indicate the reverse.

PROFILE MATCHING SYSTEM

A third approach to employee selection and placement is the profile matching system. There are numerous versions of this method which differ primarily in terms of the way profiles are matched. However, the remaining aspects of the procedure are rather invariant from version to version.

The method itself is rather simple. If one has k variables (predictors) which are accepted as important to success on the job, then one measures all "successful" employees on the job on each of these k predictors. The scores are then averaged to obtain a "typical" profile of a successful worker. A hypothetical typical profile is shown in Figure 3.11. In this example ten predictors have been used to describe

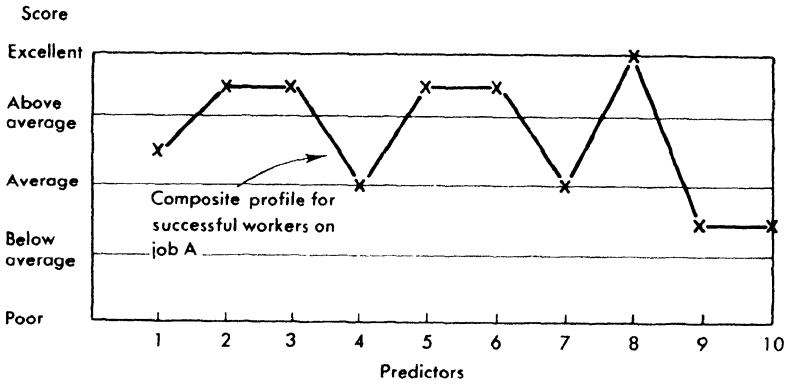


Figure 3.11. Example of a hypothetical profile composed of the average score for all workers considered successful on job A for each of ten predictors. Thus predictor 1 might be intelligence, predictor 2 might be clerical ability, etc.

the typical successful worker on job A. As the data indicate, a successful worker on job A will tend to have high scores (relative to other workers) on variables 2, 3, 5, 6, and 8. His scores on variables 1, 4, 7, 9, and 10 do not differ much from the average performance of workers in general. Once this sort of *ideal* profile has been obtained, it is then used as a standard against which the individual profiles of all new applicants are compared.

At this point two rather important questions arise in the profile method. First, how does one decide what predictors are relevant, that is, which ones should be included on the profile? Second, given that the profile elements have been successfully selected, how does one adequately judge the degree to which any individual's profile matches the ideal profile? The manner in which these two problems are solved can tremendously influence the eventual soundness and validity of any profile matching system.

SELECTION OF PROFILE ELEMENTS Each profile element is used as a predictor of job success, as are the predictors in the methods previously discussed. It is then certainly essential to establish the validity of each profile element prior to using it as a means of selecting and/or placing individuals on the job. What assurance do we have, for example, that poor or unsatisfactory workers do not have a composite profile that looks exactly like that shown in Figure 3.11? We have none at all, really, unless we proceed to find out empirically what the unsatisfactory com-

posite does look like by actually measuring a group of these people on the same traits and computing group averages.

It should be obvious that only those predictors which demonstrate a significant difference in mean scores between the satisfactory and unsatisfactory groups should be included on the ideal profile. Any trait which does not clearly differentiate between "good" and "poor" employees will only add error and confusion by being interjected into the selection process. Since validation of each trait is a necessary (but all too often ignored) step in profile item selection, it might be a legitimate question to ask why not simply use all profile predictors in a multiple regression equation (or perhaps even a multiple cutoff). Actually, the answer to this depends upon what method is employed for comparing profiles, as will be seen in the following section.

METHODS OF COMPARING PROFILES There are two quite different procedures which can be adopted in comparing the profiles of each individual to the ideal profile. One method selects those people who have profiles that *match* the composite most closely. This in turn results in a choice of procedures, depending upon how the term *match* is defined.

One way of defining a good match is to say that the closer the points of one profile are to the points of the other profile, the better the match. This method, then, uses the differences between the two scores on each trait to obtain a measure of similarity (or dissimilarity). The most usual procedure computes these differences, squares them, and then adds them up to get a similarity measure. Thus, if we have a profile with k traits, and if we further define

X_{ij} = Score of person i on trait j
X_{sj} = Score of standard profile on trait j
then $D^2 = (X_{ij} - X_{sj})^2$

and $\sum_{1}^{k} D^2$ would represent the degree to which the profile of person i matched the standard profile. The larger ΣD^2, the poorer the match. It is important to realize that the D^2 method does not concern itself at all with whether or not the scores of person i fall above or below the composite, that is, direction is *not* important with this matching procedure. All that counts is the closeness of profile points.

A second method of defining profile similarity is expressed in terms of our old friend the correlation coefficient. A high correlation between the scores of individual i's profile and the scores of the ideal profile indicates that the two profiles have similar *patterns*, i.e., individual i scores high on those traits on which the ideal profile also has high scores and he scores low on those traits where the ideal profile also has low scores. Figure 3.12 shows examples of profiles which illustrate how the use of different methods for assessing similarity can result in different individuals being selected for the job. Examination of Figure 3.12 quickly reveals that the general *pattern* of person B's scores duplicates that of the ideal or standard profile much more closely than do the scores of person A. However, the actual scores obtained by person A appear to be *closer* in value to those of the standard profile than the scores on individual B's profile. We can, therefore, hypothesize that individual A ought to have the lower (more desirable) ΣD^2 score while individual B should have the higher correlation (more desirable) with the standard. As the data in Table 3.3 indicate, such indeed turns out to be the case. When the values given in Figure 3.12 are used to compute ΣD^2, person A's *score* (ΣD^2_{as}) is 500, while

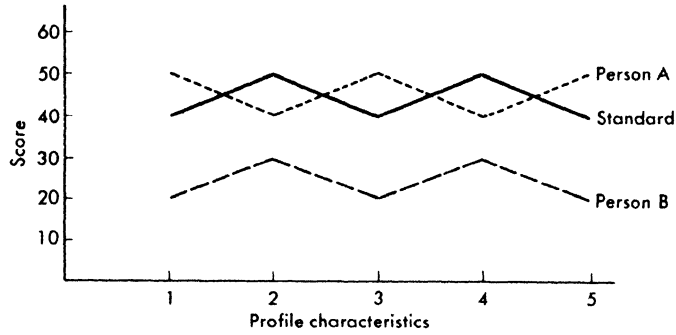


Figure 3.12. Example of effect of using different indices for profile matching. D^2 procedure would select person A, while correlation procedure would select person B.

person B's score (ΣD^2_{bs}) is much larger, having a value of 2000. On the other hand, when correlations between profiles are computed, the correlation between profile A and the standard profile is computed $r_{as} = -1.00$, while the correlation between profile B and the standard, r_{bs}, turns out to be 1.00—complete agreement. Thus, if the D^2 method were used as a selection criterion, we would select person A; if we were to use correlation between profiles as a method, we would select person B.[7]

CHOOSING A PROCEDURE Which procedure is best is a question that can only be answered by empirical means in a particular setting. In all likelihood, however, neither the D^2 nor the correlation method is the best technique. If the traits on the profile have been selected on the basis of having significantly discriminated between good and poor employees (as indeed they most certainly ought to be selected), then the logical deduction is that high scores on a trait are to be desired and low scorers are to be avoided (or vice-versa, depending on the trait). If we assume, as we have generally, that the significant relationship between each trait on the

TABLE 3.3 *D^2 Values and Correlations Between Profiles of Individuals A and B and Standard Profile Shown in Figure 3.12*

	Score						
Item	*Person A*	*Person B*	*Standard*	D_{A-S}	D_{B-S}	D^2_{A-S}	D^2_{B-S}
1	50	20	40	10	− 20	100	400
2	40	30	50	− 10	− 20	100	400
3	50	20	40	10	− 20	100	400
4	40	30	50	− 10	− 20	100	400
5	50	20	40	10	− 20	100	400
	230	120	220	10	− 100	500	2000

$r_{as} = -1.00$ $r_{bs} = 1.00$

[7] It is beyond the scope of this book to explore the problems of profile analysis in more than a cursory way. It is hoped that the reader will become aware of some of the major difficulties which arise when we attempt to classify or select applicants using profiles, and that he may be interested in exploring the topic further via some of the references suggested at the end of the chapter.

profile and job success is positive and linear, then we would want to select people according to one of the following procedures:

1. Select those people whose profile points tend to be the highest, i.e., their *average* profile score is used as a selection index. Using this procedure a person could have a large ΣD^2 score and still be selected, as long as his profile points tended to be above the corresponding profile points for the standard. This procedure is equivalent to using a multiple regression selection model where each profile trait is a predictor and the regression weights are assumed equal for each predictor.[8] Low profile scores on one trait can be compensated for by high profile scores on another trait.

2. Select those people who have profiles with the highest average profile score and whose points *all lie above* their corresponding ideal profile counterparts. This, of course, is equivalent to a combination of the multiple cutoff selection method and the multiple regression method. The ideal profile points are used to establish minimum acceptable score values. All people thus qualifying are then evaluated via the multiple regression system. Such a procedure can probably only work in cases where the selection ratio is sufficiently small to enable one to employ rather stringent cutoff values. Certainly to use the average score on each trait for a group of successful employees as minimum acceptable values is creating a stiff hurdle for new applicants.

Either of these latter procedures seems a somewhat more justifiable way to use profiles for selection than the first two procedures, D^2 or r. The concept of an "ideal" profile in which deviations in any direction are considered "bad" can be seriously questioned on logical grounds.

MULTIPLE HURDLE SYSTEM

Most selection situations involve attempts to predict later success on some task through the use of one or more predictive measures obtained at the time of job application. However, some selection situations such as management training include somewhat lengthy periods and ultimate evaluation after quite some time, but with interim evaluations or hurdles at various points of progress.

Consider the situation illustrated in Figure 3.13. Here we have diagramed a training program that might be used by a large corporation as a means of screening, training, and placing new college graduates within the corporation. The company initially hires a given number of college graduates, perhaps using college grades, interviews, letters of recommendation, and tests as a means of selecting people. All hires are told that their selection is on a probationary basis and that they will be continually evaluated during their training program. If performance during training is not satisfactory, they may be released from the program.

It is certainly in the company's interest to make an accurate decision about each individual as early as possible. Similarly, it is equally in the best interest of the employee that a decision be made as early as possible. However, the degree to which it is possible to predict success as a result of the training program increases in correctness (that is, validity increases) the longer we are able to observe the performance of the individual during training. By the end of the third evaluation period we should certainly be able to predict much more accurately whether a trainee is going to finish the course successfully than we were able to do at the time he was hired. The situation is quite analogous to the problem of predicting the final

[8] It would also be necessary that each predictor have an equal variance.

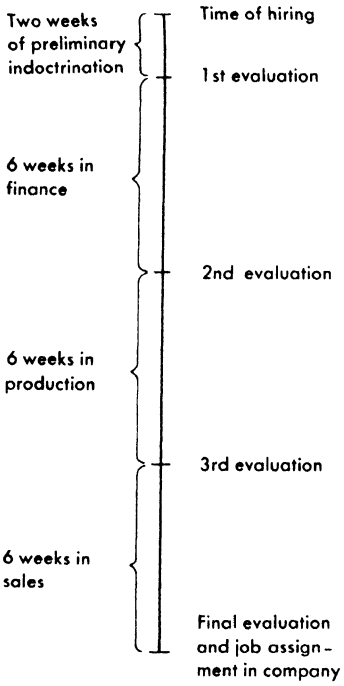


Figure 3.13. Example of a typical management training program used to train, evaluate, and place newly hired college graduates.

grades of college students. Obviously, one can make better predictions by the time the student is starting his senior year than at the time he enters college. Figure 3.14 illustrates the change in validity that one might logically expect to occur in a situation such as that diagramed in Figure 3.13.

In one sense the mechanics of a situation as shown in Figure 3.13 are identical to the more usual multiple predictor situations: A number of predictors of success are available, but to obtain each additional predictor it is necessary that additional time and money be invested into that trainee. Sequential predictors are used in several ways. Most frequently, one of the following methods is employed.

1. A person must score above some minimum desirable score at each evaluation stage. Thus, each stage becomes a hurdle which the trainee must clear if he is to be kept in the program.

2. A composite multiple regression is computed at each successive evaluation point, and the probability of success is computed for each person remaining in the program. Whenever this probability drops below some arbitrary value (for example, 25 percent), he is dropped from the program.

PROBLEM OF RESTRICTION OF RANGE One difficulty that emerges in sequential selection situations is a problem known as the effect of "restriction of range" upon validity estimates. If we have used predictor 1 to select people initially, and then if we subsequently compute the correlation between predictor 1 and the criterion or compute the correlation between some other predictor 2 and the criterion, our computed validity coefficients r_{1c} or r_{2c} are likely to be underestimates of the validity we would have obtained had *no* preselection taken place. By preselecting we have restricted the range of ability (and therefore predictor scores) which will reduce

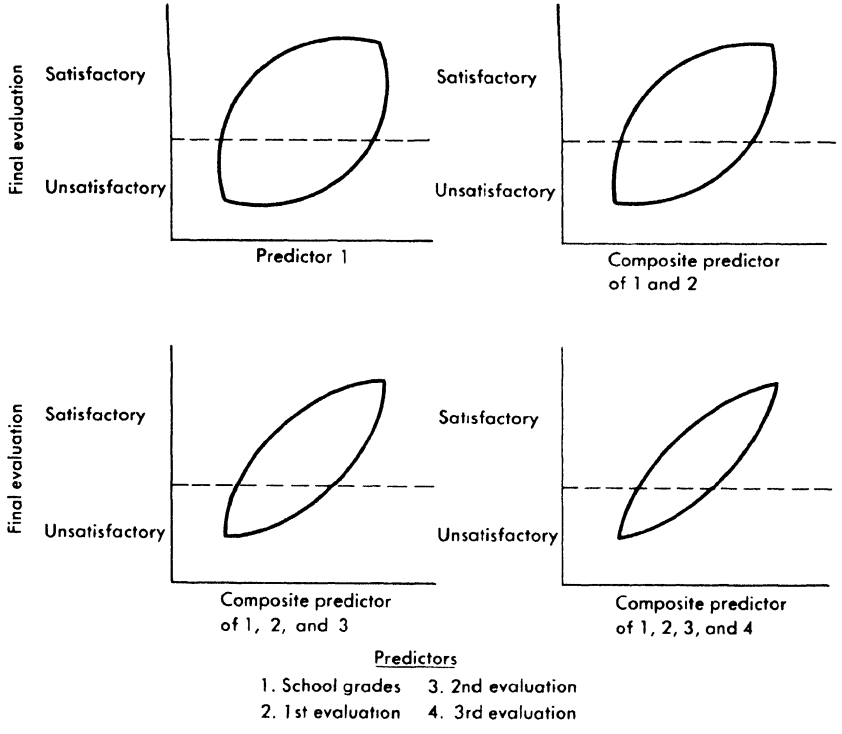


Figure 3.14. Diagrams showing the increase in validity to be expected as one obtains additional information about trainees as they progress through the different training phases shown in Figure 3.13.

the correlation coefficient. Indeed, our predictor 1 acts in a manner similar to the control variable in partial correlation; since it has already accounted for a portion of the variance, the correlation r_{2c} will be reduced. To get an estimate of what the validity R_{2c} truly is, one can use the correction formula.[9]

$$R_{2c} = \frac{r_{2c} + r_{12}\, r_{1c}\left(\dfrac{S_1^2}{s_1^2} - 1\right)}{\sqrt{\left[1 + r_{12}^2\left(\dfrac{S_1^2}{s_1^2} - 1\right)\right]\left[1 + r_{1c}^2\left(\dfrac{S_1^2}{s_1^2} - 1\right)\right]}} \tag{3.8}$$

where R_{2c} = corrected validity of predictor 2 for entire group of applicants
r_{2c} = computed validity for predictor 2 based upon restricted group
r_{1c} = validity of initial screening instrument, predictor 1
r_{12} = correlation between predictors 1 and 2
S_1^2 = variance of predictor 1 in original group
s_1^2 = variance of predictor 1 in restricted group

[9] For a more extensive discussion of this topic and its importance to selection and placement see R. L. Thorndike. *Personnel selection*. Wiley, New York, 1949.

MODERATORS

One of the more important concepts in selection and placement theory is the concept of the *moderator variable*. Sometimes referred to as a population control variable,[10] a moderator variable may be viewed as *any variable which, when varied systematically, has an effect upon the magnitude of the relationship between two or more other variables.*

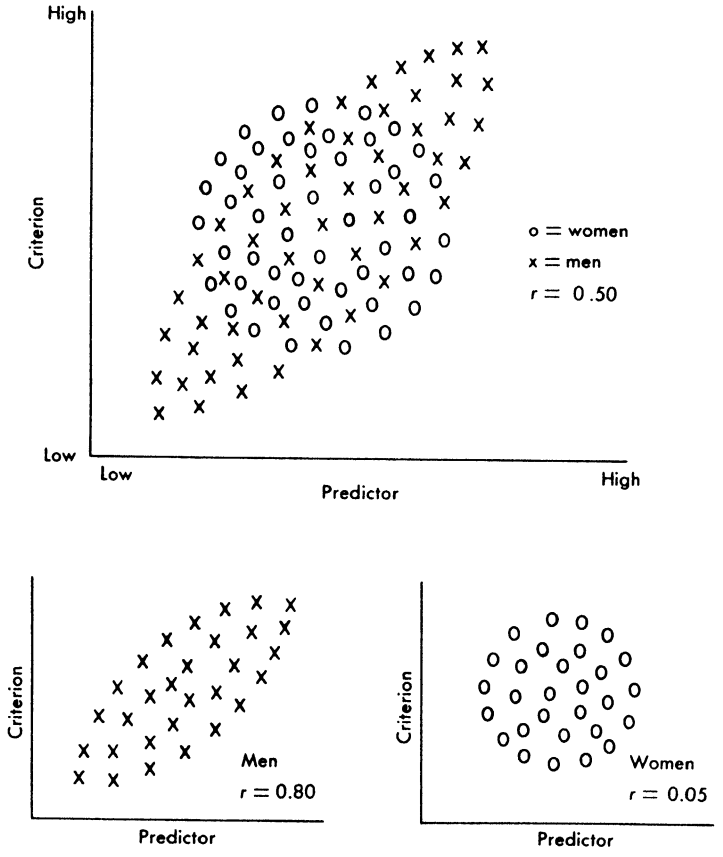


Figure 3.15. Diagrams illustrating the effect on validity which can occur through the utilization of a moderator variable such as sex.

Perhaps a hypothetical example (Figure 3.15) of how a moderator might function will serve to illustrate its influence upon the selection process. The top scatterplot illustrates a general validity of 0.50 between the predictor and a criterion. However, the "population" represented in the scatterplot is one which includes both sexes, viz., both men and women are grouped together in determining validity. Even a casual inspection of the top scatterplot indicates (if men and women are coded differently as has been done here) that the pattern of scores observed for men differs from that observed for women. To get a clearer picture of exactly how they

[10] R. H. Gaylord and J. B. Carroll. A general approach to the problem of the population control variable (abstract). *American Psychologist*, 1948, 3, 310.

differ, the two lower scatterplots in Figure 3.15 show the predictor-criterion relationships separately for men and for women. Now the difference is striking. For the men we observe a high positive relationship—one that produces a validity of 0.80. For the women, on the other hand, we see that there is virtually no relationship between the predictor and the criterion. The validity for women is 0.05.

The moderator variable in the above example is, of course, the variable of sex. The relationship between predictor and criterion is drastically affected by varying the moderator. The question "what is the validity of my predictor" clearly becomes more complex. What initially appeared to be a moderately respectable validity has now turned into two quite distinct and separate validities—one very high and one very low. One name for these latter validities might be *conditional validities*, that is, the validity of the predictor *given that the population consist of women* or *given that the population consists of men.*[11] An interesting characteristic of moderator variables is that a moderator need not have any direct relationship with either the predictor or the criterion variable (that is, r_{ym} and $r_{tm} = 0$).

EXAMPLES OF MODERATORS

Actual examples of moderators have been found in a number of research investigations. Vroom (1960), for instance, found quite marked moderator effects using degree of motivation of managers and first-line supervisors as the moderating variable. All men studied were employees in either the Chicago or New York plant of a national delivery service company which specialized in delivering small packages and parcels from department and other retail stores to private residences. Data from the study which best illustrate the moderator concept are given in Table 3.4.

TABLE 3.4 *Relationships Between a Measure of Nonverbal Reasoning and Ratings of Job Performance as a Function of the Motivation of the Worker*

		Correlations Between Nonverbal Reasoning Scores and Supervisory Ratings of			
	N	*Overall Performance*	*Summary Appraisal*	*Overall Results (Prod. Cost Qual.)*	*Overall Results (Skill, Methods, Effort)*
High motivation group	31	0.47**	0.56***	0.33**	0.21
Moderate motivation group	28	0.06	− 0.04	− 0.19	− 0.05
Low motivation group	32	− 0.07	− 0.23*	− 0.17	− 0.31**

* = Chance probability less than 5 times in 100
** = Chance probability less than 1 time in 100
*** = Chance probability less than 1 time in 1000

SOURCE: Adapted from Victor H. Vroom. *Some personality determinants of the effects of participation.* Prentice-Hall, Englewood Cliffs, N.J., © 1960, p. 55.

[11] The reader who has had some exposure to the Analysis of Variance and multiple regression analysis in statistics may recognize that the presence of a moderator would be indicated by an interaction between the moderator and predictor in the former, or by a significant cross-product term in the multiple regression equation.

All supervisors were divided into three groups based upon their assessed degree of motivation using a composite of several motivation indices obtained in the research. Validities for a test of nonverbal reasoning ability were then obtained for each of four different types of supervisory ratings of these men. This was done separately at each motivation level. As Table 3.4 shows, the test was apparently a quite valid predictor of how high a man would be rated by his supervisor *if only men with high motivation were considered.* If we systematically vary motivation by moving down to the groups having only moderate or low motivation levels, we see a corresponding systematic change in the relationship between the test and the criterion. The lower the motivation of the employee, the less the validity of the predictor; in fact, the validities even become negative for the low motivation groups.

Other examples of moderators can be found in studies by Dunnette and Kirchner (1960) and Ghiselli and his coworkers (1956, 1960). The work of Dunnette and Kirchner has been directed primarily at identifying job-related moderators, that is grouping people into jobs which are similar in terms of their responsibilities to get maximum prediction within each job group. Ghiselli's method might be called a "variable-free" moderator system: People are grouped simply on the basis of how well their success can be predicted with no direct reference to any external variable. Fredericksen and Gilbert (1960) have also done research on moderators to determine the degree to which a moderator's effect is likely to be consistent over time. They found that a moderator identified in a 1954 study (Fredericksen and Melville, 1954) was still operating in a 1960 follow-up.

MODERN VERSUS TRADITIONAL SELECTION THEORY

The concept of the moderator variable perhaps best illustrates the modern trend in selection and placement emphasis. Traditionally, selection and validation have been problems which were viewed as being best solved by simply establishing a criterion which appeared to be reliable and a predictor which could best predict that criterion. The emphasis was almost completely upon the establishment of a high validity with little or no thought toward exploring the many additional variables which, when varied, might add to or subtract from the obtained correlation. The general motto which all too often seemed to typify selection methodology was the slogan "If it works, use it!"

Without question, this policy was responsible for quite different developments in industrial psychology. First, it probably contributed to the degree to which psychologists were accepted into industry. Management is generally oriented toward positive results as represented by improved selection, and is not overly concerned with how it is accomplished. Unfortunately, however, this orientation is also probably responsible for the fact that validities in prediction have not risen substantially (if at all) during the past 50 years—a rather disturbing commentary upon the efforts of psychologists engaged in this type of work. In a 1955 review of a large number of validity studies, Ghiselli (1955) indicated that it is indeed an unusual event to obtain a validity coefficient of 0.50 or better. Figure 3.16 presents frequency distributions presented by Ghiselli of validity coefficients of varying magnitudes for different types of jobs. Note that only in the distribution of validities for clerical workers using intelligence tests as predictors and proficiency measures as criteria are there a large number of validities above 0.50.

The current interest in moderators is representative of a broader and somewhat

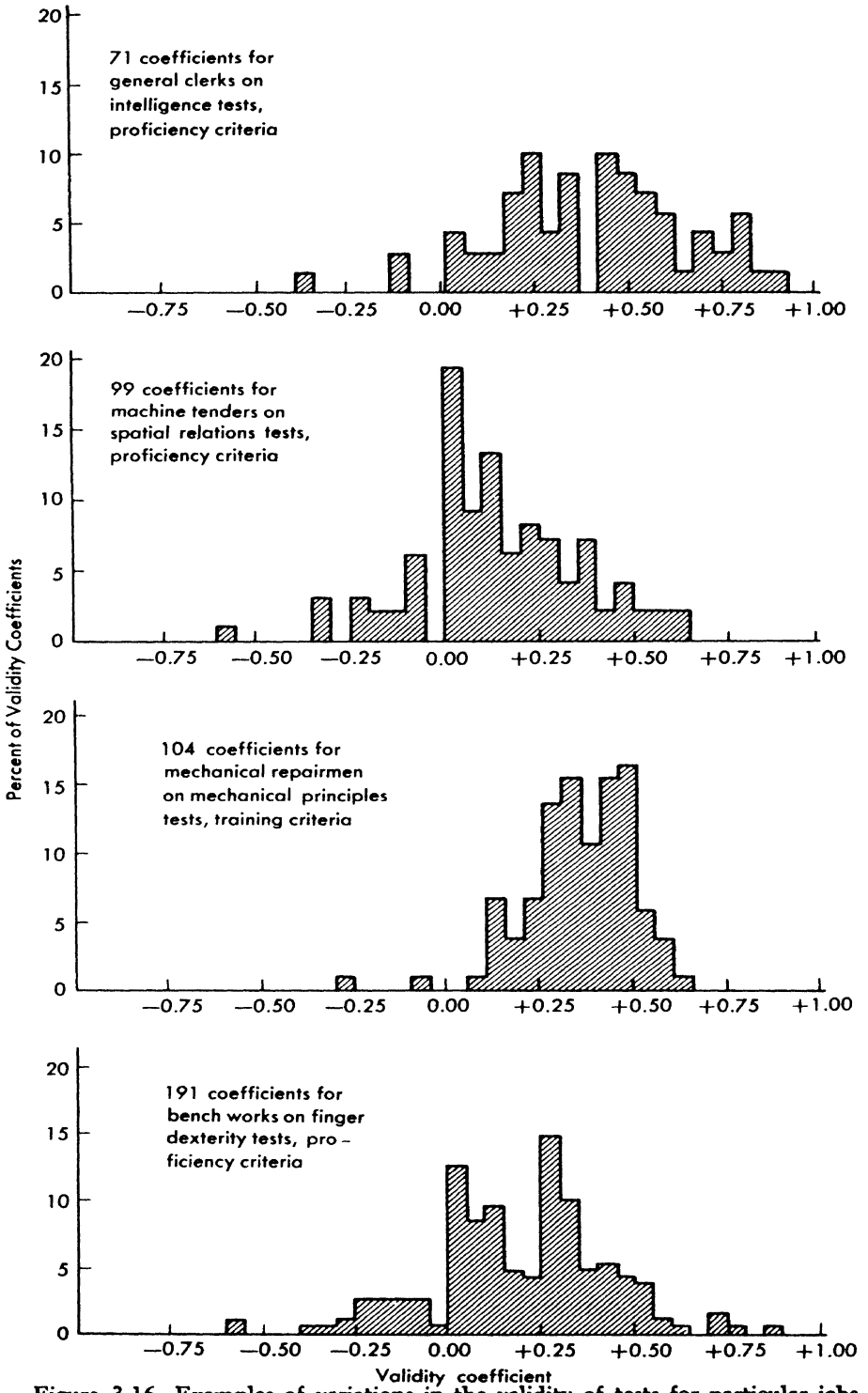


Figure 3.16. Examples of variations in the validity of tests for particular jobs. (From E. E. Ghiselli. *The measurement of occupational aptitude*. University of California Press, Berkeley, 1955.)

more sophisticated approach toward selection. It can be traced about to when Toops (1948) made an appeal for psychologists to consider the possibility that by stratifying people (for example, workers) systematically according to personal variables, one should be able to improve prediction. His method of classification, which he referred to as the *addend procedure,* is the forerunner of moderators.

DUNNETTE'S SELECTION MODEL

Perhaps the current view toward selection methodology can best be represented by the selection model proposed by Dunnette (1963). This model is shown in the diagram presented in Figure 3.17 and is designed to point out the maze of complexities and interrelationships which exist in the selection situation. The model may be viewed as more than an attempt to merely point out the dynamic nature

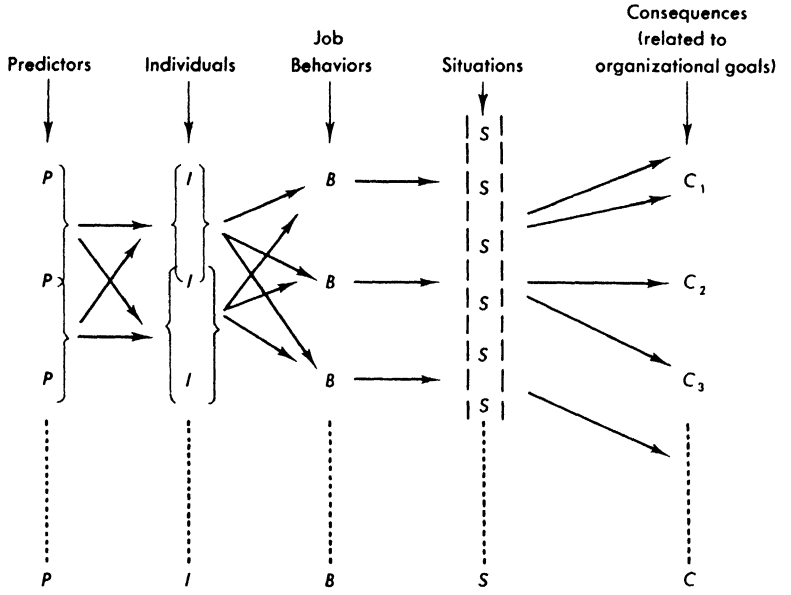


Figure 3.17. Modification of Dunnette's selection model, a diagram showing the complexities of any selection system. Every validation is subject to moderating effects of predictors, individuals, and situations. (Adapted from M. D. Dunnette. A modified model for selection research. *Journal of Applied Psychology,* 1963, 47, 317–323.)

of selection—it also represents a plea for psychologists to take advantage of these dynamics and use them to best advantage in order to improve predictability.

One can probably understand the point of view represented by the model in terms of the exact description used by Dunnette (1963, p. 318):

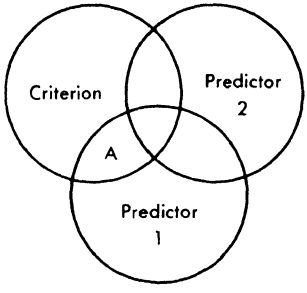
> Note that the modified prediction model takes account of the complex interactions which may occur between predictors and various predictor combinations, different groups (or types) of individuals, different behaviors on the job, and the consequences of these behaviors relative to the goals of the organization. The model permits the possibility of predictors being differentially useful for predicting the behaviors of different subsets of

individuals. Further, it shows that similar job behaviors may be predictable by quite different patterns of interaction between groupings of predictors and individuals or even that the same level of performance on predictors can lead to substantially different patterns of job behavior for different individuals. Finally, the model recognizes the annoying reality that the same or similar job behaviors can, after passing through the situational filter, lead to quite different organizational consequences.

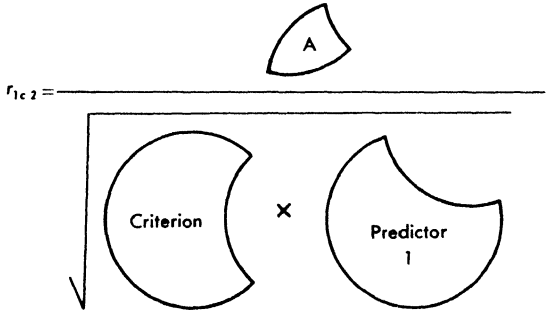
The current trend in selection represented by the awareness of moderators and by Dunnette's selection model should result in progress in both the increased efficiency of selection and the degree of understanding of the dynamics of accurate prediction.

SUPPRESSOR VARIABLES

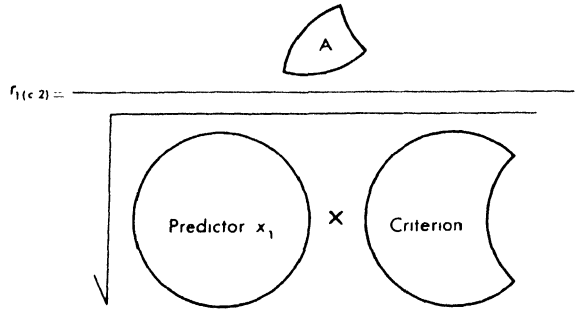
No discussion of selection would be complete without some mention of *suppressor variables*. In one sense a suppressor variable is similar to a moderator variable in that it is defined as "a variable which can have an effect upon the magnitude of a given predictor-criterion relationship even though it has little or no relationship to the criterion variable itself." The dynamics of a suppressor variable in prediction can best be understood by reviewing again the concept of a partial correlation and its related measure, the *semipartial* correlation. In Chapter 2 it was stated that if one had two predictors and a criterion that were intercorrelated as shown here



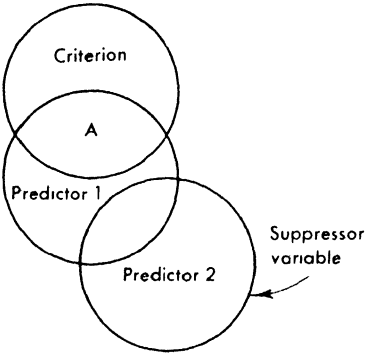
then the partial correlation between the criterion and predictor x, which is $r_{1c\,2}$, was defined as the correlation between x_1 and C after the effects of x_2 have been partialled out of both, so



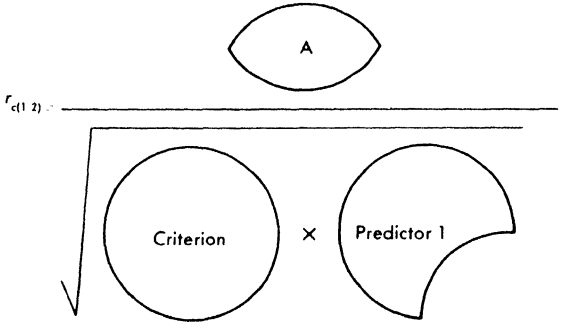
Suppose we only want to remove the effects of x_2 from the criterion prior to computing the correlation. Such a correlation is called a semipartial or part correlation. For example, we might be interested in the correlation between intelligence test scores (our predictor x_1) and final skill level at the end of a typing training program (the criterion). x_2 might represent the initial skill level of all employees in terms of their typing speed prior to taking the training course. Thus, we want to remove the effects of initial skill level upon final performance before computing the validity of our intelligence test. Our semipartial correlation now becomes



The mechanism of a suppressor variable is identical to that shown above except (1) generally, variable x_2 has only a slight (if any) relationship to the criterion, and (2) one is interested in removing its effects from predictor x_1. The general situation can therefore be diagramed as



and the "suppressed" correlation is



One cannot predict with complete certainty whether partial or semipartial correlations will be larger or smaller than the simple correlation existing between the variables, since the size of both the numerator and the denominator is affected by the partialling process. The only time this is *not* so is when the variable being partialled out is only related to one of the two other variables, as in the case of the suppressor. In such a situation only the denominator is subsequently affected (variance is removed) and the resulting semipartial correlation is larger than the simple unpartialled correlation between variables.

CROSS-VALIDATION

One feature of most multiple prediction selection systems is that in their development one typically tends to capitalize upon the chance variation that exists in the sample of employees being used for purposes of validation. This is particularly true with the multiple regression model, but applies as well to the multiple cutoff procedure. Because the multiple regression model has least-square properties, i.e., we deliberately minimize the errors in predicting *our particular sample*, it is likely that if we now apply our equation to a new sample (from the same population) we will not find our prediction as efficient as before. Thus, our computed R^2 is an overestimate of what the future validity of our prediction system is apt to be, since using our equation for purposes of prediction automatically implies applying it to new samples of workers. This expected drop in R^2 is known in statistics as the *shrinkage problem* and can best be illustrated by examining Figure 3.18.

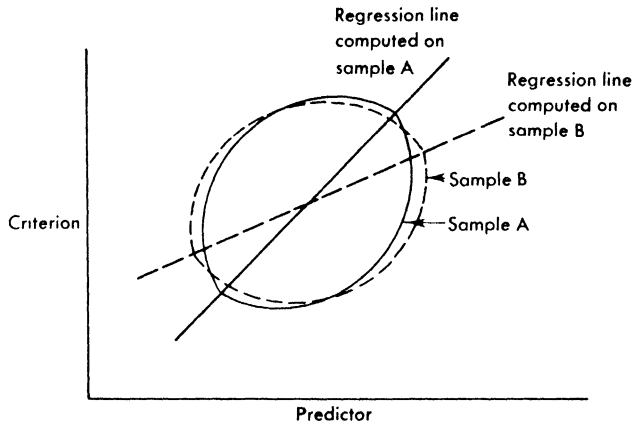


Figure 3.18. Regression lines of best fit for two different samples of workers belonging to the same overall "population."

In Figure 3.18 we have two samples of individuals. Each represents a random sample drawn from or belonging to the same population. For example, sample A might represent all job applicants for job X during the odd-numbered months, and sample B might represent all job applicants during the even-numbered months for a particular year. It would be highly unusual, even with very large numbers of applicants in each sample, for the two samples to be identical in terms of their scatterplots. Since their scatterplots can be expected to vary due to sampling error,

the correlation between the predictor and criterion (validity) can also be expected to vary somewhat, as can the regression equation computed on each sample.

Suppose we took the regression equation computed on sample A and used it to predict scores from sample B. We obviously could not do as good a job in minimizing Σd^2 using the A line with sample B as we could using the B regression line—after all, the B line by definition minimizes Σd^2 *for that sample*. Any other line will therefore have a larger error associated with it. Thus R^2 must be reduced correspondingly.

There are formulas available for estimating the amount of shrinkage one can expect when using this equation on a new sample. One such formula is

$$R_s^2 = 1 - \left[(1 - R^2) \frac{n - 1}{n - k - 1} \right] \tag{3.9}$$

where R_s^2 = shrunken multiple correlation squared
R^2 = multiple correlation squared obtained from validation sample
n = number of people in validation sample
k = number of predictors in regression equation

It is best, however, to cross-validate the equation by obtaining a second sample and trying it out to see how well it does predict. If there appears to be a very large drop, one may want to revise the equation (perhaps by combining both samples in one group). Large shrinkage is most often found when the sample sizes are small and/or the number of predictors is large relative to the sample size.

Mosier (1951) has discussed a number of *types* of cross-validation that can be conducted depending upon the design of the study and whether one is concerned about generalizing only to a new sample or if broader generalizations concerning the prediction equation are desired (for example, to different sexes, different criteria, etc.). The former is called a case of *validity generalization*; the latter is a case of *validity extension*. Of course, greater shrinkage would be expected in the latter case, and formula 3.9 only applies for cases of validity generalization.

REFERENCES

Catell, R. B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, *14*, 279–298.

Cronbach, L. J., and G. C. Gleser. Assessing similarity between profiles. *Psychological Bulletin*, 1953, *50*, 456–473.

Dunnette, M. D. A modified model for test validation and selection research. *Journal of Applied Psychology*, 1963, *47*, 317–323.

Dunnette, M. D., and W. K. Kirchner. Psychological test differences between industrial salesmen and retail salesmen. *Journal of Applied Psychology*, 1960, *44*, 121–125.

Fredericksen, N., and A. C. Gilbert. Replication of a study of differential validity. *Educational and Psychological Measurement*, 1960, *20*, 759–767.

Fredericksen, N., and S. D. Melville. Differential predictability in the use of test scores. *Educational and Psychological Measurement*, 1954, *14*, 647–656.

Ghiselli, E. E. *The measurement of occupational aptitude.* University of California Press, Berkeley, 1955.

Ghiselli, E. E. Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology,* 1956, 50, 374–378.

Ghiselli, E. E. The prediction of predictability. *Educational and Psychological Measurement,* 1960, 20, 3–8.

Horst, P. *The logic of personnel selection.* University of Washington, Seattle, 1961, 66.

Lord, F. M. Cutting scores and errors of measurement. *Psychometrika,* 1962, 27, 19–30.

Mosier, C. I. Symposium: The need and means of cross-validation. I. Problems and designs of cross-validation. *Educational and Psychological Measurement,* 1951, *11,* 5–11.

Nunnally, J. The analysis of profile data. *Psychological Bulletin,* 1962, 59, 311–319.

Sawrey, W. L., L. Keller, and J. J. Conger. An objective method of grouping profiles by distance function and its relationship to factor analysis. *Educational and Psychological Measurement,* 1960, 20, 651–674.

Toops, H. A. The use of addends in experimental control, social census, and managerial research. *Psychological Bulletin,* 1948, *45,* 41–74.

4 TESTING IN INDUSTRY

Industry generally recognizes that psychological testing is the function of the trained psychologist. Whereas many other areas of work such as incentive development, training, accident reduction, market research, and advertising effectiveness are studied by professionals from other disciplines, testing is most often reserved for the psychologist.

As stated in Chapter 1, tests were used in industry prior to 1910 by Hugo Münsterberg in connection with various problems in his research for the Boston Elevated Railway Company. Testing in industry has steadily increased since that time. During the two world wars, emphasis was placed on psychological tests, and the significant contributions made resulted in a tremendous impetus to the field during and immediately following these periods. Since 1945 the use of tests has become commonplace. In fact, as we will discuss a little later, the problem is often more of restricting the use of tests than of trying to promote their acceptance.

This chapter is intended to introduce the reader to the use of tests in industry by examining some of the advantages and disadvantages of psychological tests, by examining various types of tests which are currently in vogue, and by looking at some of the special problems which accompany the use of tests. Since testing is so widespread and the number of different tests used are so numerous, it should be pointed out that this chapter only introduces the student to the world of tests.

CHARACTERISTICS OF PSYCHOLOGICAL TESTS

Merely assembling a batch of questions does not produce a psychological test. A test may best be defined as measuring a *standardized sample of human behavior.* As such, it must meet certain basic requirements. First, the sample of behavior should be both large enough and representative enough of the class of behavior we are measuring to allow us to generalize and predict from the test results. Second, by requiring the test to be *standardized,* we mean that the series of questions or tasks should be uniformly administered whenever subjects are given the test. Finally, all psychological tests must have the prerequisite characteristics of validity, reliability, and norms.

Throwing together a series of questions regardless of whether they concern psychological topics does not elevate them to the level of psychological testing. For example, many newspaper columns and popular books of the parlor-game variety

present a series of questions; the person scores the answers and rates himself as "excellent," "good," "bad," or "awful." These are not psychological tests; they are merely batches of questions. Of course, an individual often has an overwhelming curiosity to find out something secretly, and so the self-testing plan is very popular. The subject matter of these tests usually has wide appeal. For example, a column may ask you to determine whether you are an outstanding husband by answering a few questions. A person may have inscrutable wisdom, but the only way he can develop a *test* to measure such qualities is to have a series of questions that have been answered in one way by successful husbands (whoever they are) and in a significantly different way by unsuccessful husbands.

TEST NORMS

Since the topics of reliability and validity were covered in detail in Chapter 2, no further explanation of their characteristics seems necessary here. However, the topic of test *norms* is one of considerable importance and needs additional clarification. Norms should be developed as a source of reference on all tests used in selection.

A norm is a standard of reference; it enables one to understand the meaning of a test score. Depending on the test, a raw score may be reported in various ways: Total time to complete the test, number of items correct, or number of items attempted are a few of the raw scores obtained on tests. Actually, the raw score by itself on a psychological test is usually a meaningless figure. For example, on one test a score of 240 seconds may be poor, whereas on another test a score of 75 seconds may be exceptional. The problem is further complicated when a score of 180 seconds must be compared with a score of 95 items correct. Without the use of a norm, such comparisons would be impossible; at best, they would resemble an attempt to compare apples and peaches.[1]

The two most widely used systems of norms are *percentiles* and *standard scores*. Both measures indirectly give information on the test performance of the individuals relative to a known population. They also show the relative position of one person in the group to the group as a whole. On a test to measure stenographic ability, an individual's raw score might be 105 words per minute with one error in transcription. Unless we had information about the range of speed based on many people, we could not know whether this rate of speed was good or poor. However, if this test has been given to 155 gainfully employed stenographers and this individual's score placed her in the 90th percentile, we would be in a position to estimate her ability to take shorthand as "exceptional" because she exceeds 9 out of 10 girls in shorthand speed. If this same individual types at a speed of 45 words per minute, and if this score is equivalent to the 20th percentile, we could conclude that 8 out of 10 employed stenographers are faster typists. Actually then, shorthand speed and typing speed can be compared. On the basis of the norms available, we can make a direct comparison of this girl's performance on these two tests, and our conclusion that she is very good in taking dictation but poor in typing is justified. Whether she is hired depends on the needs of the job; it is most probable that she would have to improve her typing speed before job placement would be possible.

[1] While norms are necessary to attach any real meaning to a test score, they are not necessary to the process of predicting some criterion of job success from the predictor scores.

For some clerical jobs speed is of great importance, but in others accuracy is more important. For example, speed rather than accuracy might be sought for addressing circulars. On the other hand, extreme accuracy would be necessary for filing important papers. Sometimes separate norms on speed and accuracy are necessary, as well as norms based upon a combination of the two measures.

In the establishment of norms, specific information should be available not only about the size of the group measured, but also about such facts as age level, whether the group is gainfully employed, whether it is a college population, and whether it consists of a normal unselected sampling. For example, a person who tests in the 70th percentile of a normal population on an intelligence test can be considered as having above average intelligence. However, his probable success in a Grade A college or graduate school would be questionable. In other words, a norm is meaningful only when the characteristics of the population upon which it is based are known.

TYPES OF TESTS

Psychological tests may be categorized into various groupings depending upon the particular purpose of the classification system. Before proceeding further, it might be valuable to consider some of these classification systems.

TYPE OF BEHAVIOR MEASURED

Probably the most frequently used classification system is based on the type of behavior that tests purport to measure. Thus we find tests designated as intelligence tests, personality tests, interest tests, vision tests, music tests, art tests, mechanical tests, verbal tests, etc., where each group is fairly behavior specific.

ACHIEVEMENT AND APTITUDE TESTS

Very frequently authors will distinguish between tests of *aptitude* and tests of *achievement*. The former is supposedly a measure of a person's potential in a given area, while the latter is a measure of a person's current skill or ability at the moment of testing. Since the same test can often be considered both an achievement test and an aptitude test depending upon use, this classification system is often a fuzzy one. Thus, with many tests one can (1) measure the amount of present skill, and (2) use the present score to predict future performance.

PAPER-AND-PENCIL TESTS AND PERFORMANCE TESTS

Many tests are of the *paper-and-pencil* variety: The testee simply receives a test paper or printed booklet containing the test questions, and he records his responses in some written manner on the answer sheet that is usually provided. Many other tests, however, do not require a written response—rather, they involve some sort of manipulation activity such as handling pegs or blocks, or assembling mechanical objects. These latter tests are referred to as *performance* tests.

SPEED AND POWER TESTS

Some tests are constructed so that every item is very easy—the task is to complete as many items as possible in a short time. When test performance is based primarily upon the speed with which one works, the test is referred to as a *speed* test. The other extreme would be a test where the items were difficult and the person was given as much time as necessary to complete the items. In such tests a person's score is based exclusively upon his ability to answer the questions correctly, no matter how long (within reason, of course) it takes. This type of test is called a *power* test.

INDIVIDUAL AND GROUP TESTS

There are a number of tests which are designed to be administered individually; that is, they cannot be given simultaneously to two or more people by a single examiner. An example would be the Stanford-Binet Intelligence Scale. Very often *individual* tests are used for clinical evaluation. *Group* tests are those which may be taken by many people at the same time. For industrial testing, group tests are generally preferred because they are more economical to administer.

LANGUAGE AND NONLANGUAGE TESTS

Sometimes it is important to distinguish between those tests which require a knowledge of a particular language (such as English) in order to understand either the test instructions or the test items themselves. All such tests are called *language* tests because performance on them depends partially upon the language ability of the testee, regardless of the type of ability the test is designed to measure. In some cases it is desirable or necessary to avoid the language bias of a test. For example, to test the mechanical ability of people who are illiterate using a test that has complicated written instructions would be quite inappropriate. To solve this problem, tests have been constructed in certain areas which are "language-free" tests: They require no language skill on the part of the testee. These are referred to as *nonlanguage* tests.

TESTS VERSUS OTHER SELECTION DEVICES

In this chapter our attention is devoted exclusively to examining psychological tests. However, tests are by no means the only predictive tool employed by the psychologist in a selection situation. Other standard selection aids are letters of recommendation, application blanks, and interviewing. Comprehensive selection programs will use these in addition to tests as part of the total selection process. In Chapter 5 these other methods will be discussed in greater detail.

PROS AND CONS OF TESTING

Psychological testing in industry should be approached with caution. Testing is often regarded as a fascinating subject by the novice. In many instances the desire to introduce a testing program is fostered in the hope that a solution will be

obtained to a "no solution" problem. Even when tests are used, it is advisable to conduct an interview, check the application form information, and use the various other techniques recommended by good personnel procedure.

The expected result of improving the selection of applicants is that training costs and labor turnover are decreased, production is increased, accidents and the probability of accidents are reduced, and morale may be increased. However, the mere introduction of a valid testing program cannot act like a magic wand to wave away all difficult problems. Testing is no panacea, although testing in industry can have value if used correctly. However, an improvement in job performances of 10 to 20 percent is often all that can be expected. Although better results are sometimes obtained, this is the exception rather than the rule.

Table 4.1 shows some data collected by Ghiselli (1955) on the efficiency of testing

TABLE 4.1 *Average Validity Coefficients for Various Types of Tests*

	Type of Criterion	
Type of Test	*Training*	*Job Proficiency*
Intellectual abilities		
Intelligence	38 *c*	19 *c*
Immediate memory	29 *c*	19 *c*
Substitution	26 *c*	21 *c*
Arithmetic	41 *c*	21 *c*
Spatial abilities		
Spatial relations	31 *c*	14 *c*
Location	24 *c*	15 *c*
Perception of details		
Number comparison	26 *c*	21 *c*
Name comparison	25 *c*	21 *c*
Cancellation	29 *c*	20 *c*
Pursuit	19 *c*	17 *c*
Perceptual speed	39 *c*	27 *c*
Mechanical comprehension		
Mechanical principles	34 *c*	26 *c*
Motor abilities		
Tracing	16 *c*	16 *c*
Tapping	12 *c*	14 *c*
Dotting	14 *c*	15 *c*
Finger dexterity	22 *c*	19 *c*
Hand dexterity	38 *a*	14 *c*
Arm dexterity	30 *b*	17 *c*
Personality traits		
Personality	16	21 *c*
Interest	14 *a*	27 *c*
Personal data	44 *c*	41 *c*

Less than 100 cases = no letter symbol
a = 100 to 499 cases
b = 500 to 999 cases
c = 1000 or more cases

SOURCE: E. E. Ghiselli. *The measurement of occupational aptitude.* University of California Press, Berkeley, 1955.

for various types of industrial tests. Ghiselli examined large numbers of validation studies and classified them in terms of the type of test being used and the type of criterion involved. All validities within a particular category were then averaged to give the figures shown in Table 4.1. Note that only three of the values in Table 4.1 exceed 0.40 in size, and that two variables are *not* tests but are personal history items. While the process of averaging used by Ghiselli definitely masks some very substantial validities, it should be sufficiently clear that testing is not a complete solution to the selection problem.

From the point of view of the statistician, such correlations are, of course, too low for individual prediction purposes. The relationship is nevertheless useful, for even such slight help in predicting success is better than if no tests were given, as was made clear in the earlier discussion of the selection ratio.

In the discussion of job satisfaction and morale in later chapters, we will find that many other factors, in addition to an individual's ability, help to determine performance. Compelling reasons for continuing on a job can compensate for an employee's limited ability. On the other hand, for many reasons a person may be a failure on a job even though he has the ability to be successful. Success or failure depends upon other things besides ability. Most industrial testing programs limit the functions to be measured to certain specific abilities and do not consider the other contributing factors. Furthermore, the claim cannot be made that even ability is perfectly measured.

It is no wonder that psychological tests in industry cannot do a perfect job at the present time. Anyone who expects perfection is unaware of the real situation. For example, a girl who worked in a certain watch factory had an abnormally poor score on a finger and tweezer dexterity test. In this instance better test scores were correlated with better job performance. This girl actually had a poor production record, and since salary was determined on a piece-rate basis, her earnings were very low. She was also rated as a poor worker by her supervisor. Nevertheless, she was among the workers who had the longest tenure of employment. An interview with her revealed that she was under extreme financial pressure and simply could not afford to be out of work one day. Moreover, experience had taught her that she did not last at jobs very long, except for her present one, and so she was content to stay on even though the psychological test predicted poor ability and her production records verified the prediction. An opposite example is offered by the man who made an excellent score in a battery of tests for salesmen. He possessed all the traits and abilities considered desirable but had worked out very poorly on the job. An interview revealed that he was married to a wealthy woman and did not need money; he had taken the job in order to acquire a cloak of respectability. The fact that he did not sell as much as the next fellow was of little concern to him.

The practical expediency of costs often determines the extent of the test battery. For the average job, the cost of testing would be excessive if the test period required more than twenty or thirty minutes. Since all aspects of ability cannot be measured in this time period, the effectiveness of the test battery is necessarily lowered. Psychological tests are useful but limited in scope. We repeat this over and over, not to defend or attack testing, but to give a clear and honest understanding of the possibilities. Testing is desirable, but care must be exercised that testing, when applied in industry, is appropriate.

For example, some years ago an employer in a certain factory was interested in introducing a testing program and consulted one of this text's authors. The em-

ployer complained of his high turnover, long training period, and many other difficult problems. He wanted to introduce a battery of psychological tests so that his selection of employees would be improved. He now believed that testing would solve all his problems. A subsequent review of the type of work done in the factory indicated that psychological tests might be of value. In the discussions it was found that his labor supply was rather limited. A high-school education, either general or academic, was considered a basic requirement for hiring. The question was raised as to why a high school diploma was necessary, and why graduates of a general rather than a vocational high school were considered desirable. Probing revealed that the policy of the firm was to avoid union organization. In line with this policy, the factory never hired experienced people because they might be union members, and graduates of vocational high schools were not hired because it was believed that they were directly or indirectly connected with the unions. The point of view in this text is neither for nor against unions, and it is not the intention to introduce extraneous material during a presentation of psychological testing. However, this illustration serves to indicate that this employer wanted to use psychological tests to be able to continue avoiding unionization and still improve the selection of the applicants. This was called to his attention and he was advised to reconsider his so-called basic qualifications, since there was an available supply of experienced people. But he was adamant; he wanted the battery of tests to select the inexperienced people he considered hiring.

In many situations a person does not know what to do either because the problem he faces is too difficult or because there may be more than one correct solution. Similar situations arise in connection with hiring practices, and in such cases an employer may grasp at the idea of psychological testing as a drowning man supposedly grasps at a straw. If two applicants are available for a job paying $30,000 a year and both appear to be exceptionally good, the executive officer may prefer to "pass the buck" so that he cannot be blamed if the one employed does not prove to be as successful as predicted. At such a time he will believe that psychological testing or even fortunetelling might be good; for once the decision has been made, it is impossible to know whether the other applicant would have been any better. Hence a psychologist may be called in to give some tests and recommend hiring one of the two applicants. The extent of help a psychologist can give in such a predicament is limited. At best, he can only analyze and interpret the test results of the two candidates and point out any differences the tests reveal. He may not know—and in most instances cannot guess—whether the tests are related to successful job performance.

SOME DANGERS IN TESTING

The use of tests in the competitive and profit-conscious world of modern industry results in certain problems. One should always be aware of and on guard for these problems.

OVERPROMOTION A great danger in psychological testing is the possibility of its being oversold to a prospective client. Many psychological testing organizations exist, and some make overly enthusiastic claims to potential clients. When these organizations employ professional psychologists, the likelihood of exaggerated claims is not too great; but when they employ people who are immature professionals or pseudopsychologists, there may be trouble. Psychological testing is a good thing,

will eventually be tried. However, tests and personal history data will probably continue to be the most effective predictors; and interviews and letters of recommendation, in spite of their problems and biases, will probably continue to be an integral part of the selection process.

REFERENCES

Albright, L. E., and J. R. Glennon. Personal history correlates of physical scientists' career aspirations. *Journal of Applied Psychology*, 1961, *45*, 281–284.

Anderson, C. W. The relation between speaking times and decision in the employment interview. *Journal of Applied Psychology*, 1960, *44*, 267–268

Anderson, R. C. The guided interview as an evaluative instrument *Journal of Educational Research*, 1954, *48*, 203–209.

Assessment of men. OSS Assessment Staff Rinehart, New York, 1948.

Bass, B. M Selecting personnel by observation *Personnel*, 1950, *26*, 269–272.

Bass, B. M Situational Tests I Individual interviews compared with leaderless group discussions. *Educational and Psychological Measurement*, 1951, *11*, 67–75.

Bellows, R. M., and M. F. Estep *Employment psychology The interview* Rinehart, New York, 1954.

Bingham, W. V. D., and B. V. Moore. *How to interview*, 3rd rev ed Harper & Row, New York, 1941.

Bolster, B. I., and B. M. Springbett. The reaction of interviewers to favorable and unfavorable information. *Journal of Applied Psychology*, 1961, *45*, 95–103.

Bonneau, L. R. An interview for selecting teachers. *Dissertation Abstracts*, 1957, *17*, 537–538.

Buel, William D Voluntary female clerical turnovers The concurrent and predictive validity of a weighted application blank *Journal of Applied Psychology*, June 1964, *48* (3), 180–182.

Campbell, J T., E. P. Prein, and L. G. Brailey. Prediction performance evaluations. *Personnel Psychology*, 1960, *13*, 435–440.

Chaney, Frederick B., and William A. Owens Life history antecedents of sales, research, and general engineering interest *Journal of Applied Psychology*, April 1964, *48* (2), 101–105.

Daniels, H. W., and J. L. Otis. A method for analyzing employment interviews. *Personnel Psychology*, 1950, *3*, 425–444.

Dunnette, M. D., W. K. Kirchner, J Erickson, and P Banas Predicting turnover among female office workers *Personnel Administration*, 1960, *23*, 45–50.

England, G. W. *Development and use of weighted application blanks* William C. Brown, Dubuque, Iowa, 1961, 55.

Fleishman, Edwin A., and J. Berniger One way to reduce office turnover. *Personnel*, 1960, *37*, 63–69.

Goheen, H. W., and J. N Mosel. Validity of the employment recommendation questionnaire· II—Comparison with field investigations. *Personnel Psychology*, 1959, *12*, 297–301.

Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology*, 1955, *68*, 409–416.

Harrell, T. W. The validity of biographical data items for food company salesmen. *Journal of Applied Psychology*, 1960, *44*, 31–33.

Himelstein, R., and T. L. Blaskovics. Prediction of an intermediate criterion of combat effectiveness with a biographical inventory. *Journal of Applied Psychology*, 1960, *44*, 166–168.

Hinrichs, J. R. Technical selection How to improve your batting average *Personnel*, 1960, *37*, 56–60.

ADVANTAGES OF TESTS IN INDUSTRY

The major advantage of psychological tests when used in industry is that the test is an objective and standardized behavior sample which lends itself well to statistical evaluation. Generally, it is easier to determine the value of a test than it is to evaluate other selection devices such as interviews, letters of recommendation, etc. Also, tests tend to be less subject to bias—particularly tests of aptitude and achievement. Tests of interest and personality (sometimes referred to as "inventories" rather than tests) may be more susceptible to either unconscious or deliberate bias on the part of the applicant under certain conditions. (This problem will be discussed in more detail in a later section.)

In most cases the cost of tests is not particularly high relative to the costs of other personnel evaluation aids, especially if the tests are of the paper-and-pencil rather than the performance type. A great deal of information about a person can usually be gathered in a relatively short period of time by using tests. Testing in industry is a useful aid in modern business and deserves serious consideration, despite the shortcomings mentioned earlier.

The great advantage in testing is that it can improve the selection process. The problem of hiring inexperienced people is faced by all firms to some degree. Such employees need many months of training. The cost of hiring plus the cost of training often total between $300 and $400 per employee, and in many cases may even total several thousands of dollars. If an employer expects to continue in a competitive business, such costs must be kept to a minimum. Psychological testing can reduce the costs of hiring people who will be successful, for its goal is measuring aptitude and predicting ultimate success among inexperienced applicants. However, as Brown and Ghiselli (1952) point out, it is not safe to assume that tests which best predict trainability will predict job proficiency equally well.

In an exhaustive review of all studies reporting the correlation between test performance and the two criteria of trainability and job proficiency, Brown and Ghiselli found only a slight tendency for a test that proves useful in predicting trainability also to be useful in predicting job proficiency, and vice versa. This means that factors important in learning a job may differ considerably from those that are important in maintaining proficiency on the job. After all, a bright person with certain skills may learn a job very quickly. However, if the job is not personally challenging or satisfying, he may terminate or perform in a manner considered unsatisfactory by his superiors.

Psychological tests can also be useful in the selection of experienced applicants. In many instances these tests serve as the only good check on the possibly exaggerated claims of an applicant. Some applicants successfully bluff their way through an application blank and an interview, but faking a test can be more difficult. Many girls who claim to have experience as power sewing machine operators base their claim on nonindustrial experience; there is a sewing machine in their home, and their mother has taught them to use it. They also claim to know the operations of the machine as well as various types of stitching. An interview cannot easily check such claims, but a standardized test can establish the accuracy as well as the speed of performance. Further, when norms are available the individual's ability can be judged more accurately than is possible when judgment is based solely upon an interview.

Psychological testing is also useful in selecting people with promotional ability, discovering reasons for job failure based upon personality traits, and even determining susceptibility to accidents, as we shall see in later sections of the text.

PUBLIC ATTITUDE TOWARD TESTING

It must be remembered that all people who take tests are human beings and must be considered as such. It is safe to say that the average individual, regardless of his previous success or failure with psychological or school tests, does not relish the idea of being tested. He is likely to be nervous and frightened, and to exhibit anxiety and manifest considerable tension. In giving a test to an individual, the objective is to obtain that person's most typical functioning. The behavior mentioned can interfere with normal functioning. However, the extent of interference can only be estimated. Chances are that the inexperienced examiner underestimates this interference to the same extent that the subject overestimates it; the happy medium is somewhere in between. The experienced examiner takes pains to develop rapport. He is frank and interested and explains the purpose of the test within the limits of the directions. He attempts to encourage relaxation or at least reduce the applicant's tension.

An individual displays fear of psychological tests for many reasons. He may feel that the test will be unfair and will therefore prevent him from getting the job. Also, he may be unable to estimate whether he is doing well or poorly. He often thinks that the test is an imposition to which he should not be subjected. He recognizes that the test is a challenge, and like most people he prefers not to be challenged in this manner. For all these reasons the attitudes of the person taking the test must be taken into consideration.

The employer who sponsors a testing program has still another point of view. Frequently his motives are not altruistic but are concerned with dollars and cents. For him the real question is how many dollars will be saved as a result of the test program. Naturally, the cost of the program can be determined. But few psychologists, regardless of their experience, are able to estimate the savings which may accrue with anything resembling reasonable accuracy unless they are provided with a means of converting criterion data to dollar signs. To make a reasonably accurate estimate, the psychologist would have to know the cost of the training program per individual, the profits before and after the introduction of the program, and other related figures. Most employers do not give this information to strangers, even psychologists; some do not even know.

Some employers are willing to introduce a testing program and pay the psychologist a fee based upon a certain percentage of the money saved. The ethics of such a proposal are questionable, and certainly the arrangement is unfair to the psychologist and should be avoided. After all, a surgeon is paid regardless of whether the patient lives or dies; in fact, he is often paid in advance. Also, as we said earlier, an employer sometimes grasps at psychological tests as a last straw to help him with a problem that psychological testing really cannot solve. Unless he is acquainted with the possibilities of the tests and warned that success is not certain, he may become annoyed, dissatisfied, and even adverse to psychological testing.

Another group of people whose attitudes toward psychological tests should be mentioned are union officials and members. It is safe to say that, up to the present time, unions have not been very sympathetic to psychological testing programs. Part of this reaction can be accounted for by the unions' mistrust of "scientific" management, for they erroneously consider psychological testing as a phase of "scientific" management. Business organizations which have union contracts are likely to encounter difficulties in introducing a test battery as a means of selecting employees. Correctly or incorrectly, the union figures that this is an attempt to

hedge on the various clauses in its contract that deal with the closed shop or general hiring practices. Many union executives believe that psychological testing serves only the interests of the employer; the fact that it advances the interests of the employee is open to debate. The unions' general disapproval of psychological testing programs must be reckoned with in a unionized industry.

Unfortunately, psychological testing has been subjected to more than its share of public criticism in recent years. It seems at times to be a favorite target for any group or person who needs a focal point to engender dissatisfaction. While the use of psychological tests in the public school system has borne the brunt of these attacks, industrial testing has received its share. Examples of such critics are Hoffman's *The Tyranny of Testing* (1962), Black's *They Shall Not Pass* (1963), Gross's *The Brain Watchers* (1962), and Whyte's *The Organization Man* (1956). The last two authors were particularly critical of using tests to determine the fate of an individual as a job applicant. Whyte has even gone so far as to provide "tips" on how to get the "kind of score" psychologists "want" with their tests.

There is no question that tests are being misused and abused in some instances —indeed, we have already given examples and will give more later. However, test users today tend to be a rather competent lot, and employers themselves are becoming more sophisticated. Most test publishing houses maintain certain standards concerning who may purchase their tests. While these standards are sometimes difficult to enforce, they do discourage the incompetents from getting involved in testing. The major difficulty generally occurs with those pseudoprofessionals who have developed their own test which they peddle to anyone interested in buying it as a panacea. The real tragedy is that there are enough gullible management people to make the charlatans' efforts prosper.

As an example of this gullibility, one of the authors received a visit one day from a plant manager and his personnel director. Both were from a city several hundred miles distant. The reason for their visit was that a friend of the personnel manager who resided in the author's city had sent him a 15-question test of introversion-extroversion which he had cut out of the daily paper. Although the personnel manager had misplaced it before he could show it to the plant manager, the latter was so interested in its potential that both men had travelled several hundred miles in an attempt to get more information about the test. They wanted to use the test to screen applicants for their industrial sales force. Appalling as this illustration is of the degree to which management is all too often naive about tests and their appropriate use, it is only one of many such incidents that the authors have experienced.

Stagner (1958) has experimentally demonstrated the degree to which personnel managers can be "duped" by fallacious tests. He gave a group of personnel managers a published personnel inventory, but instead of giving each man his obtained scores, a fake "personality profile" was returned to each man with thirteen personality traits circled in red as being descriptive of that manager. As shown in Table 4.2, all managers had the same thirteen traits circled. These items had been collected from horoscopes, dream books, etc., and the thirteen circled items were chosen randomly.

Each manager was then asked to read over the items marked for him and decide how accurate each one was, using a scale from (a) amazingly accurate to (e) almost entirely wrong. The results showed that half of the managers felt their profile was an amazingly accurate overall description of them, 40 percent thought it was a "rather good" description, and the remaining 10 percent said their profiles were just average descriptions of their personality. The results for each item are given

TABLE 4.2 *Evaluations of Items by 68 Personnel Managers When Presented as a "Personality Analysis"*

Item	*Judgment as to Accuracy of Item* Percent[a] Choosing				
	a[b]	*b*	*c*	*d*	*e*
1. You have a great need for other people to like and admire you.	39	46	13	1	1
4. You have a tendency to be critical of yourself.	46	36	15	3	0
5. You have a great deal of unused capacity which you have not turned to your advantage.	37	36	18	1	4
7. While you have some personality weaknesses, you are generally able to compensate for them.	34	55	9	0	0
9. Your sexual adjustment has presented problems for you.	15	16	16	33	19
10. Disciplined and self-controlled outside, you tend to be wornsome and insecure inside.	40	21	22	10	4
12. At times you have serious doubts as to whether you have made the right decision or done the right thing.	27	31	19	18	4
15. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations.	63	28	7	1	1
16. You pride yourself as an independent thinker and do not accept others' statements without satisfactory proof.	49	31	12	4	4
18. You have found it unwise to be too frank in revealing yourself to others.	31	37	22	6	4
20. At times you are extroverted, affable, sociable, while at other times you are introverted, wary, reserved.	43	25	18	9	5
21. Some of your aspirations tend to be pretty unrealistic.	12	16	22	43	7
23. Security is one of your major goals in life.	40	31	15	9	5

[a] Not all percentages add to 100% because of omissions by an occasional subject.

[b] Definitions of scale steps as follows: *a*. amazingly accurate, *b*. rather good, *c*. about half and half, *d*. more wrong than right, *e*. almost entirely wrong.

SOURCE: R. Stagner. The gullibility of personnel managers. *Personnel Psychology*, 1958, *11*, 347–352.

in Table 4.2. The moral of the story, of course, is that you can impress anyone with the wisdom of your ability to assess personality if you keep your statements on a generally favorable level—a fact fortune tellers have been exploiting for hundreds of years.

ETHICAL STANDARDS OF TESTING

To prevent, as much as possible, the misuse of psychological tests, the ethical code officially adopted by the American Psychological Association (1959) devotes considerable space to test distribution and use. Some of the major aspects of this code are as follows:

1. The sale and distribution of tests should be restricted to qualified users. Who

is qualified will vary, of course, with the type of test and the background requirements of the psychologist.

2. Test scores should only be released to those individuals qualified to interpret them.

3. A person should not be subjected to testing under false pretenses. For example, the psychologist in industry may be placed in a conflict situation between loyalty to the person being tested and loyalty to his employer. In many cases he must guarantee anonymity to the person taking the test in order to get his cooperation, only to find pressure put on him from management to release the scores of certain individuals who were tested. It is the responsibility of the psychologist to keep the testee clearly informed at all times of the purposes of the testing and the uses to which his scores will be put.

4. Tests should not be released for public use without adequate objective data to support their merit. If a test is being given for research (testing the test), this should be clearly indicated to all concerned.

5. Test manuals containing normative data and validity and reliability information should be made available for all tests. This information should be kept current.

6. Tests or parts of tests should *never* be published or presented in a public media such as newspapers, magazines, etc.

Besides the above, several additional standards might be stated.

7. All testing sessions should be carried out under standardized conditions with a qualified test examiner present.

8. One should never "prep" a person for a particular exam. To do so invalidates the test results, as the score received is no longer a true representation of the individual in comparison with those individuals who were not "helped."

INTRODUCING A TESTING PROGRAM

Chapter 2 discussed in some detail the basic steps necessary in the proper establishment of a selection program. Since a testing program is one type of selection program (remember that there are other kinds of predictors of job success), these steps bear repeating:

1. Examine the job in question.
2. Select a criterion.
   Select a predictor.
3. Measure job performance.
   Measure test performance.
4. Relate test performance to job performance.
5. Accept or reject the test as being of value.

The introduction of a testing program in a business organization demands that a research program accompany it. Attempts to select applicants with the aid of psychological tests based upon opinion and judgment rather than statistical facts are rarely, if ever, justified.

Before introducing a test battery, it is extremely advisable for employers to consult an industrial psychologist. He not only has training and background in tests but also has had experience with different testing programs. He knows that there is

no shortcut in the process, and that a testing program must be accompanied by research. There is a practical reason for research to parallel a testing program. A decision must be made as to which applicants will be recommended for hiring and which will be turned down. It is not at all safe to assume that the applicant with the highest score will be the best person for the job. For example, it has been unequivocally determined that in many occupations the people with the highest scores on an intelligence test are not necessarily the most successful employees; in many jobs scores above a certain point are conducive to inefficiency and high turnover. In many occupations the best employee, from the point of view of a prediction based upon an intelligence test score, is the one who ranks in the middle range. It is often necessary to establish a maximum test score above which it is unsafe to hire; a minimum score must also be established, below which it is unsafe to hire. In other words, test results must be compared with success on the job in such a manner as to determine adequately the statistical validity of that test as a selection instrument.

Another reason for employers to utilize the services of a psychologist is that some tests have misleading names: They can be used for selecting workers in occupations other than that implied by the name of the test. For example, the Minnesota Clerical Test is a name-checking and number-checking test, but it has been found useful in the selection of inspectors and also department store packers. In both these instances the usefulness of this test was greater than when it was used to select clerical workers.

It cannot be expected that a battery of psychological tests which has been used successfully in one company will give the same degree of success in another. The type of personnel in two neighboring plants manufacturing the same type of product may vary tremendously. If known to the applicants, such factors as working conditions, morale, and selection technique used may predispose the better applicants to seek jobs in one plant rather than the other. Moreover, age, sex, skin color, and religion as a basis of selection are often artificial factors and may vitiate the validity of the testing program.

Because of relatively recent school experience, young people are often more familiar with pencil-and-paper tests and as a result do better on them. Older people may be frightened at the very thought of writing, as a result not only of passive decay but also of certain unpleasant memories connected with school tests. If factory A hires young people and gives a pencil-and-paper test, such a test may work for that factory. Factory B may be just as successful from the point of view of production but may hire older people who are equally efficient. This factory will probably not be able to use the pencil-and-paper test which was successful in factory A. Thus the principle that test batteries should not be borrowed is important. Naturally, this does not mean that one should not attempt to benefit from the experience of others, for such information may provide useful leads.

Also, as mentioned in Chapter 3, one may obtain differential prediction with the same test for different subgroups of people on the same job (moderator effects). In a study on the selection of sewing machine operators, Blum (1943) devised a zigzag pattern which the subjects were required to follow, using an actual machine with a needle but no thread. They were also required to sew a zigzag pattern in between two lines.

Before this job sample could be claimed to be a valid test for selection, it was necessary to test the test. In this particular study it was found that this sewing task

was moderately helpful in selecting experienced operators, but was practically useless in selecting inexperienced operators, that is, the test was a valid predictor of success for experienced applicants but not for inexperienced ones.

PROBLEMS OF ADMINISTERING TESTS IN INDUSTRY

A very important reason why a psychologist experienced in psychological testing should be employed for a test program in industry lies in the nature of the problems that arise in giving tests. Applicants or employees who take tests are likely to be cautious. They need encouragement in order to give their most typical performance, and they must be given honorable guarantees. Regardless of their success on the job, employees worry about the test, for they feel that poor test performance may cost them their job. If the precaution of dispelling such ideas is not taken, they are likely to show antagonism and resentment, and the employer will have a serious morale problem on his hands. Much of the anxiety of employees can be relieved by statements to the effect that the testing program to be used in the company in the future will depend upon whether the established employees do well on the tests, and that if the employees do not do well it will mean that the tests are unsuitable for the company. It is also desirable, when possible, not to give the employer the exact test results in any individual case; knowledge that individual test results will not be revealed gives the employees further reassurance.

The attitude of the examiner while administering the tests is important. He must be friendly and show a personal interest in the subjects but still maintain the standardized conditions for the test. He must know how to cope with unanticipated responses during the testing—laughing fits, talking, and even cursing. Ability to handle these so that they cause a minimum of interference is a prerequisite of an experienced examiner. The person administering the test program not only should be trained in psychological testing techniques but also should spend an apprentice period with an experienced examiner. This is true regardless of whether he has a master's degree in psychology or not. Some years ago one author hired an outstanding graduate student to help administer an industrial testing program. The student phoned him that the tests could not be given because of a broken stopwatch. The author's hurried taxi ride with a spare stopwatch proved to be unnecessary, for the student's watch merely needed winding. In his many courses the young man had never been taught to anticipate an unwound stopwatch.

CURRENT TESTING PRACTICES

The development and use of tests in industry are widespread at the present time, and they promise to continue on an increasingly broad scale. The remainder of this chapter will deal with just how extensive industrial testing is and will give examples of several testing programs and specific tests used in industry.

EXTENSIVENESS OF TESTING

To list all the companies in the United States that use psychological tests would be a lengthy and impossible task—impossible because published test results of studies are often described without mention of the specific companies involved. Moreover,

the companies that pay for research sometimes regard the material as "top secret" or as their private property, and so their policy is to discourage reports in the literature. Such a policy often gives a company an advantage over its competitors, and from this point of view the policy is wise. But at the same time it constitutes an obstacle in the development and use of tests. An incomplete list of the companies which have used psychological testing can be made, however, and in some respects may be said to comprise an industrial *Who's Who*. This list has not been organized on the basis of either the type of company or the type of test problem because it was believed that a random order would best illustrate the wide and varied use of the tests. A random list of some of the companies and the tests used is as follows:

The Johnson & Johnson Company, well-known manufacturers of surgical bandages, have used tests to select foremen and clerical workers. The Vega Airplane Company and the Curtiss Wright Corporation use tests in their selection process. R. G. Le Tourneau, Inc., has developed a battery of tests to measure the primary mental abilities needed for job success with them. The Sun Oil Company has been interested in developing a battery which would be suitable for selecting people who could be trained to operate machine tools. The Lockheed Aircraft Corporation has had successful experience with personality, intelligence, and mechanical aptitude tests in the selection of many different types of employees. The Hawaiian Sugar Refining Corporation has found the use of tests for the promotion of employees successful. Aptitude tests as well as intelligence tests have been used by the A. C. Spark Plug Division of General Motors, the Woodward Governor Company, and Martin and Schwartz, Inc. The Harwood Manufacturing Company and the Kaiser Glove Company are two firms in the needle trades which have hired operators with the aid of tests. The American Aluminum Company selected a group of prospective foremen for training on the basis of psychological test scores. Apprentice toolmakers are hired by the Scovill Manufacturing Company with the aid of tests. The Milwaukee Electric Railway and Light Company uses a battery of many tests to select a category of employees known as "electrical trouble men." The Philadelphia Electric Company has found tests helpful in the selection of substation operators. Westinghouse Electric Company uses tests as part of its promotion program for clerical workers. The Atlantic Refining Company selects men for responsible positions, such as accountants and statisticians, with the aid of tests. Most of the insurance companies —among them Liberty National Life Insurance Company, Aetna Life Insurance Company, and Metropolitan Life Insurance Company—have found test batteries helpful in selecting agents.

The Harris Trust and Savings Bank of Chicago uses both personality and intelligence tests in its selection process. Considerable research has been done by the Kimberly-Clark Corporation; this concern has developed its own tests, such as the Kimberly-Clark packing and inspection test, which have been made available to the profession.

W. T. Grant hires clerical workers on the basis of tests. The Tremco Manufacturing Company has done much research on the selection of salesmen. R. H. Macy has used psychological tests for years in hiring employees for many of its departments. Procter & Gamble has found tests useful in selecting salesmen and sales managers. Aptitude tests have proved valuable to the Todd Company in hiring salesmen.

It is very difficult to obtain any really accurate set of figures indicating, in percentage terms, the frequency with which tests are employed in industry. One major

difficulty is that many firms may use tests only occasionally for specific openings which occur at infrequent intervals. Other firms may rely on some state or government agency such as the State Employment Service to do most of their testing and preliminary screening for them. Thus, one must really differentiate between formal and informal testing programs before attempting to determine the frequency with which industrial firms are using tests as a method of screening job applicants.

One such survey was carried out in 1959 by the *Industrial Relations News*, a trade magazine (1959). They surveyed 200 companies to see what percent used formal testing programs for job applicants. Sixty-five percent reported having such a program. Of those with a testing program, *all* agreed that their programs were highly useful and that they fully intended to continue them. When the nontest users were queried, 60 percent felt that testing could be useful, 2 percent felt testing was of little value, and 38 percent had no opinion. Thus, of the total sample only about 14 percent were not of the opinion that testing was of considerable value in an industrial setting.

SOURCES OF INFORMATION ABOUT TESTS

A major difficulty often encountered by a person wanting to install a testing program is where to find information about various kinds of tests—particularly information related to test use in industrial settings. There are a number of excellent sources for such information, and anyone who is apt to become involved in a testing program should be aware of these references.

Probably the best single source of test information is a series of *Mental Measurement Yearbooks* edited by Oscar K. Buros. Published periodically, Buros' books represent a continuing attempt to keep reference material related to tests as up to date as possible. They are a very complete source because they embody a *critical* survey of tests. They are not limited to industrial tests but include practically all work which has been done in the entire testing field. In each volume a group of contributors carefully selected by Buros reviews all the available material on practically all the psychological tests. For example, Buros' *Sixth Mental Measurements Yearbook* (1965) presents information on 1219 different tests and includes 795 original critical reviews.

Several extremely useful general textbooks are available which cover the general topic of psychological testing. These include Anastasi's *Psychological Testing* (1961), Cronbach's *Essentials of Psychological Testing* (1949), and Guion's *Personnel Testing* (1965). While they are not devoted exclusively to tests in industry, they present information on many of the tests currently being used in industry. In addition, they provide a great deal of general information on testing which is quite useful to a person involved in any sort of testing process.

A third source of readily available test information is provided by test publishers. While there are a large number of such sources, many of whom are quite small and who publish only one or two tests, the larger test publishers have catalogues which often are a convenient starting place for someone interested in establishing a test program.

The psychology journal *Personnel Psychology* has a section devoted exclusively to the exchange of information concerning industrial applications of various tests. It is called the *Validity Information Exchange* (VIE). Its purpose is to stimulate the publication of validity information in industry and to stimulate the use of such

information in guiding personnel actions. Every validity study is organized in the following succinct format for easy evaluation by the reader:

1. *Firm:* The sponsor for the study and the organization or firm within which the study was done (if not proprietary information).
2. *Problem and setting:* The problem to which the study was addressed is described, and the social, economic, organizational, and related elements in the setting for the study may be outlined.
3. *Job title and code:* A job title and job code are given. The title and code describe the job performed by the people who were studied. The code is taken from the United States Employment Service's *Dictionary of Occupational Titles* (Third Edition), Volumes I and II, U.S. Government Printing Office, 1965, or from supplements and new editions as they may become available.
4. *Job description:* A description supplementing the job description given in the *Dictionary of Occupational Titles* is given when necessary.
5. *Criteria:* The means by which job performance was observed, recorded, evaluated, and quantified are described.
6. *Sample:* The kind of people studied are described, such as by sex, age, educational level, job service, and marital status.
7. *Number of cases:* The number of people in the sample, the number of people in subsamples, and the uses of subsamples in the study are described.
8. *Predictors:* The kinds of data being investigated for their usefulness in guiding personnel actions are described. When tests are used as predictors they may be described by title and a reference number, such as "Miller Analogies Test (6:472)." This example refers to a description of the Miller Analogies Test found in review number 472 in Buros' *Sixth Mental Measurements Yearbook* which was described previously in this chapter. Other references may be made to the fifth, fourth, and third yearbooks, to subsequent issues of the yearbooks as they may appear, or to bibliographies such as Buros' *Tests in Print* by the Gryphon Press, Highland Park, N.J., 1961.
9. *Technique and results:* The methods used in evaluating the data are described and the results are summarized in statistical form.

Jones (1950) reviewed over 2100 references on employee selection and selected 427 studies as representing the "cream of the crop." The ten classifications of workers most frequently studied were: salesmen (75), clerical workers (60), teachers (49), assemblers (23), executives (23), inspectors (23), supervisors (21), typists (17), stenographers (14), and machinists (9). However, the important point of the article is that Jones found only eight reported studies which met the criteria of adequacy in both experimental design and report. So that the more serious student may refer to these studies, they are here listed:

Bellows, R. M. Studies of clerical workers, chap. 8. In W. H. Stead, C. L. Shartle, *et al. Occupational counseling techniques.* American Book, New York, 1940, 144–146 (study of coding clerks).

Blum, M., and B. Candee. The selection of department store packers and wrappers with the aid of certain psychological tests. *Journal of Applied Psychology,* 1941, *25,* 76–85.

Guilford, J. P., and A. L. Comrey. Prediction and proficiency of administrative personnel from personal history data. *Educational and Psychological Measurement,* 1948, *8,* 281–296.

Holliday, F. The relation between psychological test scores and subsequent proficiency of apprentices in the engineering industry. *Occupational Psychology,* 1943, *17,* 168–185.

Otis, J. L., O. L. Endler, and L. E. Kolbe. Data-analysis methods, chap. 7. In W. H. Stead, C. L. Shartle, *et al. Occupational counseling techniques.* American Book, New York, 1940, 113–136 (study of department store salespersons).

Rundquist, E. A., and R. H. Bittner. Using ratings to validate personnel instruments: A study in method. *Personnel Psychology*, 1948, 1, 163–183.

Sartain, A. Q. Relation between scores on certain standard tests and supervisory success in an aircraft factory. *Journal of Applied Psychology*, 1946, 30, 328–332.

Selover, R. B. The development and validation of a battery of tests for the selection of clerical workers. *American Psychologist*, 1948, 3, 291–292 (abstract) and personal communication.

Jones concluded by listing five requirements of a research report on employee selection using psychological tests:

1. Detailed job description, with each group treated separately.
2. Complete description of the sample: *N* (sufficiently large); what proportion of the total population this represents and how selected; factors involved in hiring such as age, length of time on the job (preferably with widely differing employees treated as separate groups), and total experience in jobs of similar nature; and use of two samples, with one an applicant group.
3. Exact test titles; period in the employment experience when the tests were administered; whether the tests were a factor in hiring; where the tests were given; under what conditions and incentives the tests were given; reliabilities of tests with comparable groups.
4. Detailed description of the criterion; length of time on the job when the criterion measure was applied (with widely differing employees treated as separate groups); reliability of the criterion; some discussion of the validity of the criterion selected; if ratings were used, some estimate of the amount of contact the rater had with the employee; if production records were used, the duration of the period and whether there were any unusual factors operating at that time.
5. Adequate statistical treatment, with assurance that the assumptions governing the use of the given measures were met, and an actual report of the numerical results together with an appropriate measure of significance.

Dorcus and Jones (1950) have published a volume of abstracts which is essentially the record of the 427 "cream of the crop" studies referred to in Jones's study. It includes over 200 occupations listed alphabetically from accountant to YMCA secretary, and reports such items as subjects, tests used, criterion, validity of results of the study, and reliability in the criterion. A sample abstract follows:

Surgent, L. V.: The use of aptitude tests in the selection of radio tube mounters. *Psychological Monographs*, 1947, *61* (283), p. 40.

1. Subjects: 233 radio tube mounters, female. Tests used as part of application procedure but results not used in selection.
2. Tests: (1) Minnesota Rate of Manipulation Test: Placing, (2) Minnesota Rate of Manipulation Test: Turning, (3) O'Connor Finger Dexterity Test, (4) Purdue Pegboard, (5) O'Connor Tweezer Dexterity test.
3. Criterion: a single over-all rating based on the pooled judgment of the supervisor of training and 2 instructors; 5-point scale; rating during training period.
4. Validity: biserial correlation between test scores and criterion (groups divided into satisfactory and unsatisfactory):

Test	*Correlation*	*t*
1	.56	6.56
2	.50	5.79
3	.48	5.60
4	.64	7.39
5	.59	6.81

Multiple correlation of test scores and criterion (optimum order of test addition):

Test	*Correlation*
4	.64
5	.72
1	.76
3	.76
2	.76

Follow-up of 35 of the 233 mounters: prediction of criterion scores on basis of a regression equation correlated with over-all rating by immediate supervisor on the job:
Two-test equation, involving Tests 4 and 5: .60
Three-test equation, involving Tests 4, 5, and 1: .43

5. Reliability: rank-order intercorrelations ($N = 51$) of 2 over-all fitness rankings of quality and quantity of work produced in training (latter 2 one week after first):

Items	*Correlation*
Quality and quantity	.66
Quality and over-all fitness rating	.74
Quantity and over-all fitness rating	.85

Note the similarity between the format used by Dorcus and Jones in their *Handbook* and that currently used by the VIE.

In summary, the preceding sources of test information provide an excellent introduction and information base relating to test data, standardization, and research.

SPECIFIC EXAMPLES OF TESTS

In the following pages examples are given of different types of tests used today in industry. No attempt at a complete listing can be made because of limited space. Rather, the intent is to provide the reader with a feel for the typical test, using as examples some of the more commonly used test instruments of each type.

MENTAL ABILITY TESTS

There are a number of intelligence tests which have either been adapted or specially developed for use in industry. The primary function of most of these tests is to serve as a preliminary screening or evaluation device without taking too much time. Most such tests are designed to be completed in 15 or 20 minutes, and the items included provide as much face validity as possible. Some of the more frequently used tests will be briefly discussed.

OTIS TESTS OF MENTAL ABILITY Probably the most widely used industrial screening tests are the *Otis Self-Administering Tests of Mental Ability* (1922–1929). These

consist of two series of four equivalent tests. The first series is designed for high school and beginning college individuals, while the second series is designed for grades four to nine. In industry these tests have been used for selecting applicants for almost every conceivable type job. The success of the tests, as one might suspect, varies greatly from situation to situation (Dorcus and Jones, 1950). They generally are more useful with lower-level jobs, since they do not discriminate well at upper ability levels (Anastasi, 1961).

WONDERLIC PERSONNEL TEST This test is an abridged adaptation of the Otis tests (higher series). It has a time limit of only 12 minutes, making it particularly attractive to industrial users. In spite of its shortened form, it correlates above 0.80 with the original Otis test. The term *personnel* has been used in the title to reduce the threatening characteristic of any test that appears to deal with intelligence. Since the items on the Wonderlic are those which were found to differentiate between good and poor workers on a variety of industrial jobs, it is not surprising that the test often is found to be a better predictor than the Otis, in spite of its shortness. Excellent normative data are presented in the Wonderlic manual. As with the Otis, its validity varies greatly depending upon the situation, but its highest validities have occurred in the selection of workers for clerical jobs. Parallel-form reliabilities of 0.82 to 0.94 have been reported (Wonderlic and Hovland, 1939).

THURSTONE TEST OF MENTAL ALERTNESS. One of the interesting features of this test is that it provides two subscores, verbal (V) and quantitative (Q), plus a total score (Thurstone and Thurstone, 1943). The test has a time limit of 20 minutes, and consists of 126 verbal and quantitative items in alternate order and ascending difficulty. Reliabilities near 0.90 have been reported. It is available in two forms and appears to have its greatest validity in sales and clerical jobs.

THE ADAPTABILITY TEST One of the few selection instruments which was developed specifically for industrial research, this test has a 15-minute time limit and also has been named to avoid a negative reaction by applicants to having their mental ability measured (Tiffin and Lawshe, 1942). It provides good face validity. It has been used primarily with clerical workers and first-line supervisors.

MECHANICAL APTITUDE TESTS

Tests dealing with mechanical aptitude can be classified into two subgroups: measures of mechanical reasoning and/or information, and measures of spatial relations. The reason the latter tests are included is that many mechanical tasks involve the rapid manipulation of pieces and parts relative to one another. Thus the ability to perceive geometric relationships between physical objects and to be dexterous in the manipulation of such objects is generally perceived as being part of the general complex of mechanical aptitude.

MINNESOTA SPATIAL RELATIONS TEST This was designed to test both dexterity and spatial relations (Paterson, 1930). A picture of the test is given in Figure 4.1. As the picture indicates, the test consists of a series of 58 geometric shapes which

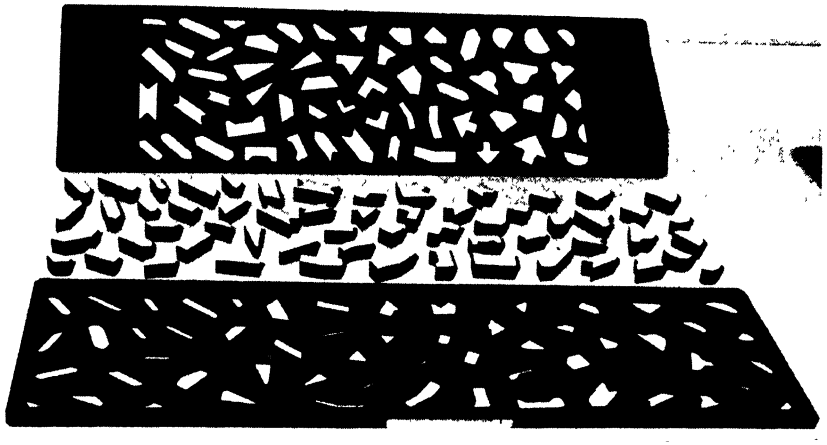


Figure 4 1. Minnesota spatial relations test. (Courtesy Educational Test Bureau Division, American Guidance Service, Inc., 420 Washington Ave S E., Minneapolis, Minn)

have been cut from a large board. The task is to place the cutouts into their appropriate slots as fast as possible. Both amount of time and number of errors are scored. Four boards are used (A, B, C, and D) with the same set of blocks fitting boards A and B and another set fitting boards C and D. The first board serves as a warm-up for the following three boards which are used to obtain the person's score.

MINNESOTA PAPER FORM BOARD TEST This test is a paper-and-pencil instrument, thus eliminating dexterity as a variable in performance (Likert and Quasha, 1941–1948). It consists of 64 multiple-choice items. Each item presents the parts of a geometrical figure "cut up" into pieces, followed by five assembled geometric forms. The task is to select the form which represents what would be obtained if the cutup parts were assembled. Reliabilities for this test are in the mid-eighties. Ghiselli (1942) obtained a correlation of 0.58 between test scores and the success of inspector-packers. The test does *not* appear to be highly correlated with either intelligence tests or with other mechanical aptitude tests.

MINNESOTA MECHANICAL ASSEMBLY TEST This test is a revision of an earlier test (Paterson, 1930). It is designed as a test of mechanical knowledge and reasoning, and also involves dexterity and spatial relations ability. The test is of the work-sample type, wherein a person is given a series of disassembled mechanical contrivances (e.g. a clothespin, a doorbell, etc.) and asked to reassemble them. The total test comes in three boxes, with 11 objects in each box. The entire test takes nearly one hour. Reliabilities range from 0.72 to 0.94.

BENNETT TEST OF MECHANICAL COMPREHENSION This is one of the most widely used of all tests of mechanical knowledge (Bennett, 1940). A paper-and-pencil test, it uses pictures to present the testee with various questions about mechanical facts and principles. Separate forms are available for trade school applicants, engineering applicants, and women. Reported reliabilities range from 0.77 to 0.84.

TESTS OF SENSORY ABILITY

Success in certain types of jobs is often found to be related to such physical attributes as visual acuity, color vision, and sensitivity of hearing. Several standardized tests or test instruments are currently available for evaluating the degree to which a person possesses each of the above abilities. Such tests are commonly referred to as tests of sensory ability.

TESTS OF VISION Vision is a rather complex ability consisting of many different types of visual skills. Thus we can distinguish between the ability to see things far away (far acuity) and the ability to see things close to us (near acuity). People also differ in terms of how well they can perceive relative distances (depth perception) and how well they can differentiate colors (color vision). Any or all of these abilities may be found to be important for job success.

Probably the best-known test of far visual acuity is the *Snellen Chart*, a display consisting of rows of letters which gradually diminish in size from top to bottom of the display. The person is usually placed at a distance of 20 feet from the chart and asked to read successive lines until he is no longer able to distinguish the letters. A ratio is then computed which expresses his acuity as follows:

$$\text{Acuity} = \frac{\text{distance at which person } i \text{ can read the line of print (usually 20 feet)}}{\text{distance at which the average person can identify the same line}}$$

Therefore, if the smallest line which person *i* can read at 20 feet is a line which most people can read clearly at 80 feet, his score is 20/80. Normal vision is, of course, 20/20.

The Snellen Chart is generally considered to be only an approximate measure of a person's total visual ability. Several general purpose acuity instruments are available commercially which measure a variety of visual skills. The most widely used instruments of this type are the Ortho-Rater (Bausch and Lomb), the Sight-Screener (American Optical Company), and the Telebinocular (Keystone View Company). All three instruments are general purpose in that each provides measures of depth perception, near and far acuity, phorias (a phoria exists when the eyes fail to converge naturally to an appropriate focal point distant from the individual) and color discrimination.

Probably the best-known test of color blindness is the *Ishihari* test. It consists of a series of cards or plates which have designs comprised of colored dots. The designs are constructed so that the background dots consist of one color while the design is made up of another color. The colors were deliberately selected so that a person with normal color vision will perceive the correct figure (usually a number),

while the color-blind person will perceive either no figure or an incorrect design. Appropriate responses to each plate for normal vision, red blindness, green blindness, and total color blindness are given in the test manual.

TESTS OF HEARING Auditory acuity, like visual acuity, has many dimensions. However, the most common dimension is simply the degree to which a person is sensitive to sounds of different pitch. The usual instrument employed to measure auditory sensitivity is the *audiometer.* This generates pure tones of different frequencies which can be varied in terms of their loudness by the examiner. The individual being tested listens for the tone, which is gradually raised in intensity. When he hears the note he signals to the examiner who records the intensity. Usually on half the trials the examiner starts with a loud tone and decreases the intensity until the testee signals he can no longer hear it, while on the other half he starts below the audible point. The average of a series of such trials is taken as the person's *threshold* for that frequency. Figure 4.2 shows the general relationship that exists between the

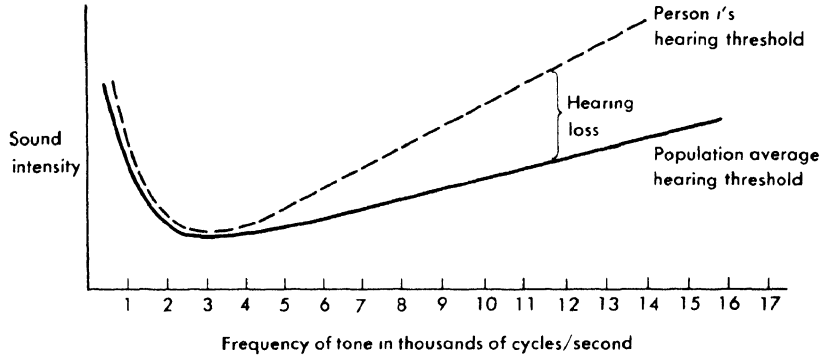


Figure 4.2. General functional relationship existing between the pitch of a tone and the auditory threshold (amount of sound intensity necessary for that tone to be audible) for the population average. An acuity curve for a person with a hearing loss at the upper frequency levels is shown for comparison.

frequency of a tone and the threshold for that tone. Notice that we are most sensitive to sounds in the two-to-three-thousand cycles/second range. For comparison, a hearing curve for a person having a substantial hearing loss in the upper frequency levels is also included in Figure 4.2. The amount of loss at any particular frequency is the distance between the population average and the individual's own curve.

Using the audiometer it is possible to prepare a chart, known as an *audiogram,* which presents a clear profile of a person's auditory acuity. One such audiogram or profile is prepared for each ear.

TESTS OF MOTOR ABILITY

Quite a variety of standardized performance tests are available for measuring fine and/or gross motor coordination. Some of the frequently used tests are listed below.

PURDUE PEGBOARD This timed test requires pins to be placed into small holes in the pegboard, using first the right hand, then the left hand, and finally both hands together (Tiffin, 1941). In the second part of the test, the pins must again be

placed in the holes but this time collars and washers must be assembled with the pins as they are placed (both hands are required for this part of the test). The two parts of the test are designed to measure manual dexterity and fine finger dexterity respectively.

CRAWFORD SMALL PARTS DEXTERITY TEST There are two parts to this timed test (Crawford and Crawford, 1946). In part 1, pins are placed in holes with a tweezer and then metal collars are placed over the pins. In part 2, a screwdriver is used to screw down small screws after they have been placed by hand into threaded holes.

MINNESOTA RATE OF MANIPULATION TEST Like the Crawford Test, this test consists of two discrete parts. In the first part the task is to place 58 round blocks into a board having an equal number of holes. The blocks are 1½ inches in diameter and are slightly thicker than the holes are deep, i.e., they still protrude after being correctly placed. The first part is referred to as the "placing" test. The second part consists of having the subject turn all the blocks over and replace them in their holes. This is called the "turning" test. A person's score is the time required to complete each task. An interesting aspect is that the two parts do not correlate very highly. Correlation between part scores is generally in the order of 0.50.

INTEREST INVENTORIES

The likelihood of success in an occupation in which one is interested should be greater than the likelihood of success in an occupation which a person finds not particularly interesting. For this reason it would seem sensible to obtain some measure of a person's interests prior to placing him on a job or suggesting any particular vocation. Of course, the most obvious way to ascertain a person's interest is simply to ask him what he is interested in. However, for a variety of reasons, the direct approach has proven to lack the necessary reliability. All too often people lack sufficient knowledge about specific vocational choices to make even an intelligent estimate as to whether or not they would really find it an interesting career. Instead, they can merely express their interest in terms of gross estimates such as "That sounds like it might be fun, from what little I know about it."

For this reason, interest inventories have been developed over the years to accomplish the assessment in a more subtle and sophisticated manner. The development of most such inventories has followed the same basic pattern with only minor variations on the overall general theme. The procedure for developing such tests might be as follows:

1. Prepare a variety of items dealing with how well one likes all kinds of activities, ideas, objects, and types of people. Some typical items might be:

I like to go swimming.
I enjoy going to concerts.
People who talk a lot are fun to be with.
Writing school reports is interesting work.

2. The items are then given to people in different occupations or occupational groups, and people in these occupations are asked to indicate whether these items describe *them*, that is, do they agree or disagree with the statement.

3. *Keys* are then developed for each occupation using those items which statistically discriminate that occupation from all others, that is, people in that occupation respond to it significantly more (or less) often.
4. These keys can then be used to get an interest score for each occupation for a person taking the test. To the extent that a person checks those items on a particular key, he is considered to have interests similar to people already in that occupation. The implicit assumption is that the more similar the interest pattern, the more apt one is to succeed in that occupation.

STRONG VOCATIONAL INTEREST BLANK (SVIB) Developed by E. K. Strong, Jr., the SVIB is made up of 400 items divided into eight different parts (1938). There are 47 different occupational keys available for use with the men's form of the test and 28 keys for scoring the women's form. Most of the keys are for higher-level occupations. (Relatively little success has been achieved in deriving keys for low-level jobs.) In addition to the keys for the various occupations, the SVIB also provides a number of scoring keys which can be used to evaluate interest maturity, occupational level, masculinity, and specialization level. The latter scale is relevant for today's highly complex job world, since it contains items which have been found to differentiate between the interests of men who have specialized in a given field and those who have remained "general practitioners." The SVIB scales have reliabilities in the mid-eighties.

KUDER PREFERENCE RECORD A somewhat newer interest inventory is the Kuder Preference Record. Its two most frequently used forms are the Vocational Form (Kuder, 1934) and the Occupational Form (Kuder, 1956). The former is made up of a number of forced-choice alternatives in which the person must select from three listed activities the one he would like most and the one he would like the least. The scales on the Vocational Form include outdoor, computational, mechanical, persuasive, scientific, artistic, literary, musical, social service, and clerical. The Occupational Form was constructed in a manner similar to the SVIB and has 38 different occupational keys.

PERSONALITY INVENTORIES

More controversy has surrounded the use of personality measures as predictors of job success than with any other type of test. As a rule, the validity of such instruments has tended to be somewhat less than impressive, although at times rather substantial predictability has been obtained. Numerous reasons have been suggested to account for the general lack of success in industrial situations of personality inventories, such as lack of reliability, inappropriateness of context, and susceptibility to faking. Even those tests developed specifically to counter these objections have failed to produce validities of the magnitude one might wish. The authors are of the opinion that "personality" contributes to a significant degree in determining the job success of many individuals, particularly in higher-level jobs. Nevertheless, successful personality assessment for a specific job is the exception rather than the rule.

Measures of personality may be conveniently classified into either self-reporting questionnaires or projective tests. Self-report inventories generally involve responses to a question in terms such as how much the testee "agrees" with the question, how

well it "describes" him, or which of *k* alternatives he "prefers most." The essence of the response is that of self-description. Typical test items are:

Do you worry over possible misfortunes?
Does discipline make you discontented?
At a reception or tea do you feel reluctant to meet the most important person present?
Do you often worry about your health?

Projective tests, on the other hand, usually involve obtaining an unstructured or "free response" from the individual. For example, the test may provide an incomplete sentence and asking the testee to finish it in any way he wishes. Another major difference between self-description inventories and projective measure is that the scoring and interpretation of the latter are much more complex and require much more training. Examples of both types of personality measures are described in the following sections.

GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY This test is a combination of three earlier personality inventories (Guilford and Zimmerman, 1949). It was developed using a statistical procedure known as factor analysis, a technique which groups things (in this case test items) into homogenous clusters which are usually relatively independent of one another.[2] The Guilford-Zimmerman gives a score on ten different factors. Each trait or personality factor is based upon thirty different items, and each trait is supposedly independent of the other traits. The traits are as follows:

G.	General Activity	O.	Objectivity
R.	Restraint	F.	Friendliness
A.	Ascendance	T.	Thoughtfulness
S.	Sociability	P.	Personal Relations
E.	Emotional Stability	M.	Masculinity

The reliability of the different scales range from 0.75 to 0.85.

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI) Probably the most widely known self-description inventory is the Minnesota Multiphasic Personality Inventory (MMPI). It consists of 550 affirmative statements which the person taking the test must classify as *true*, *false*, or *cannot say* (Hathaway and McKinley, 1943). Using standard item analysis procedures in which a control group of normals had to respond significantly more (or less) frequently to an item than a corresponding group of clinical patients in order for that item to be considered discriminating, ten different "clinical scales" have been developed as follows:

Hs.	Hypochondriosis	Pa.	Paranoia
D.	Depression	Pt.	Psychesthenia
Hy.	Hysteria	Sc.	Schizophrenia
Pd.	Psychopathic deviate	Ha.	Hypomania
Mf.	Masculinity-femininity	Si.	Social introversion

[2] For example, suppose one had an intelligence test partly made up of vocabulary items and partly of arithmetic items. A factor analysis of the relationship between these items would probably indicate two distinct factors or clusters. This would indicate that the verbal items have a lot in common, as do the arithmetic items, but the verbal items have little in common with the arithmetic items.

In addition to these scales, the MMPI also provides four "validity" scales which can be used to indicate the extent to which the respondent understood what was required of him and cooperated in taking the test.

The test has had limited industrial use, particularly because of its length and possibly because of clinical terminology. However, it has been used in governmental agencies and has been the subject of congressional inquiries related to "invasion of privacy." The reliabilities of the subscales tend to be quite low—thus, one may be investing a great deal of time for some data of uncertain consistency.

BERNREUTER PERSONALITY INVENTORY Like the Guilford-Zimmerman, the Bernreuter Personality Inventory is a combination of previous personality measures (four, in this case). It consists of 125 yes-no items and gives scores on the following scales:

B1N.	Neuroticism	B4D.	Dominance
B2S.	Self-sufficiency	F1C.	Confidence
B3I.	Introversion	F2S.	Sociability

The first four scales were developed by Bernreuter. However, they were found to be highly correlated (e.g., B1N and B3I correlated as high as 0.95) and therefore somewhat redundant. Flanagan (1935) subsequently factor-analyzed the test items and found two relatively independent scales, F1C and F2S. These latter keys should be used in place of, rather than in conjunction with, the first four scales.

RORSCHACH TEST This test is a projective instrument rather than an inventory (Rorschach, 1947). The Rorschach test presents a series of ten standardized inkblots to the subject. In each instance the subject states freely what he sees, either in the parts of each blot or in the complete blot. The examiner, who must be trained in this technique, analyzes the results in numerous ways. Scoring is based on the type of response—whether the subject reports movement, human figures, animate or inanimate objects, etc. The rigid supporters of this technique exhibit an almost religious fervor in their claims, and conversely the attackers insist that the test is useless. The authors believe that the test does have diagnostic value in analyzing personality and emotional aspects of the individual, but it does depend upon the skill of the test interpreter.

THEMATIC APPERCEPTION TEST (TAT) A second well-known projective test, the Thematic Apperception Test (TAT) was developed by Murray (1943). A series of pictures is presented to the subject and he is asked to tell an extemporaneous story about each picture. Analysis of the themes of the stories, especially when the same themes are repeated, reveals dominant drives as well as conflicts and inhibitions. In this test as well as the Rorschach, special training is necessary.

HUMM-WADSWORTH TEMPERAMENT SCALE A personality inventory, the Humm-Wadsworth Temperament Scale consists of 318 questions to which the testee responds *yes* or *no* (Humm and Wadsworth, 1935). Seven different aspects of temperament are obtained from the inventory. Very little validity information has come to light regarding its value in industry.

ACTIVITY VECTOR ANALYSIS A highly controversial test is the Activity Vector Analysis. According to its author, W. V. Clarke (1953), it is used in over 100 companies in business and industry. The test consists of a sheet containing 81 descriptive words. The respondent "places an X (in column 1) before every word that has ever been

used by anyone in describing you." In column 2 the respondent "places an X opposite every word which you honestly believe is descriptive of you." This results in a scoring of four vectors—aggressiveness, sociability, amiability, and avoidance—and an overall score or activity level. The novelty is the claim that after only ten days of training any person who is a college graduate and has had five years of industrial experience can become an analyst. We wonder whether the popularity of this test is based upon "You too can become an expert and in only ten days."

A STUDY IN CONTROVERSY

Perhaps no single test better illustrates the controversial problems involved in industrial testing than does the Activity Vector Analysis. A brief review of the controversy surrounding this assessment device should serve to illustrate some of these difficulties. Locke and Hulin (1962) did a comprehensive review of the use of this test as an industrial selection device. They reviewed a total of 18 studies, 17 of which were relevant to the value of AVA in industry. The majority of these studies, however, were of the concurrent validity type in which the AVA was given to present employees. Locke and Hulin reported that there was evidence that it could differentiate between

1. Managers and production workers
2. Executives and a mixed worker group
3. Seven different occupational groups ranging from machine operators to company presidents
4. Good and poor employees in several different occupations

However, in all the present-employee studies, validities were obtained by developing a key based upon that sample and then applying the key back to the sample upon which the key was developed—a procedure known as "fold-back" validation.[3] Rarely was any attempt made to conduct the necessary process of cross-validation on a different group of present employees. Locke and Hulin believe that all too often the authors of the AVA studies exceeded accepted limits in discussing the value of their findings, often leaving the reader with the impression that their validities were indicative of the *predictive* power of the test.

Only one study, in the opinion of Locke and Hulin, used a true predictive validity procedure where a previously developed scoring key was given prior to employment and then checked later for validity. In this instance no validity was obtained.

It would appear that a decision that the AVA has predictive value must at least be postponed until more substantial evidence is available. At issue here, however, is not only the actual value of the test, but also the discrepancy between the claims made for its value and the actual empirical findings. Locke and Hulin are not alone in taking the AVA to task on the basis of misrepresentation of research findings. Dunnette and Kirchner (1962) have strongly protested the "colored" reporting of AVA data by Meranda and Clarke (1959). Bennett, in his review of the AVA for Buros' *The Fifth Mental Measurement Yearbook* (1959) is also rather uncom-

[3] The term "fold-back validation" in general refers to the procedure of validating an empirically deprived scoring key using the same data from which the key was obtained. This is clearly not a good procedure and one which will lead to artificially high validities.

plimentary, stating, "The mumbo-jumbo of allegedly sophisticated statistical procedures is no substitute for demonstrated validity." The interested reader may enjoy the Dunnette and Kirchner article (1962) and the corresponding reply by Meranda and Clarke in the same *Journal of Applied Psychology* issue.

OVERVIEW OF PERSONALITY TESTING IN INDUSTRY

Ghiselli and Barthol (1953) reviewed 113 studies dealing with the validity of personality inventories in employee selection. Their results are summarized in Table 4.3. They conclude that under certain circumstances scores on personality inventories correlate better with proficiency on a wider variety of jobs than might have been expected. The authors recognize both the potential value of personality testing in industry and the need for a vast increase and improvement in research and development.

TABLE 4.3 *Weighted Mean Validity Coefficients of Personality Inventories for Various Occupational Groups*

Mean r	*Total Number of Cases*	*Total Number of r's*	*Occupation*
0.14	518	8	General supervisors
0.18	6433	44	Foremen
0.25	1069	22	Clerks
0.36	1120	8	Sales clerks
0.36	927	12	Salesmen
0.24	536	5	Protective workers
0.16	385	6	Service workers
0.29	511	8	Trade and crafts workers

SOURCE: E. E. Ghiselli and R. P. Barthol. The validity of personality inventories in the selection of employees. *Journal of Applied Psychology,* 1953, *37,* 18–20.

EXAMPLES OF INDUSTRIAL TESTING

A few of the many studies that have been conducted will be reported rather completely to illustrate the systematic discipline required in the use of tests for industrial purposes. None of these studies is to be regarded as illustrating an ideal testing situation; each was conducted within the limitations imposed by the particular industrial situation, but each attempted, to the degree possible, to conform to the rigid standards of research.

BLUM STUDY

A study on the use of dexterity tests to select watch-factory workers was conducted by Blum (1940). The first step in the research called for job analysis of the different types of work performed in the factory. This brought to light the follow-

ing important job requirements: fine finger movements, the manipulation of tweezers, and the ability to continue performing delicate and sometimes intricate tasks over long periods of time without any increase in tension or maladjustment.

A review of the available test material indicated that in view of the factory's requirement of a brief testing period, the Johnson O'Connor finger dexterity and tweezer dexterity test would probably be best. Prior to embarking on the testing program, however, the various available criteria of success were surveyed and three measures were obtained: length of employment, salary ratio, and foremen's ratings.

The length of employment criterion was divided into four categories. The first was the "less than one week" group. The majority of employees in this category left or were dismissed within one week because of poor ability for the job. From the point of view of this criterion, this group was the poorest. The second category was the "one week to four months" group. Management believed that the job could be learned adequately within four months. Those who could not learn it in this time either were dismissed or resigned. From management's viewpoint, such an employee was inefficient; according to the employee, he could not earn enough. In any event, the employment relation was severed. The third category was the "four months to one year" group. These employees were regarded by management as moderately successful. The fourth and final category was the "one year or longer" group. These employees were considered to be most successful; the training costs for them were lowest and the group had the highest production.

The second criterion was salary ratio. Since all employees were on a piece-rate basis, earnings were a direct reflection of production. The figure used was the average of the weekly earnings over a three-month period. This time was considered long enough to average out minor peaks or slacks in business and minor difficulties in production. The actual criterion was expressed in the form of an index, with X dollar earnings per week equal to a base of 100.

The third criterion was the rating of employees by the foremen. Each foreman was asked to give an overall rating on the basis of his opinion of the employee's usefulness and efficiency. In accordance with this rating, each employee was classified as excellent, good, average, poor, or unsatisfactory.

Five measures of test performance were established. Two were based on speed (total time required to do each test). Another measure was the rating of the subject's qualitative performance on each test. Since this adds meaning to any objective score such as total time or items completed, it will be described in detail in the hope of encouraging the use of such a measuring technique.

The qualitative rating is the examiner's overall estimate of the test performance. It includes how well the subject follows directions, tension during the test, and the method used to complete the performance. A rating scale of "good," "average," and "poor" was used for the following qualities: accuracy of selection, grasp of pins, positioning of pins, placing of pins, hand tremor, conditions of board, pace, position and movements of arm, and body posture.

The fact that there is an appreciable distribution in qualitative performance ratings is illustrated by the findings in this particular study as presented in Table 4.4.

This measure of test performance results in an overall rating by the examiner of the subject's manner during the test. Most performance tests and many pencil-and-paper tests readily lend themselves to such a rating. These ratings are useful in considering a person for hiring.

The fifth measure was a comparison of the time taken to do the second half of

TABLE 4.4 *Distribution of Quality Ratings on Finger and Tweezer Dexterity Tests*

	Finger Dexterity Test		*Tweezer Dexterity Test*	
Quality Ratings	*Employees in Watch Factory*	*Applicants for Factory Jobs*	*Employees in Watch Factory*	*Applicants for Factory Jobs*
Total Number of Individuals	163	420	143	420
Excellent	5.0%	3.6%	13.3%	5.9%
Good	44.1%	39.2%	55.9%	50.1%
Average	48.5%	50.9%	28.0%	36.3%
Poor	2.4%	5.8%	2.8%	7.1%
Unsatisfactory	0.0%	0.2%	0.0%	0.5%

SOURCE: M. L. Blum. A contribution to manual aptitude measurement in industry. *Journal of Applied Psychology*, 1940, *24*, 381–416.

the finger dexterity test and the time taken on the first half. It was assumed that the difference between the total time on each half might measure ability to improve the rate of speed and might therefore be useful as an indicator of rate of improvement on the job.

In this study 258 subjects were used. Of this number, 137 were tested subjects, 84 were referred without testing and constituted a control group, and 37 were used in the pilot study which preceded the major study. This last group was also considered a follow-up group. All testing was done under favorable environmental conditions, that is, in a testing room; and all the subjects knew that they were taking a test to be used by a specific watch factory for hiring purposes.

The data were intensively analyzed in three ways. First, the characteristics of the various criteria as well as their interrelations were studied. Second, the characteristics of the various indicators and their interrelations were studied. Third, the relation between the criteria of success and the test results was analyzed; this was the crucial step.

The correlations among the criteria indicated that each was measuring a relatively different aspect of success on the job. For example, the correlation between foremen's ratings and salary ratio was + 0.13. The correlation between length of employment and foremen's ratings was + 0.25. The correlation between salary ratio and length of employment was highest, + 0.44.

Four of the five measures of test performance—the time scores and qualitative performance on both tests—were established as reliable; but one measure, improvement, was not established as reliable. Comparing each intercorrelation of each of the five measures showed that all were low. Eight were less than + 0.20, and one was only slightly higher. The only high intercorrelation was between quality rating and total time of the tweezer dexterity test; this was + 0.71. However, quality ratings generally cannot be considered intimately related to speed because the correlation between the quality rating on the finger dexterity test and the total time on that test was only + 0.13. Low correlations among test measures and low intercorrelations among job criteria are actually more favorable than high intercorrelations. When such correlations are high, their predictive value is limited since all will be predicting the same thing. By the same token, if all the job criteria are highly

intercorrelated, they will all be measuring the same thing instead of different aspects of success. The total time score on the finger and tweezer dexterity test generally had the highest prediction value of the job success criteria. Quality ratings on test performance were valuable for prediction in some instances, but improvement on the second half of the finger dexterity test was not predictive for any of the criteria. Some of the specific findings are as follows (Blum, 1940).

1. Quality ratings during testing for finger dexterity were not indicative of length of employment. But those with "average or below" ratings on the tweezer dexterity test were found in the shorter categories of employment. Sixty-one percent of the group who received such ratings were no longer employed after four months, and only 27 percent of those who received "above average" ratings left or were dismissed within this period. The difference was statistically significant. When the quality ratings for both tests were combined and distributed according to length of employment, no statistically significant differences existed between those rated high and those rated low in test performances.

2. Low correlations were reported between the quality ratings on either dexterity test and salary ratios. The correlations were $+ 0.17 \pm 0.11$ for the finger dexterity test and salary ratios, $+ 0.15 \pm 0.11$ for the tweezer dexterity test and salary ratios, and $+ 0.05 \pm 0.12$ for the combined quality ratings and salary ratios.

3. Quality ratings on the finger test were related to foremen's ratings of job performance with a coefficient of contingency of + 0.50. The C for tweezer quality ratings and foremen's ratings was + 0.24. A coefficient of contingency of + 0.30 was obtained when foremen's ratings were correlated with the combined quality ratings on both tests. (Maximum C = + 0.86.)

4. Improvement on the second half of the finger dexterity test was not predictive of length of employment.

5. Improvement correlated with salary ratio $- 0.06 \pm 0.13$. This is the only comparison of the entire investigation that showed a negative though unreliable relationship between test indicators and the criteria of proficiency.

6. Of the workers who received "A" ratings by foremen, 100 percent showed improvement on the second half of the finger dexterity test. Of the workers who received "D" ratings, only 50 percent improved. Eighty-four percent of those who received "B" ratings improved, and 61 percent of those who received "C" ratings improved. The difference in percentage between those who received "A" and "D" ratings is not statistically reliable, possibly because of the limited number of subjects.

7. Time scores on both the finger and tweezer dexterity tests were faster on the average as length of employment increased. The difference for the average time on the finger dexterity test between the "less than 7 day" and the "more than 1 year" groups was statistically significant. In the same comparison on the tweezer dexterity test, the D/σ difference was 2.5. Combining the finger and tweezer dexterity time scores did not increase the statistical reliability of the difference, and the D/σ difference between the "less than 7 day" and "more than 1 year" groups was 2.3.

8. The correlation between finger dexterity time score and salary ratio was $+ 0.26 \pm 0.10$; between tweezer dexterity time score and salary ratio it was $+ 0.32 \pm 0.10$; and between the combined test times and the salary ratio it was $+ 0.39 \pm 0.09$.

9. The "above average" group according to foremen's ratings was 5 seconds

faster on the finger dexterity test and 9 seconds faster on the tweezer dexterity test than the "average and below" group. The difference for the combined test scores between the two groups was 12 seconds in favor of those receiving above average ratings. These differences were not statistically reliable.

10. The practical value of the critical scores (time score of 5 minutes, 30 seconds or better on the tweezer dexterity test and 7 minutes, 30 seconds on the finger dexterity test) which were suggested in the pilot study (Candee and Blum, 1937) is clearly indicated in this investigation. These scores discriminate employees in the watch factory with a considerable degree of exactness according to the criteria of proficiency.

11. A comparison according to length of employment showed that 7 percent of the group that "passed both tests" left within one week, whereas 23 percent of the "no test" group and 24 percent of the group of workers who "failed either or both tests" were unemployed after one week. The differences in the percentages between the first and the last two were statistically significant with critical ratios of 3.2 and 3.4. The greatest possibility of prolonged employment was found in the group that "passed both tests." Of this group, 72 percent remained four months or longer. This percentage was significantly different from that of the "no test" group (D/σ difference of 3.1), and from that of the group which "failed either or both tests" (D/σ difference of 4+).

12. A comparison according to salary ratios indicated that the group that "passed both tests" earned the most money. The earnings of this group were statistically different from the earnings of the group that "failed one or both tests." The "no test" group was superior in salary ratio to the group that "failed either or both tests." The group that "passed both tests" was not significantly superior to the "no test" group.

13. A comparison according to foremen's ratings showed only a trend. The group that "passed both tests" was rated by foremen as "better than average" in 34 percent of the cases. The group that "failed either or both tests" was rated as "above average" in 25 percent of the cases. This difference was not statistically reliable. No differentiation between the "no test" group and the "passed both tests" group is possible according to foremen's ratings.

14. A follow-up of the subjects in the pilot study supports the findings of the present investigation that time scores on the tests are indicators of proficiency. Two years prior to the "follow-up," 20 workers were selected by foremen as superior and 17 as mediocre workers in the watch factory. These groups were originally significantly different in their test scores on both the finger test and the tweezer test. Significant differences were later obtained between these a priori groups in terms of (a) percentage discharge and (b) salary ratio.

Before presenting a review of other studies, it is desirable to take stock and consider the implications of such research. The above study was not an "ideal" one. Because of factory conditions, it was impossible to use desirable scientific controls. For example, equal-sized groups in the various categories could have been arranged for in a laboratory more readily than in the factory. The demands of research must often conform to prevalent conditions, but only insofar as planning and methodology are concerned. The statement does not hold for research conclusions, for these must be reported independently of plant conditions and must conform very strictly to the data obtained.

The industrial psychologist is justified in conducting laboratory studies only as a

preliminary to the research in which the industrial situation necessarily becomes the "laboratory." Laboratory conclusions cannot be generalized and applied to the industrial scene without checking them in every way possible. Whether the problem involves tests for selection or any other purpose, the industrial psychologist often has to modify procedures and demonstrate a flexibility in manner that may be unorthodox from the scientific point of view. If, for example, he will not forsake rigid controls in some cases, he may be forced to give up the entire problem. To the industrial psychologist, the lesser of the two evils is to conduct research in industry under the best conditions that industry can provide.

A further illustration of this point concerns the desirability of having large numbers of subjects for an experiment. However, an industrial organization is often not in a position to hire on a large scale. In this case it is better to sacrifice the number of subjects rather than the entire study.

This principle does not imply that a form of research euphemistically called "directed research" should be tolerated. Research reports must agree with the conclusions based upon the facts obtained. An industrial psychologist who does not report negative findings because this would create disrespect and might lead to his dismissal is as wrong as the business executive who wants the research to back up his sales ideas and therefore "directs" the research and its findings.

The points just discussed are important and should serve as a means of evaluating not only test research but also other types of research in the field. To this end, a few additional studies on test selection will be described.

GHISELLI STUDY

In his study of tests for selecting inspector-packers, Ghiselli (1942) used a battery of several tests on a group of 26 women employees working as inspector-packers in a pharmaceutical concern. There were five main duties in this job: filling containers, inserting stoppers, examining the contents by eye, labeling the containers, and packaging. The job, although routine, is extremely important, for the presence of extraneous material or incorrect labeling might result in serious illness or even death for anyone who happened to be sold the wrong product. The criterion of job efficiency was the combination of ratings by the forelady and the supervisor.

The job analysis indicated that the important abilities that should be measured in the preliminary battery of tests were:

1. Dexterity of fingers, hands, and arms
2. Eye-hand coordination
3. Estimation of the size and form of objects
4. Ability to observe difference in details

To this end a battery of six tests was administered to the employees. It was found that the average performance of the inspector-packers was vastly superior to that of samples of an adult population in the Minnesota Placing and Turning tests and the Paper Form Board test. Although this information is useful, selection is justified only when a relationship is found to exist between test performance and job performance. Such a relationship is called a validity coefficient. The Minnesota

Paper Form Board test was found to be more closely correlated with the criterion than any other test in the battery; this test correlated to the extent of +0.57. The Pegboard test correlated with the criterion to the extent of −0.50, and the Turning test to the extent of −0.40. The other tests had lower correlations with the criterion.

This study, like the one previously reported, has shortcomings. The criterion of proficiency was a rating by superiors; but, since it apparently was the only criterion available, it had to be used or no tests could have been evaluated. Too often in industry such ratings comprise the only available criterion. The number of subjects, 26, is small; but when we remember that this constitutes a large number of people performing the same task, it can be considered a satisfactory sample. The study presents no evidence on the success of its recommendations that were adopted for future use; such information, as well as recommended critical scores, is often helpful in evaluating the use of tests.

HAY STUDY

Edwin N. Hay (1943), using a battery of tests, was able to predict 91 percent of the better bookkeeping machine operators and 72 percent of the less satisfactory operators. Prior to establishing these results, all the procedures involved in test research and development were followed. These included job analyses, establishment of adequate job criteria, preliminary selection of the test battery, and statistical investigations. Only after many years was it possible to recommend the particular test battery as successful for prediction purposes.

The job analyses revealed that the outstanding requirement for machine bookkeeping is bimanual ability. A detailed time and motion job analysis showed that there were 5 distinct operations which could be broken up into 18 motions. The average time for the five operations—selecting ledger card, inserting ledger card, picking up previous balance, post-checking amount, and returning card—was 6.8 seconds. In most of these operations the eyes and both hands were used. Speed and accuracy were required for the successful performance of this job. The difference between the ability of the best and the ability of the poorest operators was greater than 2 to 1; that is, the best operators did more than twice the work done by the poorest operators.

The tests used included the Otis Intelligence test, the Minnesota Clerical test, the Ziegler Rate of Manipulation tests, and other tests of clerical ability such as filing and name finding. In all, 22 tests were administered. Although six of them were of the hand, arm, and finger dexterity variety none of these six showed any considerable relationship with the criterion "speed of posting." However, the Otis test, the Minnesota Clerical Test-Numbers, and the Alpha Number series correlated +0.50 or higher with this criterion. The multiple correlation technique indicated that the Otis test and the Minnesota Clerical had a correlation of +0.65 with the criterion. The highest multiple correlation between the test battery and the criterion was +0.71 and included the Otis, the Minnesota Numbers, the Alpha Number series, and the Fryer Name Finding.

Hay's study is valuable from many points of view. The evidence presented in the follow-up indicates that in this case the tests worked. Table 4.5 shows the steady improvement in the average production of bookkeepers since selection based on testing was introduced.

TABLE 4.5 *Average Production of Test-Selected Bookkeepers Showing Influence of Selection Techniques*

Date	*Number*	*Average Production*
October 1937	43	105.0
November 1939	40	108.7
April 1940	30	109.7
December 1940	32	110.2
December 1941	26	110.9

SOURCE: E. N. Hay. Predicting success in machine bookkeeping. *Journal of Applied Psychology*, 1943, *27*, 483–493.

SARTAIN STUDY

A. Q. Sartain (1945) administered a battery of seven tests to a group of 47 employees in the inspection department of an aircraft factory. The criterion was set by the ratings of the instructors in a refresher course being given to inspectors. The instructors were also familiar with the performance of these employees and their ratings undoubtedly reflected this knowledge. The test battery included the MacQuarrie, the Otis, the Cardall Test of Practical Judgment, the Minnesota Paper Form Board, the Industrial Classification Training Test, the Bennett Test of Mechanical Comprehension, and the O'Rourke Test of Mechanical Aptitude. The multiple correlation of this entire battery with the criterion was +0.787. However, three of the tests—the MacQuarrie, the Cardall, and the Minnesota Paper Form Board—yielded results equally good, the multiple correlation being +0.780. Since the difference between these two correlations is negligible, it would be advisable to use these three tests, rather than all seven.

SHUMAN STUDY

John T. Shuman (1945) administered a battery of tests to a group of employees and applicants for various jobs at the Lycoming Division of the Aviation Corporation. The tests included the Otis, the Minnesota Paper Form Board, the Bennett Test of Mechanical Comprehension, the O'Rourke Test of Mechanical Aptitude, and the Minnesota Vocational Test for Clerical Workers. The employees tested included inspectors, engine testers, machine operators, foremen, and job setters, among others. Shuman reports that the average improvement in selecting excellent workers was 18 percent with the Bennett test, 15 percent with the Otis, and 13 percent with the Minnesota Paper Form Board. The critical scores that made this improvement possible meant the elimination of one out of every four persons tested.

The Bennett test when compared with the criterion (ratings) was found to be highly correlated to the job of job setter, the correlation being +0.73. The Minnesota Paper Form Board correlated +0.59 with this same job. The Otis test had the highest correlation with the job of engine tester, +0.57. Shuman found that the tests in his battery correlated better with jobs that require skills such as working on machinery precision parts or testing aircraft engines, rather than purely manual skills. He believed that tests can be more useful in assigning employees to job category levels than to specific jobs. An incidental finding, but one that deserves mention, is the fact that one test had to be dropped because of dissatisfaction on

the part of the subjects. This was the O'Rourke Test of Mechanical Ability, which was too long and proved to be almost useless with women applicants, many of whom did not even try to take it. This emphasizes that a test battery must be favorably received by applicants if any value is to be attached to the results.

SUMMARY OF EXAMPLES

Many other studies have been as effective as the few described above. However, further examples would only belabor the point. The use of psychological tests in industry demands research and development in the light of the particular problem and the specific plant. The fact that this work has been done before does not make it unnecessary. Rather, it means that the chances of success are improved each time. The essential point is that test results must be correlated with job success; they can never be taken for granted.

UNITED STATES EMPLOYMENT SERVICE TESTING

A chapter on the use of psychological tests in industry would be incomplete without reference to the considerable work done by the United States Employment Service (USES). Some state employment services send to employers, without any charge, applicants who have been given psychological tests. Of course, in many instances the testing is not tailor-made to the requirements of a particular company and this may prove a disadvantage. But, desirable as it is to have the testing program fit the specific needs of a specific company, it must be recognized that many firms may not wish to spend money on psychological testing. Under such circumstances, the facilities offered by public agencies may be more beneficial than detrimental, provided some check on the usefulness of the test in the specific company is made.

A battery of tests known as the General Aptitude Test Battery has been made available to the various state employment services by the United States Employment Service (1947). It consists of eight pencil-and-paper tests and four apparatus tests.

TABLE 4.6 *Aptitudes Measured by Tests in the GATB*

Aptitude	*Test Name*	*Test Number*
Intelligence (G)	Three Dimensional Space	3
	Vocabulary	4
	Arithmetic Reason	6
Verbal aptitude (V)	Vocabulary	4
Numerical aptitude (N)	Computation	2
	Arithmetic Reason	6
Spatial aptitude (S)	Three Dimensional Space	3
Form perception (P)	Tool Matching	5
	Form Matching	7
Clerical perception (Q)	Name Comparison	1
Motor coordination (K)	Mark Making	8
Finger dexterity (F)	Assemble	11
	Disassemble	12
Manual dexterity (M)	Place	9
	Turn	10

The twelve tests measure nine of the aptitudes that are often related to the requirements in the performance of jobs. Table 4.6 presents the test names and the aptitudes measured.

The General Aptitude Test Battery is intended to be used in two ways:

1. As an integral part of counseling where a measure is needed of the applicant's abilities in relation to the various fields of work in which he may have interest but no practical experience.
2. As an approach to the problem of developing specific batteries for the countless occupations for which selection tests may be needed. Those tests in the General Aptitude Test Battery which measure abilities significant to the successful performance of a given job can be administered as a specific aptitude test battery, and the other tests omitted.

According to the guide in the use of the GATB, certain of the tests and their cutoff or minimum scores are recommended for use in the selection of applicants for a wide variety of occupations. For example, the job of all-round mechanical repairing would include testing for G, N, S, and F. Computing work requires tests for N, Q, K, and F. Plumbing would include tests for N, S, and M. Typewriting requires V, Q, K, and F.

The battery has been administered to a very large number of persons employed in a wide range of occupations. Over 500 employers and many vocational schools and colleges have cooperated in the development of occupational norms. It is obvious that problems of establishing reliability, validity, and norms for a test battery such as this with its intended wide coverage of so many occupations are stupendous if not insurmountable. Indeed, the procedure often adopted by the USES has been that of synthetic validity, a procedure discussed in Chapter 2. The reader will remember that synthetic validity involves *assuming* test validity simply because the job in question has elements identical to those found in a job for which the test has previously demonstrated statistical validity. While in many practical situations it is just not possible to evaluate a test in the proper manner, synthetic validity must never be used with complacency—it must always be used with caution. While the GATB has demonstrated its validity with a variety of situations and jobs, it is probably being used in many situations where its value as a predictor is highly questionable.

Blum has been rather critical of this battery in a review in Buros' *Fourth Mental Measurements Yearbook* (1953). Since that time (no cause and effect relationship implied) additional data, conferences, and published material have tended to improve the situation somewhat. The GATB has promise, but additional work is constantly needed.

TESTING OF EXECUTIVES

Traditionally, industrial testing has been oriented toward the blue-collar, sales, and clerical workers, since it was felt that this was where the largest increase in efficiency could be achieved. These individuals usually represented the bulk of the labor force employed by a company, and by improving the average quality (as defined by one or more criterion dimensions) large benefits should accrue to that company. In more recent years the emphasis has been shifting to an intensified concern for selection of individuals higher up in the managerial hierarchy. American business

is realizing the importance of the manager to the system and is becoming aware that one bad person in a decision-making function may cost the company much more, in the long run, than 20 to 30 poor persons at lower-level jobs.

Management selection poses many problems that are not found with the more traditional testing and selection situations. Gellerman (1958) has listed what he considers the "seven deadly sins" of executive selection. These are:

1. Careless treatment of candidates
2. Overdependence upon expert opinion
3. Misjudging job requirements
4. Making "stop-gap" appointments
5. "Pigeon-holing" prospective candidates
6. Disregarding the company personality
7. Overlooking personal compatability

The major problem in predicting executive success is in establishing some valid criterion of it. Attempts to use tests to predict success, once success has been arbitrarily defined for research purposes, have been less than successful—particularly tests of ability. Gaudet and Carli (1957) have estimated that seven times as many executives fail due to personality problems than for lack of technical competence. As Taylor and Nevis (1957) point out, this is not really too surprising. First, the executive job is a very complex one. Thus, to isolate specific abilities may not be sufficient. Second, the abilities required are generally cognitive rather than physical. Third, intelligence measures are probably not good predictors because to get into management requires a fairly bright person in the first place.

An example of the use of tests with supervisory level personnel is provided by Neel and Dunn (1960). They used the *How Supervise* test, the *F Scale* measure of authoritarian personality, and the *Wonderlic* to predict the degree of success that 32 supervisors would have in a supervisory training program. The results are given in Table 4.7.

TABLE 4.7 *Correlations of the Tests with Training Success*

	(1)	(2)	(3)	Criterion[a]
(1) How Supervise		0.08	0.33	0.69
(2) F Scale			−0.23	0.39
(3) Wonderlic				0.25
(Multiple correlation = 0.77)				

[a] Criterion was grade at end of course.

SOURCE: R. G. Neel and R. E. Dunn. Predicting success in supervisory training programs by the use of psychological tests. *Journal of Applied Psychology*, 1960, *44*, 358, 360.

A further example of the frustration which is so often experienced in attempting to predict the success of executives can be found in a study by Wagner (1960). He attempted to predict ratings of on-the-job success of 150 executives using a total of 31 different variables. These variables included measures of both intelligence and personality. The only correlation which was found to be high enough to be of any value was the correlation between ratings and amount of education ($r = 0.39$).

However, the chance likelihood of this correlation is substantial. At present it is safe to say that this is the area where testing has had perhaps its smallest impact, with tests of personality and temperament appearing to offer the greatest potential value. Even these have not been particularly promising to date. Taylor and Nevis (1957, p. 473) summarized in their review of projective techniques in management selection, "All too often, we do not know how effective projective devices actually are. Nor can we say—unless on the basis of a hunch, prejudice, or individual intention—which are most appropriate and which are not useful at all. In short, personnel assessment is at the moment much more of an art than a science."

TESTING OF SCIENTISTS

America is fast becoming the society of science. Billions of dollars each year are allocated to applied and basic research in the Federal budget. In addition, the money spent each year by private industry for research and development reaches astronomical proportions. Since the end of World War II there has been a premium on scientific talent of all types which has shown no signs of diminishing. This premium has led, by simply a supply and demand process, to even larger numbers of scientifically trained professional people. While most of these people have certificates or degrees which, at one time at least, were sufficient testimony to professional competence, many employers are currently attempting to select the "cream of the scientific crop" by using tests as a selection aid.

As in the case with executives, the criterion problem immediately raises its ugly head. What is a "successful" scientist? Ignoring that problem temporarily, how efficient are tests in separating good scientists from poor scientists?

At best, the available evidence on this question is exceedingly sparse. Much of the research has used predictors other than tests such as biographical information. Several of these studies will be discussed in detail in Chapter 5, which covers selection methods other than tests. A few notable exceptions are available which have attempted to predict scientific success using paper-and-pencil tests.

TAYLOR *ET AL.* STUDY

One of the more elaborate research projects has been conducted by Taylor, Smith, Ghiselin, and Ellison (1961). They carried out extensive interviews with over 200 physical scientists concerning the nature of scientific productivity and the characteristics of effective scientists. Using the interview suggestions as a base, data were obtained on 52 different criteria. These measures were then clustered (using factor analysis) into 14 relatively independent dimensions by which the performance of scientists could be evaluated. A number of tests were then used to predict how well a scientist would score on each of the 14 dimensions. In addition to the 14 dimensions obtained in the factor analysis, three other criteria based upon supervisor and peer judgments were included, since these are the most often used methods of performance evaluation. Thus the criteria used were:

1. Productivity in written work (effectiveness in completing paper work)
2. Recent quantity of research reports (number of articles, research reports, etc., in a two-year period)
3. Quality (without originality) of research reports

4. Originality of written work
5. Scientific and professional society membership
6. Actual quantity of work output as judged by peers, supervisors, and laboratory chiefs (higher level supervisors)
7. Creativity rating by laboratory chiefs (higher level supervisors)
8. Overall performance (quality ratings by supervisors on ten different scales)
9. Likableness as an effective member of the research group
10. Visibility of the scientist (well known by person or by name)
11. Recognition for organizational contributions (organizational awards)
12. Status-seeking, "organizational-man" tendencies
13. Current organizational status
14. Contract monitoring load (number of research contracts supervised)
15. Peer ranking on productivity as a scientist
16. Supervisory rating of drive-resourcefulness
17. Supervisory rating of creativity

A total of 130 different predictors were validated against *each* of the 17 different criteria. Many of these predictors were simply different a priori scoring keys which could be applied to the same questionnaire or test instrument. For example, the Personality Research Inventory was given to all scientists in the validation sample for purposes of prediction. However, this single test yielded a total of 23 subscores, each of which was separately validated. The validation sample consisted of 107 scientists.

By relating the 130 predictors to the 17 criteria, 2,210 validity coefficients were obtained—a testimony to modern computer technology. Of these, 568 were found to have an individual chance probability of less than 5 times in 100.[4]

The different predictors varied considerably in their efficiency as Table 4.8 shows.

TABLE 4.8 *The Percent of Valid Scores Obtained from Each of the Major Categories of Predictor Information*

Type of Test	*Percent of Scores Valid*
Biographical Information Blank (BIB) (with empirically keyed scores)	47
Biographical Information Blank (BIB) (a priori keyed scores only)	34
Self Ratings (SR)	33
Grade-Point Average	24
Minimum Satisfactory Level (MSL)	22
Profile Matching	20
Motivational Analysis Test (MAT)	8
Personality Research Inventory (PRI)	8
Creative Process Check List	6
Aptitude Tests	4

SOURCE: C. W. Taylor, *et al.* Explorations in the measurement and prediction of contributions of one sample of scientists. ASD-TR-61-96, Personnel Research Laboratory, Lackland Air Force Base, Texas, April 1961.

[4] One would expect at least 110 of the correlations to be significant at the 0.05 significance level simply due to chance. While there were many more than that found, the question remains "which of the 568 are indicative of true validity and which are the 110 chance validities?"

As the Table clearly indicates, the most efficient predictor was the Biographical Information Blank (the empirically keyed scores being exceptionally valid due to the spurious fold-back validity obtained when applied back to the same sample). Of the paper-and-pencil tests used, only the MAT showed any potential as a predictor, having 8 percent of its correlations above chance level (where 5 percent would be expected by chance alone). The high-level aptitude tests were paper-and-pencil tests designed to measure such things as flexibility, originality, conceptual foresight, sensitivity to problems, etc. These showed little or no predictive promise with this sample of scientists.

The general conclusion that one is forced to draw from a close examination of the study is that paper-and-pencil tests must take a back seat, at least at present, to other types of predictive tools in the identification and selection of scientific personnel.

CREATIVITY

The topic of creativity has received a good deal of attention in conjunction with industrial testing since World War II and needs some mention. While it is an exceptionally broad and oftentimes diffuse topic, two aspects of creativity are pertinent to any discussion of testing in industry—particularly in conjunction with a discussion of the testing of executives and scientific personnel, since in many cases creativity is felt to be an essential component to success in these high-level jobs. These two aspects are: (1) Can tests of creativity be useful in predicting a criterion of scientific (or managerial) success, and (2) can other tests be used to predict a criterion of scientific (or managerial) creativity?

The study cited above by Taylor *et al.* provides an example of both situations. Some of the high-aptitude tests used in that study could be classed under the general heading of creativity tests. These were in turn validated against a variety of criteria, some of which were clearly success criteria (for example, "recognition for organizational contributions"). The same study is also an illustration of studies of the second type. Consider, as an example, the correlation between the Motivational Analysis Test and the criterion of supervisors' ratings of creativity.

Jones (1964) also attempted to predict the creativity of industrial scientists (as defined by supervisory ratings) using both aptitude and attitudinal test scores. A multiple correlation of 0.67 was obtained which the author reports as holding up satisfactorily under cross-validation. Datta (1964) also reports a correlation of 0.31 between the Remote Associates Tests and supervisory ratings of creativity for engineers.

A study by Mullins (1959) provides a final illustration of both types of approaches. He administered an interest questionnaire, a vocabulary test, and nine tests of the Guilford Creativity Battery to 131 research personnel in the physical sciences. Forty-two different test scores were derived from the battery, only four of which were found to relate significantly to a global creativity rating. Seven of the predictor scores were found to be significantly related to a criterion of publication frequency.

A POINT OF VIEW

The selection and evaluation of executives and scientists is a serious matter, and while psychologists cannot claim always to provide perfect answers (many times there is no perfect answer), they *can* help to eliminate some of the elements of risk.

The trend in business and industry has been toward recognition that the industrial psychologist is well-qualified to aid management in the evaluation and selection of high-level personnel. Many consulting firms are finding that this activity occupies more and more of their time. It is too bad that, as in all new developments, misunderstanding is widespread and that "gold brick" salesmen have a holiday. The psychologist can be helpful provided management understands that no one has a secret formula, a magic selector, or a panacea. The shortcut is not possible, at least at the present stage of development.

Not too long ago, one author attended a management development meeting at which three staff members of a utility company presented a report on the wonders of a technique they were using to select a wide range of employees from clerical to executive level. They seriously believed that they had the psychological counterpart of the wonder drugs in medicine. However, what they had was a moldy substance of another kind. It was a test administered in five minutes and scored in about the same time. They claimed it resulted in a "know it all, tell it all" possibility. In addition to the three staff members who glowingly reported the value of their work, a vice-president sat silently but smilingly on the side, enthusiastically nodding.

The few psychologists in the audience were amazed. Psychologists generally agree that they cannot validly assess and evaluate a phenomenon as complex as a human being within five minutes. How is such a thing possible? If this had been a matter of tea leaves, very few would have taken it seriously; but since this was a more substantial claim, it simply could not be dismissed without some mental gymnastics. After some deliberation, two tentative explanations were proposed. The first is that many people like to play psychologist and do not have the time to acquire the information through normal and regular channels—that is, graduate schools. The other reason is a little more practical. After all, these people should be expected to defend their position very strongly. If the system used should prove to be nonsense, it goes without saying that they would be fired. Thus it is likely that, in addition to being enthusiastic, they wanted to keep their jobs.

One of the problems is that it is currently "fashionable" to use psychological tests in selecting scientists and executives. The decision to use a psychologist should *not* rest on "keeping up with the Joneses"; it should depend on a real problem in need of a solution. Let us recognize that this is not a fad or a game, and that what is worth doing is worth doing correctly. The executive who desires to use psychological techniques to help select employees should attempt to estimate the nature of his problems and also be given estimates of the likelihood that the techniques used will help in offering solutions.

ASSESSING OF NEEDS

Management must assess its needs before embarking on a recruitment program for future top management people. The simplest way to do this is to take a look at the organization chart and estimate growth needs, then fill in the names and ages of

the people occupying the present positions and estimate the moves that are likely to be made within a ten-year period. This procedure points to the future needs for engineers, controllers, research and development people, and others. It can also help to estimate the numbers that will be needed, and so the task of the recruiter becomes more specific and meaningful. If present staffing is not likely to fulfill the needs, one must go outside the company.

DECISION TO HIRE

The decision to hire must be made. There are at least three ways of doing this: Top management makes its own selections, or it empowers personnel to do the hiring, or an outside consulting organization receives the assignment. If top management is to spend the necessary time, then other aspects of management tend to be delayed or postponed. While it is axiomatic that top management should be involved in the final decision, it does appear that the details of evaluation should be left to people with more time and also more training in the area. Whether the personnel department does this work depends upon the professional level of staff operation. In many companies personnel departments have very few, if any, staff psychologists. Obviously, because of the wide variety of duties and assignments, it is not necessary for all people in personnel departments to be psychologists. However, evaluation of potential executives should primarily be the work of psychologists. If personnel departments are not so geared, then the most economical practice is to call in a consultant.

Since the professional psychologist working in the area of selection and evaluation of executives has no secret formula, he must select his instruments from a variety of psychological tests and interview techniques that are most helpful in enabling him to draw certain conclusions. He is likely to select one or more intelligence tests, interest tests, personality tests, and so on. In addition, he will use one or more interview techniques as a means of gathering data. He is likely to be more accurate when he uses the services of a team of psychologists who gather independent data and then check the data against the hypotheses.

EVALUATION VERSUS PREDICTION

It is important for the psychologist to differentiate between describing the degree of the characteristic possessed by the man and predicting future behavior. If his measures are accurate his descriptions are accordingly accurate. In the realm of prediction he is on less substantial ground. His predictions may be accurate provided he has adequate description of behavior and provided he knows the characteristics necessary to perform the executive duties.

This, of course, means that the psychologist has need for a criterion of success. Management must furnish this either in the form of ratings of success or in their descriptions of the kinds of people needed. For example, as a result of acquiring data about an individual the psychologist can describe him as being accurate, energetic, serious, and superior in intelligence. Further, he can describe the degree to which these traits are possessed. Predicting whether this combination will be successful depends entirely upon the characteristics required in the performance of the executive duties. If management insists that a person be creative, highly imaginative, and excessively ambitious, then the previous description does not allow for the prediction. If management would rather have a person who will work accurately at all costs, the previous description could allow for a prediction of success.

ROLE OF THE PSYCHOLOGIST

In the evaluation of an executive it is important for a psychologist to recognize that he is not playing a management role whether he is on the staff of the company or is employed as a consultant. He must limit his role to furnishing information to management that will enable it to decide. Experience has indicated that the awareness that psychologists, per se, are not management people is desirable. Likewise, management people are not psychologists. In other words, psychologists can furnish descriptions of an individual, management must offer a criterion, and together they can decide the likelihood that this person meets the criterion stated. Ultimately, the decision to hire or promote must be management's and not the psychologist's.

Think of the controller, the president, and the chief engineer of different companies. It is quite apparent that very different people with the same titles can be equally successful. In other words, there is no particular degree of intelligence that is required of a president. He may or may not be the most intelligent executive on the staff. By the same token, one president may be shy and retiring and another may be a glamorous dynamo. This leads to the known fact that the characteristics required on a job are very often determined by the philosophy of management rather than by the demands of the job itself. This, of course, makes the task of predicting managerial success with conventional validation methodology somewhat difficult.

Unless management has made incredible and preposterous decisions, it often becomes the task of the psychologist merely to describe a person's characteristics so that management may get a better estimate of whether or not he will fit. The advantage of the psychologist's description is that he is not emotionally involved or biased as sometimes different executives in a company are. His description can considerably reduce the margin of error. It can definitely provide management with additional information that will make it possible to develop the candidate's potential over a period of years. It can also provide the candidate with additional insights that, in turn, will help him to change, grow, and reach his potential.

AN EXAMPLE OF EXECUTIVE "EVALUATION"

A certain amount of success has been attained in a testing program instituted by one large management consulting firm. Briefly, the approach requires the candidate to spend approximately two days taking tests and being interviewed. Some of the tests are individually administered and are of the projective technique type; that is, they are not structured to have absolute and correct answers, but the answers definitely reflect the individual involved. To be sure, considerable skill is required if interpretations of these data are to be meaningful and accurate. At least two levels of intelligence tests are given. One is rather simple and short; the other is of the power type that tests the limits of the individual. Personality tests of two varieties are given. In one, the person answers questions directly and can quite easily be the way he either wants to be or thinks he is. These results are then checked with test results that are less prone to faking. The comparison of the two measures indicates to a certain extent the degree of self-insight. For example, the person who is impulsive may or may not recognize this trait in himself. If he does, then changes are possible. If he does not, the problem is of a different order.

In addition to a lengthy battery of tests, three types of interview situations are used. One requires that the person talk about himself at length. The interviewer

says very little and makes copious notes about what the person is saying, adding his own judgments about the kind of person he is. These judgments are not regarded as fact but are considered hypotheses and checked against other data independently gathered. Another type of interview used is the "leaderless group conference." If there are five candidates for a job, these five are brought together, given a topic, and asked to hold a conference on that topic and to pay no attention to the observers. Three psychologists observe the interaction of the candidates. The third form of interview is the panel technique in which three psychologists meet with the applicant. After discussing the usual social amenities, conversation is brought around to a number of situations. The role of the psychologists is to a large extent to disagree with the candidate. The way the person behaves under such circumstances allows for certain assumptions related to descriptions as well as predictions.

Probably the most important feature of the appraisal technique used by this management consulting firm is that a team of psychologists is always involved. Depending upon circumstances and costs, there are between three and six psychologists on the assigned team. Each is involved in different aspects of the data gathering, and each is required to submit an independent report without knowing the data acquired by other sources. When hypotheses are in disagreement, data checks as well as conferences are held in order to clarify the picture. The final report is then written in nonpsychological language with division headings that are related to management performance. This report is submitted only to top management, and a conference is held at the time with the team of psychologists present. Too many written reports tend to be easily misunderstood. The conference results in clarification. The conference is necessary, and reports are never submitted unless a conference has been arranged.

Any new field of development must recognize that different procedures will be used. For evaluation, the authors favor the team approach because it allows psychologists to check their hypotheses (that is, one can assess the reliability of the evaluations). Since each gathers a set of independent data, the possibility of contamination is held to a minimum. Other techniques include only interviews, only projective techniques, only inventories of a personality variety, or only one psychologist in the act from overture to finale. Still other techniques use combinations of the above-mentioned methods. Although this is a very promising field, serious developmental work and exhausting evaluation are again necessary before the acme of perfection is reached.

RESPONSE SETS OR FAKING

Many times throughout this chapter we have referred to the problem of faking on tests. It is impossible to understate the degree to which this is a critical problem in psychological testing. It is probably the greatest single problem, particularly with personality and interest inventories. With these assessment instruments there is no "correct" response; the test attempts to solicit attitudinal rather than factual responses. Many personality inventories, such as the MMPI, have special keys which can be used to give an indication of the degree to which a person is either deliberately distorting his answers or is not paying attention to the task at hand.

It should always be kept in mind that most personality and interest inventories

were developed primarily for clinical and counseling use. In those settings one usually assumes that the patient or advisee has a vested interest in responding correctly—after all, the test information is going to be used for his benefit. He will therefore have little situational reason to motivate him into false responding. When such tests are used to determine the future of a person, as when they are given as part of the decision process in a selection setting, the motivation to distort responses will substantially increase. The testee will no longer be as concerned with the "true" response (his real attitude)—instead, he will be concerned with making that response which is most apt to be the one for which the "establishment" is looking. He believes that if he does not follow such a strategy in making his responses he is likely to hurt his chances of getting the position. His motivation to falsify, therefore, is apt to be in direct proportion to the extent to which he wants that particular position, and to this extent he "fakes" when he thinks he should answer not the way he is but the way he thinks he should be to get the job.

TYPES OF RESPONSE STYLES

The tendency deliberately to bias one's responses in a systematic fashion is usually called a *response set* or *response style*. There are many kinds of response sets, although perhaps the two most often of concern are *social desirability* and *acquiescence*.

Social Desirability refers to a general tendency to respond in the manner felt to be most socially desirable in that particular situation. The tendency will be reflected by the person responding to personality attitude and interest items in terms of how acceptable their responses are socially (that is, in the eyes of others).

Acquiescence refers to a general tendency to respond as we think others want us to respond. This tendency will be reflected by the person attempting to give the responses he thinks the test giver is hoping to receive.

All of us have let both types of response styles influence our behavior in daily living. For example, flattery is a fine example of a response style, as is politeness in many cases. Under certain conditions both of the above styles may lead to highly similar responding, while at other times they could be quite conflicting in their impact upon behavior. Rorer (1965) has pointed out that much of the currently available evidence on the two types of response styles indicates that they are not completely independent. Rather, acquiescence may only be an important consideration when the stimuli (items) are of medium social desirability.

REDUCING RESPONSE BIAS

Some tests are specifically designed to eliminate the possibility of faking. Probably the best-known procedure for doing so is the "forced-choice" method. Such inventories present the respondent with several alternatives from which he must select the one he most agrees with or is most like. These alternatives have been paired on the basis of being equally favorable or desirable in the eyes of a large number of individuals but only one response is the "correct" or discriminating answer. The hope is that the choice between alternatives will be discriminative and not based upon "how good will it make me look if I respond to this one or that one." Since response alternatives will make him "look good" to the same degree,

other factors than "looking good" must enter into the choice of an answer.

A second and less well-known but perhaps more promising procedure for controlling faking is the "error-choice" method. Chapters 7 and 9 cover both of these procedures.

STUDIES ON RESPONSE BIAS

Of particular interest to the industrial psychologist are those studies which relate to response bias in a selection situation.

BORISLOW STUDY (1958) In this study a group of students took the Edwards Personal Preference Schedule (EPPS) under standard self-assessment conditions. They were then divided into three matched groups for a retest two weeks later. Group A took the EPPS a second time under conditions identical to the initial testing session. Group B received the instructions to respond as a "perfect" person would respond. Group C was told to respond in terms of what they would "like to be" rather than in terms of what they actually were. Test-retest correlations ranged from 0.65 to 0.91 for the people in Group A, from 0.03 to 0.68 for those in Group B, and from −0.03 to 0.68 for group C. These values indicated that the people in groups B and C responded under the new instructions a great deal less like their original performance than did the controls, indicating that the EPPS is susceptible to deliberate falsification.

BRIDGMAN AND HOLLENBECK STUDY (1961) The test under investigation by these authors was the Kuder Preference Record, Occupational Form. Four groups filled out the inventory, and each group received a different set of instructions.

Group	*Instructions*
A	As if applying for the job of sanitary supply salesman
B	As if applying for the job of industrial psychologist
C	As if applying for an unspecified job in industry
D	As if tests were being taken for purposes of vocational counseling

All inventories were then scored using a scale developed for sanitary supply salesmen and the Kuder industrial psychologist scale. Verification scores (Kuder index of response bias) were also obtained.

The results were quite interesting. As Table 4.9 shows, the scores of the A and B groups did not differ significantly from the norm groups they were impersonating.

TABLE 4.9 *Mean Occupational Interest Scales for Each of the Four Groups*

Instructions	*Salesman Key*	*Psychologist Key*	*Verification Key*
Salesman	71.5	48.7	46.8
Psychologist	61.4	53.2	45.5
Typical job	64.8	48.9	48.9
Vocational counseling	61.8	44.6	51.1
Actual salesman (norm)	73.1	—	—
Actual psychologist (norm)	—	54.5	—

SOURCE: Adapted from C. S. Bridgman and G. P. Hollenbeck. Effect of simulated applicant status on Kuder Form D occupational interest scores. *Journal of Applied Psychology*, 1961, *45*, 237–239.

Also, the highest scores on each key were obtained by that group which had been instructed to assume that particular set of responding. The verification scores for all fake groups were higher than the control, indicating that it might be of use in identifying biased test results, although its actual magnitude was of questionable practical value in all cases except with the salesman group.

KIRCHNER'S STUDIES Both of the above studies indicate that faking can take place in a *laboratory* setting. Kirchner (1961, 1962) has reported two instances in which clear evidence for response faking was found in an actual selection situation. In his first study, responses made on the Strong Vocational Interest Blank for 92 retail and 64 industrial sales applicants (later hired) as part of the selection procedure were compared with responses made by 68 retail and 49 industrial salesmen who had been employed at least five years and who had completed the SVIB voluntarily as part of a concurrent validity study. Of 96 mean differences on the 48 scales, 32 were significant at the 0.05 level. Both the retail and the industrial applicants tended to score higher in social service and business occupations and lower in technical-scientific and, surprisingly, sales. In general, applicants indicated a greater liking for things than did employed salesmen, which suggests that they were completing the SVIB in the most socially acceptable fashion: liking much, disliking little.

Kirchner's second study was much the same except that this time he studied the Edwards Personal Preference Schedule. The major results are shown in Table 4.10.

TABLE 4.10 *Mean Scores and Standard Deviations Obtained by Retail Sales Hires and Retail Salesmen on EPPS Scales with Mean Differences*

	Retail Sales Hires (applicants) N=97	*Retail Salesmen* N=69	
EPPS Scale	*Mean*	*Mean*	*Mean Difference*
Achievement	17.28	16.91	0.37
Deference	12.63	13.30	−0.67
Orderliness	11.33	12.90	−1.57*
Exhibition	16.40	15.77	0.63
Autonomy	10.65	11.25	−0.60
Affiliation	15.21	14.29	0.92
Intraception	15.44	13.90	1.54*
Succorance	8.15	8.42	−0.27
Dominance	21.77	19.88	1.89*
Abasement	10.62	10.23	0.39
Nurturance	12.94	12.88	0.06
Change	15.66	16.35	−0.69
Endurance	15.56	15.91	−0.35
Heterosexuality	14.29	16.57	−2.28*
Aggression	11.92	11.52	0.40

* Significant at 0.05 level of probability or better.

SOURCE: W. K. Kirchner. "Real-life" faking on the Edwards Personal Preference Schedule by sales applicants. *Journal of Applied Psychology*, 1962, *46*, 128–130.

The EPPS scores for 97 retail sales applicants and 66 industrial sales applicants (all later hired) were compared to those of scores of 69 retail salesmen and 49 industrial salesmen (all tested on the job). Results showed that the retail applicants tended to score significantly higher than retail salesmen on the orderliness, intraception, and dominance scales and lower on the heterosexuality scale. There was no significant differences, however, between industrial applicants and industrial salesmen. Kirchner suggests that persons more oriented toward selling in terms of interests and personality (thus, retail sales applicants) are more likely to distort answers to the EPPS.

TESTING AND CIVIL RIGHTS

Testing has remained relatively unaffected by state or Federal legislation throughout the years. Certainly, laws such as the Wagner Act of 1935 affected the social climate in such a way that testing was more easily adopted by employers as a selection instrument. However, no direct legislation dealing with testing has appeared on either the state or national level.[5] Psychologists have attempted to maintain this state of affairs by an active system of self-regulation through a code of ethics and by administering professional sanctions to ethics violators. In keeping their own house clean, they hope to avoid the need for outside regulation such as is found in medicine and law.

The recent social dynamism in America referred to as the civil rights movement has brought industrial testing into a rather precarious position—one which has caused a great deal of concern to the profession of psychology. At the crux of the matter is the extremely complex question of whether psychological tests can be, and in fact are, discriminatory in nature. Since the Federal Civil Rights Act (Title VII) and many similar state civil rights acts, laws have clauses which evaluate the hiring, firing, or promotion of employees in terms of whether or not there was any "discriminatory intent" involved. It may eventually become necessary to demonstrate to the satisfaction of the courts that all tests used for such purposes are indeed nondiscriminatory.

MYART VERSUS MOTOROLA

One such instance involving industrial testing has already occurred, the case of Myart versus Motorola (1964) in the State of Illinois. In this instance Myart brought suit against the Motorola Corporation through the Illinois Fair Employment Practices Commission (FEPC), claiming that he had been discriminated against because of race during the hiring process. The case was originally heard by a FEPC examiner who ruled that Motorola should (1) cease using their low-level screening test because it did not equate for inequalities and environmental factors of culturally deprived groups and (2) hire Myart. This examiner's decision was based upon his evaluation of testimony concerning the extent to which the screening test was discriminatory in nature.

[5] The only exception would be those states where licensing or certification legislation includes a statement about testing.

When the case was reviewed by the full commission, the ruling was changed somewhat. The FEPC did not feel the problem of test discrimination was at question. Rather, the commission was more concerned with whether Myart's test score had been changed by the company so that he would fall below cutoff. (Myart's original complaint did not allege the test to be discriminatory.) However, the Commission stated it "did not ignore the possibility that tests of this nature are inherently discriminatory against persons alien to the predominant middle class white culture." The Commission's final decision was to (1) award Myart damages and (2) not require that he be hired by Motorola.

The immediate and the eventual impact of discrimination in testing is hard to determine. Certainly it is a complex problem. There is no question that many of our tests *do* favor those raised in a white middle-class culture. The normative data for these people tends to be higher than for other groups. There is always the question of cause and effect. Do white middle-class people score higher because they indeed have *more* of what is being measured than do the other cultural groups, or do they score better because the test "favors" them? This is a very difficult question to answer. Psychologists have made numerous attempts to develop truly "culture free" tests in such areas as intelligence and have achieved only a modicum of success.

If carried to extremes, the discrimination problem can become ludicrous. There are many tests which show clear-cut sex differences. Suppose a circus were to use a job-sample test requiring the lifting of heavy weights in selecting its strongman. Such a test would clearly discriminate against women, yet the right of the circus owner to establish a minimum cutoff value would not be questioned. One might

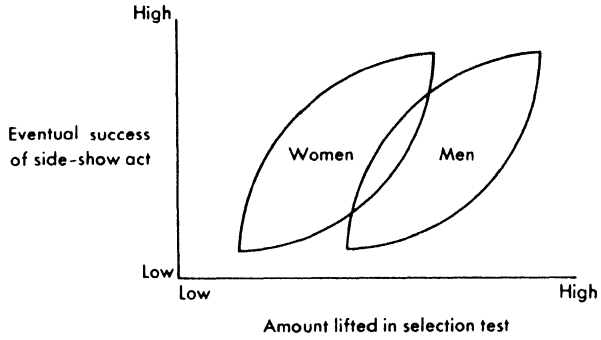


Figure 4.3. Hypothetical scatterplot of results of weight-lifting test given to men and women.

ask what would occur if a woman were found who could lift much more than a normal woman, even though not as much as a very strong man. Should the circus hire her and change the act to "the world's strongest woman"? This, of course, is one way of handling the discrimination problem and involves using different norms, standards, and cutoffs for each group. In Chapter 3 terminology, sex becomes a moderator variable in that a separate validity coefficient would be computed for each subgroup. A hypothetical scatterplot of the situation might look like Figure 4.3.

There is a difficulty in using separate norms (or equation groups by a transformation such as the Z score) to remove group differences. In most hiring situations it is now ***illegal*** to obtain information on race during the employment process.

Thus, any application blank or test booklet may not indicate the applicant's race. It is therefore impossible at any later time to develop different norms for the separate groups in order to equate for cultural factors. Suffice it to say that this is an area in industrial psychology which will be interesting to keep tabs on in the next few years.

REFERENCES

American Psychological Association. Ethical standards of psychologists. *American Psychologist*, 1959, *14*, 279–282.

American Psychological Association. Technical recommendations for psychological tests and diagnostic techniques. American Psychological Association, Washington, D.C., 1954.

Anastasi, Anne. *Psychological testing*, 2nd ed. Macmillan, New York, 1961.

Bennett, G. K. *Test of mechanical comprehension.* Psychological Corporation, New York, 1940.

Black, H. *They shall not pass.* Morrow, New York, 1963.

Blum, M. L. A contribution to manual aptitude measurement in industry. *Journal of Applied Psychology*, 1940, *24*, 381–416.

Blum, M. L. Selection of sewing machine operators. *Journal of Applied Psychology*, 1943, *27*, 35–40.

Blum, M. L., and B. Candee. The selection of department store wrappers with the aid of certain psychological tests. *Journal of Applied Psychology*, 1941(a), *25*, 76–85.

Blum, M. L., and B. Candee. The selection of department store wrappers and packers with the aid of certain psychological tests Study No. 2. *Journal of Applied Psychology*, 1941(b), *25*, 291–299.

Borislow, B. The Edwards personal preference schedule and fakability *Journal of Applied Psychology*, 1958, *42*, 22–27.

Bridgman, C. S., and G. P. Hollenbeck. Effect of simulated applicant status on Kuder Form D occupational interest scores. *Journal of Applied Psychology*, 1961, *45*, 237–239.

Brown, C. W., and E. E. Ghiselli. The relationship between the predictive power of aptitude tests for trainability and for job proficiency *Journal of Applied Psychology*, 1952, *36*, 370–377.

Buros, O. K. (ed.). *The fourth mental measurements yearbook* Gryphon Press, Highland Park, N.J., 1953.

Buros, O. K. (ed.). *The fifth mental measurements yearbook.* Gryphon Press, Highland Park, N.J., 1959.

Buros, O. K. (ed.). *The sixth mental measurements yearbook.* Gryphon Press, Highland Park, N.J., 1965.

Candee, B., and M. Blum. Report of a study done in a watch factory. *Journal of Applied Psychology*, 1937, *21*, 572–582.

Clarke, W. V. *A manual for activity vector analysis.* Clarke Associates, Barrington, R I , 1953.

Crawford, J. E., and Dorothea M. Crawford. *Small parts dexterity test.* Psychological Corporation, New York, 1946.

Cronbach, L. J. *Essentials of psychological testing.* Harper & Row, New York, 1949 (2nd ed., 1960).

Datta, Lois-Ellen. Remote associates test as a predictor of creativity in engineers. *Journal of Applied Psychology*, 1964, *48*, 183.

Dorcus, R. M., and M. H. Jones. *Handbook of employee selection.* McGraw-Hill, New York, 1950.

Dunnette, M. D., and W. K. Kirchner. Validities, vectors, and verities. *Journal of Applied Psychology*, 1962, *46*, 296–299.

Flanagan, J. C. *Factor analysis in the study of personality*. Stanford University Press, Stanford, Cal., 1935.

Gaudet, F. J., and A. R. Carli. Why executives fail. *Personnel Psychology*, 1957, *10*, 1.

Gellerman, S. W. Seven deadly sins of executive placement. *Management Review*, 1958, *47*, 4–9.

General aptitude test battery. *Labor Information Bulletin*, May 1947, 16.

Ghiselli, E. E. *The measurement of occupational aptitude*. University of California Press, Berkeley, 1955.

Ghiselli, E. E. Tests for the selection of inspector-packers. *Journal of Applied Psychology*, 1942, *26*, 468–476.

Ghiselli, E. E., and R. P. Barthol. The validity of personality inventories in the selection of employees. *Journal of Applied Psychology*, 1953, *37*, 18–20.

Gross, M. L. *The brain watchers*. Random House, New York, 1962.

Guilford, J. P., and W. S. Zimmerman. *The Guilford-Zimmerman temperament survey*. Sheridan Supply Co., Beverly Hills, Cal., 1949.

Guion, R. M. *Personnel testing*. McGraw-Hill, New York, 1965.

Hathaway, S. R., and J. C. McKinley. *Minnesota multiphasic personality inventory*, rev. ed. Psychological Corporation, New York, 1943.

Hay, E. N. Predicting success in machine bookkeeping. *Journal of Applied Psychology*, 1943, *27*, 483–493.

Hoffman, B. *The tyranny of testing*. Crowell-Collier, New York, 1962.

Humm, D. G., and G. W. Wadsworth. *The Humm-Wadsworth temperament scale*, 2nd rev. Doncaster G. Humm Personnel Service, Los Angeles, Cal., 1935.

Jones, F. E. Predictor variables for creativity in industrial science. *Journal of Applied Psychology*, 1964, *48*, 134–136.

Jones, M. H. The adequacy of employee selection reports. *Journal of Applied Psychology*, 1950, *34*, 219–224.

Kirchner, W. K. "Real-life" faking on the Edwards personal preference schedule by sales applicants. *Journal of Applied Psychology*, 1962, *46*, 128–130.

Kirchner, W. K. "Real-life" faking on the Strong vocational interest blank by sales applicants. *Journal of Applied Psychology*, 1961, *45*, 273–276.

Kuder, G. F. *Kuder preference record—occupational*. Science Research Associates, Chicago, 1956.

Kuder, G. F. *Kuder preference record—vocational*. Science Research Associates, Chicago, 1934.

Likert, R., and W. H. Quasha. *Revised Minnesota paper form board test*. Psychological Corporation, New York, 1941–1948.

Locke, E. A., and C. L. Hulin. A review and evaluation of the validity studies of activity vector analysis. *Personnel Psychology*, 1962, *15*, 25–42.

Meranda, P. F., and W. V. Clarke. The predictive efficiency of temperament characteristics and personal history variables in determining success of life insurance agents. *Journal of Applied Psychology*, 1959, *43*, 360–366.

Meranda, P. F., and W. V. Clarke. Rejoinder to "Validities, Vectors, and Verities" by Marvin D. Dunnette and Wayne K. Kirchner. *Journal of Applied Psychology*, 1962, *46*, 300–302.

Mullins, C. J. Prediction of creativity in a sample of research scientists. WADC Technical Documents Report PRL, Lackland Air Force Base, Texas, 1959.

Murray, H. A. *Thematic apperception test*. Harvard University Press, Cambridge, Mass., 1943.

Myart versus Motorola (a summary report). Fair Employment Practices Commission, State of Illinois, 1964.

Neel, R. G., and R. E. Dunn. Predicting success in supervisory training programs by the use of psychological tests. *Journal of Applied Psychology*, 1960, *44*, 358–360.

Otis, A. S. *Otis self-administering tests of mental ability.* World Book Co., Tarrytown-on-Hudson, N.Y., 1922–1929.

Paterson, D. G. *Minnesota mechanical assembly test.* Stoelting, Chicago, 1930.

Paterson, D. G. *Minnesota spatial relations test.* Stoelting, Chicago, 1930.

Personnel men view testing programs. *Industrial Relations News,* 1959, 9(38), 4.

Rorer, L. G. The great response style myth. *Psychological Bulletin,* 1965, 63, 129–156.

Rorschach, H. *Psychodiagnostics: A test based upon perception* (trans. by P. Lemhau and B. Kronenburg). Huber, Bern, Switzerland, 1947. (U.S. distributor, Grune and Stratton, 1941.)

Sartain, A. Q. The use of certain standardized tests in the selection of inspectors in an aircraft factory. *Journal of Consulting Psychology,* 1945, 9, 234–235.

Shuman, J. T. The values of aptitude tests in the aircraft engine and propeller industry. *Journal of Applied Psychology,* 1945, 29, 156–160.

Stagner, R. The gullibility of personnel managers. *Personnel Psychology,* 1958, *11,* 347–352.

Strong, E. K., Jr. *Strong vocational interest blank for men, revised.* Stanford University Press, Stanford, Cal., 1938.

Taylor, C. W., W. R. Smith, B. Ghiselin, and R. Ellison. Explorations in the measurement and prediction of contributions of one sample of scientists. ASD-TR-61-96, Lackland Air Force Base, Texas, April 1961.

Taylor, E. K., and E. C. Nevis. The use of projective techniques in management selection. *Personnel,* 1957, 33, 462–474.

Thurstone, Thelma G., and L. L. Thurstone. *Thurstone test of mental alertness.* Science Research Associates, Chicago, 1943.

Tiffin, J. *Purdue pegboard.* Science Research Associates, Chicago, 1941.

Tiffin, J., and C. H. Lawshe, Jr. *The adaptability test.* Science Research Associates, Chicago, 1942.

Wagner, E. E. Predicting success for young executives from objective test scores and personal data. *Personnel Psychology,* 1960, *13,* 181–186.

Whyte, W. H. *The organization man.* Simon and Schuster, New York, 1956.

Wonderlic, E. F., and C. I. Hovland. The personnel test: A re-standardized abridgement of the Otis S. A. test for business and industrial use. *Journal of Applied Psychology,* 1939, 23, 685–702.

5 THE INTERVIEW AND OTHER SELECTION DEVICES

The previous chapter discussed psychological tests as a method of assessing and selecting employees. We mentioned, however, that there are other selection techniques. These are generally *in addition to* tests, but sometimes are used without tests. The most frequently used of these other selection methods is the interview. Nevertheless, such aids as letters of recommendation, application blanks, ratings by former supervisors, and records of past performance are also used as adjuncts in selecting job applicants. This chapter will examine each of these techniques in terms of their utility and validity as part of the selection process.

INTERVIEWS

Interviews used in industry differ according to their objectives. Some of the different objectives are:

1. To evaluate the candidate for purposes of hiring, promotion, or transfer. Interviews conducted for such purposes are usually referred to as *selection* or *placement interviews.*

2. To obtain information from the interviewee concerning his attitudes about his job, the company, etc. These interviews are referred to as *attitude interviews.* In many companies they are carried on routinely with all employees during employment or at the point of termination. The latter type interviews are called *exit* or *separation interviews.* Their purpose is to understand the person's attitude better, even though the decision to terminate has already been made. In such situations it is felt the interviewee will be more candid since he may no longer feel that he must "please" to keep his job. Surprisingly, few individuals refuse to cooperate, probably because of the need to "get it off his chest."

3. To help the employee with any particular personal or on-the-job problem which might be harmful to his performance both on and off the job. These interviews are referred to as *counseling interviews.* Many larger firms provide services of this type for their employees, both as an individual service and in the hope that some benefit will accrue back to the company. In recent years this type of interviewing has become more frequent for top executives. The assumption is that executives often operate as "islands unto themselves" during the business day

and are not able or willing to discuss their many problems and responsibilities with associates. The interview in this setting can become a therapeutic device allowing the individual to "unload" to a person who is trained to view things objectively and who will not be threatening because he has no authority to change things. While lofty in its objective, there is little empirical evidence to prove that such interviews do serve the intended purpose. However, by intuitive judgment, the idea seems to have some merit, and those who work in this area extol its virtues. If nothing else, it may save the wife and children from having to cope with a grouchy and tense husband or father returning home from a frustrating day at work, at least on those days when the counseling interview took place.

4. To evaluate the interviewee for his ability to perform under a particular set of difficult circumstances. These interviews are often called *assessment* or *stress interviews*. The interviewer deliberately provokes stressful situations with the hope of measuring a person's ability to perform under such conditions. Such interviews were first popularized by the Office of Strategic Services in World War II, and variations of the stress interview are being used more and more often in industry. For a fascinating account of the O.S.S. experiences with the stress interview, the reader is referred to the book *Assessment of Men* (1948).

Our present concern is primarily with the selection interview. Thus, we will examine how the interview is used in the selection process, how it compares with other selection devices, and its strengths and weaknesses.

THE SELECTION INTERVIEW

In spite of the widespread increase in the use of tests as part of the selection process, testing ranks second in terms of the most frequently employed selection methodologies. Firmly established in first place and highly unlikely to tumble from this rank is the selection interview. Rarely, if ever, do people become employed today without having experienced some type of employment interview with one or more company representatives who do or do not know anything about conducting interviews. The selection interview is used for jobs ranging from itinerate fruit pickers in California to executive vice-presidents and presidents. No one is immune from the selection interview.

For some jobs and in some companies, as many as eight or ten such interviews occur with different people prior to being accepted or rejected for the given position. In fact, at one university with which one of the authors is particularly familiar, it is standard procedure for all candidates for a tenured job on the faculty to have a full day of interviews with different people such as the departmental chairman, the college dean, the graduate school dean, the university vice-presidents for research and instruction, and several of the senior faculty members of the department. Needless to say, the candidate is usually a groggy man by the time he has completed a cycle of such interviews. The assumption of multiple interviews is based on the value of pooled judgments. It may also, we should point out, involve sharing the risk.

Just how extensive is interviewing in industry? Several surveys have indicated the use of the selection interview to be almost universal. Spriegel and James (1958)

report data from both a 1930 survey of 236 firms and a 1957 survey of 852 firms. In the earlier survey, 93 percent of the firms reported using an interview as part of their selection process; by 1957 this had increased to 99 percent. Exactly how many interviews are conducted each year is a matter of pure conjecture, although a guess might place the number at close to a quarter of a billion! With so much time and money spent in interviewing, one should expect more knowledge about the interview than now exists—at least in terms of research efforts and conclusions.

LENGTH OF THE SELECTION INTERVIEW

There is no standard time limit associated with interviews for selection purposes. The length varies depending on the type of position, number of applicants, importance attached to the interview as a selection device, and personality of the interviewer. Our experience indicates that many employment interviews are conducted by untrained "executives" who believe their skill in interviewing to be exemplary.

In an early study based upon only seven employment interviews, Uhrbroch (1933) found the interview duration to be approximately twelve minutes, with the interviewer doing most of the talking. A somewhat more recent study by Daniels and Otis (1950) found the average interview length to be ten minutes. The interviewer spoke about 57 percent of the time, the applicant spoke 30 percent, and 13 percent was spent in silence. The majority of these interviews (conducted in eight different factories) were for factory jobs, but others were for office jobs. The few studies reported in the literature do not allow for a generalization about the relation between length of interview and interview value. In many interview situations, the length is predetermined by an appointment schedule. College recruitment settings are typical of this fixed time basis. As will be pointed out in a later section, one important but as yet unanswered question concerns the effect of interview length on the validity and reliability of the interview (Mayfield, 1964).

In spite of its extensive use in employment situations, the interview has not received a great deal of attention concerning its reliability, validity, and how these attributes are affected by the type or length of the interview itself. However, before examining the data that are available concerning these questions, it will be helpful to get a clearer picture of the role the interview plays as part of the selection process.

A MODEL OF THE SELECTION INTERVIEW

The interview often is difficult to evaluate as a selection device simply because of the somewhat complex fashion in which it fits into the selection process. First, it must be remembered that the basic purpose of the interviewer in any assessment interview situation is twofold: (1) to gather as much data as one can which are relevant to the selection decision, and (2) to evaluate the data available and decide to select or reject the applicant.

The data collected during an interview depend upon the "type" of interviewer, his degree of training, his biases, and the position he is attempting to fill. Thus data can be collected very systematically through the use of a highly structured interview session in which a planned series of questions is presented to the applicant. On the other hand, data can be gathered in a much more open or unstructured interview in which one simply probes and explores the applicant's

qualifications in a "play-by-ear" fashion. Similarly, the process by which one combines interview information into a single decision is also often open to numerous options on the part of the interviewer.

Perhaps the best way to view an interviewer is to envision that he acts as a sort of intuitive regression equation. His job is one of attempting to collect, combine, and weigh the interview data in such fashion that he comes up with a prediction of how successful the applicant may be should he be hired. The degree

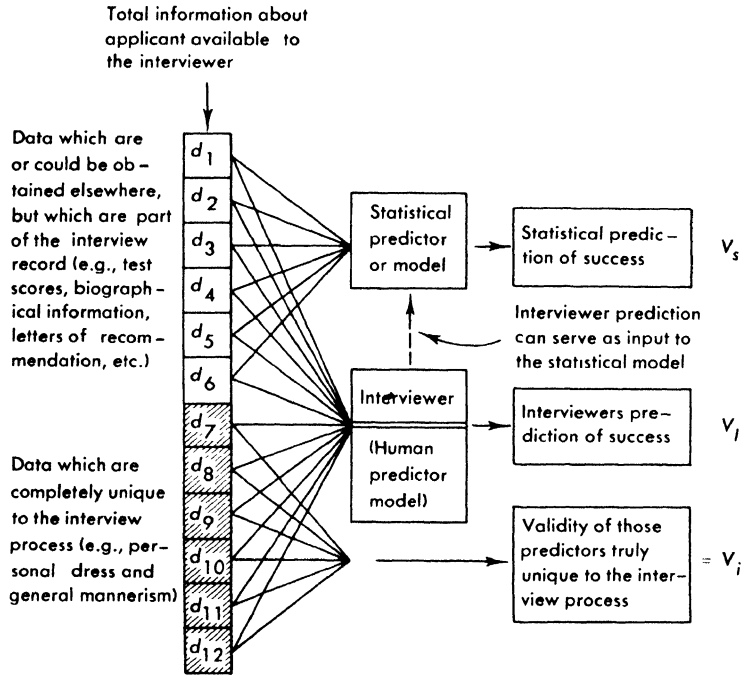


Figure 5.1. A model of the selection interview process.

to which he can successfully predict eventual job success is his *validity*. From a theoretical viewpoint it should be simple to evaluate empirically the validity of the interview both in the absolute sense and in terms of its relative value to other selection procedures. Unfortunately, as Figure 5.1 depicts, such an evaluation is not as easy as it might superficially appear.

DATA SOURCES IN THE INTERVIEW

Bellows and Estep (1954), in a valuable book on the use of the interview in industry, were the first to point out some very basic distinctions concerning the data involved in the interview. They classified data into two categories depending upon the source of the information. In the first category are data from sources not directly tied to the interview situation itself. These auxiliary sources include letters of recommendation, application blanks, psychological tests, etc. While one *can* and often does obtain such information in the interview, it is usually more conveniently obtained via other sources and is frequently provided to the interviewer prior to or concurrent with the actual interview in the hope that such information will be of

help in the interviewing process. The second major category of data is information obtained in the face-to-face situation which is unique to the interview itself. Examples might be data about the dress and mannerisms of the applicant, his speech, his way of handling himself in a social situation, and a variety of other bits of information which are apt to emerge during the interpersonal situation afforded by the interview. Both kinds of data are shown in Figure 5.1.

Personnel selection procedures often utilize the first type of data in a formal prediction model, such as the multiple regression model or profile model discussed in Chapter 3. Thus one can at least conceptualize a validity, V_s, which represents the *statistical* utility of these kinds of data when used to predict job success. In a similar fashion, one can conceptualize a second statistical validity, V_i, which represents the utility of the kinds of data which can only be obtained in the interview in predicting job success. This validity, since it is statistical, represents a potentially objective form of predictive information derived from the interview information. However, it is strictly conceptual in nature, since in practice it is almost impossible to identify, measure, and quantify the different informational cues which manifest themselves in the face-to-face situation of a selection interview.

In addition, we must consider the validity of the *interviewer*, V_I. After considering data 1–6 (the noninterview information about the candidate) and after being exposed to the kinds of cue information exclusive to the interview (data classes 7–12) the interviewer makes a global decision about the "likely success" of the applicant. The degree to which his estimate of success actually correlates with eventual job performance is a measure of his validity as an interviewer. It is *not* an indication of the validity of the interview *itself*, which is V_i. Note that even if we exclude all noninterview information from the interviewer, so that he must base his prediction solely upon cues obtained in the actual interview, we are still assessing *his* validity rather than the validity of the interview information.

This confusion about *interview* versus *interviewer* validity must be emphasized. (The authors are sorry to emphasize confusion at this point but it seems necessary.) Interview information is always channeled through the interviewer, and it is the job of the interviewer, almost by definition, to weigh the informational cues according to his best judgment and arrive at a decision about the applicant. Therefore, when one attempts to evaluate the effectiveness of the interview, one tends to evaluate the ability of the interviewer to collect relevant information and combine it in the most accurate fashion. That is, one tends to evaluate the interviewer's ability to act as a statistical prediction equation instead of evaluating the interview per se.[1] This fact has usually been completely ignored by most individuals who have attempted to ascertain the extent to which the interview can contribute to accurate selection.

CHARACTERISTICS OF THE SELECTION INTERVIEW

Before considering the validity studies, we shall review certain general characteristics of the interview. For example, it must be kept in mind that the interview is a verbal and visual interaction between two individuals. Thus many of the cues available to the interviewer will be based upon the language and appearance of the interviewee. The basic structure of an interview is "conversational"; in fact, one definition of the interview is that it is simply "a conversation with a purpose"

[1] For an interesting discussion of clinical versus statistical prediction see P. E. Meehl. *Clinical versus statistical prediction*. University of Minnesota Press, Minneapolis, 1954.

(Bingham and Moore, 1941). The task of the interviewer is to use this conversational tool to elicit as much relevant information as possible (and generally within a specified time span). It is therefore reasonable to assume that the conversational skill of the interviewer can be an important variable in the success of any interview.

It is also important to realize that since the interview is an interpersonal relationship, the behavior of the interviewer can have a very dramatic effect on the behavior of the applicant. The applicant seeks cues of his own concerning "how well he is doing" or what the interviewer "thinks of him." These cues serve as a kind of feedback or reinforcement to the applicant as shown in Figure 5.2, and he will often respond differently depending upon whether he perceives the cues as being favorable or unfavorable. Dramatic evidence of this kind of influence on perform-

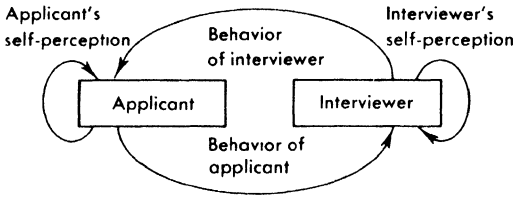


Figure 5.2. The circularity of the behavior pattern in an interview situation.

ance in a two-person situation has been provided by Verplank (1955) and Greenspoon (1955). Both have shown that the frequency of particular kinds of verbal responses made by the interviewee can be dramatically influenced by having the interviewer show approval for these kinds of responses.

In the Verplank study the interviewers were instructed to respond to any statement made by the subject which began with "I think . . . ," or "I believe . . . ," etc. In other words, any statement of opinion was to be reacted to by one of four different methods:

(1) *Agreement.* The interviewer would say "you're right," "I agree," "that's so," or something similar. In cases where he did not want to interrupt, he would nod his head and smile.

(2) *Paraphrasing.* The interviewer would repeat back to the subject the opinion statement just made.

(3) *Silence.* The interviewer refrained from any comment or action at all.

(4) *Disagreement.* The interviewer by statement or action would show disagreement with every statement of opinion.

Each interview was taped and analyzed for the number of opinion statements made. Significant differences in the numbers of these statements were found as a function of the interviewer behavior. The agreement and paraphrasing conditions increased the number of opinion responses above base rate, with agreement being most effective. Silence and disagreement, on the other hand, decreased the number of statements below base rate, with disagreement being most effective.

The fact that the interviewer can noticeably influence applicant behavior is important to the degree to which the interviewer lets his behavior differ from applicant to applicant or to the degree that interviewers differ from each other. All too often the interviewer is inclined to forget that, like other selection devices, he should be "standardized"—he should attempt to hold his behavior as constant from applicant to applicant as is possible. This, of course, is one argument for having highly structured selection interviews which permit little individuality on the part of the interviewer. Unfortunately, such highly structured situations prevent

the interviewer from exploring certain avenues of information which arise and might be fruitful data sources. In fact, some argue that to restrict the flexibility of the interviewer involves sacrificing the basic interpersonal advantage that the interview has over other selection devices, and that one might just as well scrap the interview and substitute a paper-and-pencil questionnaire if it is structure that is desired.

Symonds (1939) has listed a series of factors which he believes can influence the quality and quantity of data collected in the interview:

Factors Inherent in the Applicant

Age
Intelligence
Sex
Race
Socioeconomic level
Language ability
Emotional need
Emotional security
Applicant's attitude toward interviewer
Applicant's previous acquaintance with interviews
Purpose of applicant in coming to the interview

Factors Inherent in the Interviewer

Age
Sex
Intelligence
Race
Socioeconomic level
Position or authority with reference to the applicant
Personality (social warmth, sympathy, out-goingness, human interest, vitality)
Social outlook
Psychological understanding (his ability to sense or feel the purposes, needs, or drives of the applicant)
Previous acquaintance with the applicant
Use interviewer can be to the applicant
Interest of interviewer to applicant (degree to which interviewer enjoys the conversation of interviewer)

Factors in the General Situation Where the Interview Is Conducted

Place
Time
Persons present
First, second, or subsequent interview
Experiences of applicant directly preceding interview
Emergency character of interview
Directions given applicant preceding the interview
Voluntary versus nonvoluntary nature of interview

Factors in the Form and Content of the Interview

Content of questions
Form of questions
Interpretation, suggestions, or other reactions of interviewer

Telling applicant purpose of the interview
Relieving applicant as to the identity of the interviewer
Encouragement given by interviewer
Remarks interpolated by examiner during the interview

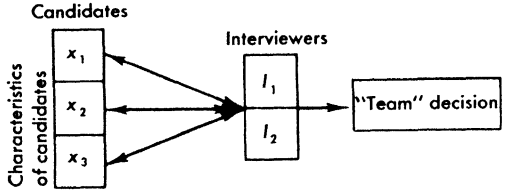
As can be seen, this is quite an extensive list. However, in reading through it one can think of additional factors which have not been included. In fact, it might be easier to make a list of things which could *not* influence the results of the interview.

FORMAT OF THE SELECTION INTERVIEW

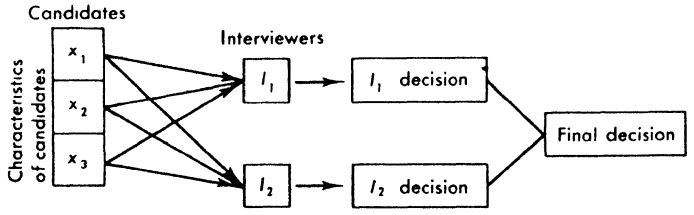
We mentioned previously that interviews differed in terms of objective and format. Although the different objectives were outlined, format differences were not explained. The reason for this was that it seemed more appropriate to discuss format differences within the specific context of the selection interview.

GROUP VERSUS INDIVIDUAL INTERVIEW Under certain circumstances the interview may involve more than two participants. A typical example would be a military board of review or a promotion board of several officers who would interview the candidates individually. It is also possible for a single interviewer to process several interviewees at the same time, although this is perhaps less frequent and probably occurs only when circumstances make it impossible to interview each person alone. Virtually no data are available concerning the efficiency, validity, and reliability of group interviews relative to individual interviews. This is really a very interesting problem, and it is hard to understand why more attention has not been devoted to it. For example, consider a simple situation where two interviewers are assigned to examine each new applicant. Two primary procedures might be suggested for their interviewing strategy.

First, both interviewers could examine each candidate together, working as a team and arriving at a joint decision as to hire or not. This might be diagramed as:



Second, each interviewer could examine each candidate separately and then only hire those whom both interviewers agreed should be hired (multiple cutoff approach) or those whose composite was above some minimum (multiple regression approach). This procedure could be diagramed as:



This strategy provides for assessing the validity of each interviewer separately plus evaluating the validity of the final decision.

Questions of primary importance in comparing interviewing strategies are (1) is the validity of the "team" decision higher than that of the combined final decision based upon their individual predictions and (2) how does the validity of a "team" decision relate to the separate validities of the respective interviewers? We will discuss this matter further in the chapter on team performance, but the question is both so important and so intriguing that it seems worthwhile to call it to the reader's attention.

STRUCTURE OF INTERVIEW When the format of an interview is mentioned, one most often refers to the *structure* of the interview situation. Structure in turn refers to the degree of formalization and stereotypic behavior of the interviewer. Highly structured interviews are sometimes referred to as *patterned interviews* (for example, see McMurray, 1947).

In a totally structured interview the questions and their sequence are predetermined in advance. The interviewer merely proceeds systematically through the list of questions, noting the replies given by the applicant. A sample interview guide which might be used by an interviewer is shown below.

Patterned Interview Guide

Name

1. What is your age?
2. What is your educational background?
3. Are you married?
4. What is the size of your family?
5. What previous work experience have you had?
6. What type of work do you desire?
7. What wages do you wish to earn?
8. What are your drinking habits?
9. What is your current financial situation?
10. Do you like working closely with others?

The reader may have noticed the similarity between a highly structured interview and the application blank. In fact, the interviewer with no leeway in asking questions might just as well be dispensed with and application blank substituted which asks the same questions in written form. The major difference is who records the answers, the interviewer or the applicant.

At the other end of the structure continuum we have those "free-wheeling" interview situations where the interviewer has no preplanned interview strategy at all. He simply proceeds to question the applicant along those lines which appear most likely to produce critical information at a given moment. A totally unstructured interview closely resembles a clinical interview in terms of format, but varies in the nature and type of question asked.

Proponents exist for both the structured and the unstructured interview. Advocates of highly formal interview situations point out that unless you have a consistent pattern from one applicant to the next it is *not* a standardized selection procedure. One applicant may therefore be forced to undergo one type of evaluation while a second applicant may experience a quite different interview evaluation. In reply, those in favor of unstructured interviews point out that if one wants complete

structure, why not use a questionnaire? In addition, they argue that one cannot necessarily determine in advance what the best kinds of information are going to be. Their point is that only by giving the interviewer freedom is it possible for him to collect unique cues which would otherwise not be obtained but which may have great relevance (validity) to eventual job success.

Both camps would agree that the ultimate test of good selection interviewing must be the degree of job success observed by those applicants selected by the particular method. Therefore it would seem that the argument between the two groups can be decided by empirical research. Unfortunately such definitive research remains undone. One major problem, of course, concerns the difficulty of deciding which questions need to be included in the structured interview.

The authors suggest that any interviewing, by definition, must allow for a reasonable degree of question variety from applicant to applicant—a position which puts us a little closer to the unstructured end of the continuum than to the structured end.

EVALUATION OF THE INTERVIEW

Of course, the ultimate test of the interview—and of any selection device—must lie in an evaluation of its reliability and validity. As we shall see, the evidence at this point remains highly equivocal. There are tremendous difficulties involved in being able to ascertain exactly what validity is actually being measured in an interview study.

INTERVIEW RELIABILITY

Wagner (1949) published a comprehensive review on the employment interview which surveyed 106 different articles. In only 25 of these did he find *any* quantitative information about the value of the interview in selection. In fact, he found

TABLE 5.1 *Summary of Selected Data from Wagner's Survey of the Interview*

Trait or Characteristic	*Reliability*			*Validity*		
Ability to present ideas	0.42					
Alertness	0.36					
Background: family and socio-economic				0.20		
Energy	0.64					
Initiative	0.57					
Intelligence or mental ability	0.96	0.87	0.77	0.58	0.82	0.45
	0.62	0.90		0.94	0.51	0.70
Personality				0.21		
Self-confidence	0.77					
Sociability	0.87	0.72		0.37		
Social adjustment				0.22		
Tact	0.26					
Overall ability	0.71	0.48	0.24	0.27	0.21	0.16
	0.20	0.26	0.43	0.87	0.23	
	0.68	0.61	0.85			
	0.55					

SOURCE: J. Tiffin and E. J. McCormick. *Industrial psychology*, 5th ed. Prentice-Hall, Englewood Cliffs, N.J., © 1965, p. 90.

a total of only 34 reliability coefficients for 174 different sets of ratings. These reliability coefficients ranged from 0.23 to 0.97 for ratings of specific traits and from −0.20 to 0.85 for ratings of overall ability. The only trait to give reliability coefficients above 0.40 was intelligence. While Wagner's summary table is too extensive to show here, Tiffin and McCormick (1965) have prepared a summary of some of Wagner's data as shown in Table 5.1.

More recently, Ulrich and Trumbo (1965) have updated the Wagner survey, examining all research articles on the interview since 1949. They discuss a number of different studies which give reliability data, as follows:

Study	*Results Listed by Ulrich and Trumbo*
Strupp & Williams (1960)	"Significant" interrater agreement on 9 different traits
Sternberg (1950)	Reliabilities ranged from 0.15 to 0.71
Bonneau (1957)	Interrater reliabilities in 0.80's
Anderson (1954)	Interrater reliabilities in 0.80's
Shaw (1952)	Reliabilities ranging from 0.71 to 0.78
Prien (1962)	Reliabilities of 0.55 and 0.62
Raines and Rohrer (1955)	Reliability of 0.15
Plag (1961)	13 of 15 reliabilities significantly greater than 0
Zaccaria, *et al.* (1956)	Reliability estimated at 0.72

In summarizing the data on interview reliability since Wagner's study, Ulrich and Trumbo (1965, p. 108) have this to say:

> It is apparent, first, that few studies have reported reliabilities, and, second, that those reported, with few exceptions, are lower than usually accepted for devices used for individual prediction. Reliability coefficients of criterion ratings were almost never reported, but probably did not exceed those reported for the interview. Therefore, unreliability remains a serious source of attenuation for any validity coefficients which might be found.

EFFECT OF STRUCTURE ON INTERVIEW RELIABILITY Not only is it difficult to acquire good information about the reliability of the interview, but there is almost a complete scarcity of research concerning what variables influence interview reliability. As the Wagner (1949) data indicate, the kind of trait being measured is apparently important. One would also suspect that the degree of structure is an important consideration in determining reliability of the selection interview. As yet, however, no systematic research is available on this vital problem. While Bass (1951) did obtain reliabilities between guided interviews conducted at two different companies and obtained inter-interviewer correlations of 0.56 and 0.74, and while these are quite high, there was no comparison data on less structured interviews performed with the same individuals. Perhaps more directly relevant to the question are the findings of Pashalian and Crissy's 1953 study on the selection of men for the submarine service. Mayfield (1964, pp. 249–250), in another recent and excellent survey of the effectiveness of the selection interview, comments on the Pashalian and Crissy study:

> In an unstructured interview material is not consistently covered. This is probably one of the causes of the unreliability of the interview. . . . This fact was clearly shown by

Pashalian and Crissy (1953) in their work with submariners. In this study it was found that out of 109 interviews, the following information was covered the percentage of times indicated:

Establishment of attendance at high school or college	86%
Marital status	75%
Reasons for leaving school—graduated, joined the Navy, etc.	64%
Kinds of duty held in Navy	55%
Place(s) of duty in the Navy	50%

In general, it was found that the items most consistently covered were of the factual, biographical type. The lowest consistency was found for items of the attitudinal type, such as attitude toward job held.

INTERVIEW VALIDITY

Even should the interview prove to be a highly reliable assessment device (which it does not appear to be), there still remains the problem of its validity. As was pointed out earlier, there are those who maintain that it would be undesirable to have an interview highly reliable, for this would necessitate restricting the information gathering of the interviewer and thus restrict his potential validity. This argument, it might be noted, has some support from classical measurement theory in the form of what is known as the "attenuation paradox." This paradox refers to the problem in which increasing the reliability of a test instrument can result in a decrease in the validity of that same measuring instrument.

One further difficulty in evaluating interview validity has to do with the criterion used—specifically, what were the interviews trying to predict? Both Ulrich and Trumbo (1965) and Mayfield (1964) emphasize that criterion inconsistencies between studies may be one of the major reasons for lack of agreement. For example, Wagner's (1949) earlier survey had reported only 22 validity coefficients which ranged in size from 0.09 to 0.94 (with a median of 0.19). The highest validities (as was the case with reliability) were found when the criterion was a rating of overall ability. Ulrich and Trumbo in their review subdivided all validity studies into three classifications depending upon the criterion used. These were (1) predictions of proficiency ratings, (2) predictions of success in training, and (3) predictions of psychiatric ratings or discharge. A brief summary of their findings for the first two criterion categories, prediction of proficiency ratings and predictions of success in training, is given on the following page.

In only a few of these research studies does one find the interview emerging as a meaningful predictor of either job or training success. While in a number of cases it has demonstrated moderate validity, in most of these it adds little to any multiple prediction over and above what can be obtained with more standardized test instruments—a disturbing fact when one considers the cost of the interview relative to the cost of these other selection methods. At the risk of seeming redundant, there would certainly seem to be a great deal of need for definitive research on the interview. Perhaps no other aspect of personnel psychology exists today about which we know less and which costs industry more money yearly than does the selection interview.

Predictions of Proficiency Ratings

Study	*Results*
Tupes (1950)	Interviews tended to increase validity above that obtained using credentials alone. Median *r*'s were raised from 0.21 and 0.28 to 0.42 and 0.46
Kelly and Fiske (1950, 1951)	Median validities were increased 0.01 by adding interview
Anderson (1954)	Validity for interview alone was found to be 0.51
Raines and Rohrer (1955)	Interview validities of 0.30 and 0.33 for prediction of two different types of proficiency ratings
Yonge (1956)	Interview validities ranged from 0.48 to 0.99
Zaccaria, *et al.* (1956)	Three-man interview boards had validities of about 0.20
Woodworth, Barron, and Mackinnon (1957)	Median correlation between interview predictions and criteria was less than 0.20
Trites (1960)	Interview predictions correlated in 0.20's with criteria
Shaw (1952)	Interview significantly increased prediction
Bonneau (1957)	0.65 validity with very structured interview
Campbell, Prien, and Brailey (1960)	No validity for interview obtained
Huse (1962)	Interview did not increase validity of prediction based on test data
Prien (1962)	Validities of interview were significant for nonsales positions but not for sales

Predictions of Success in Training

Study	*Results*
Pashalian and Crissy (1953)	Interview had high validity for success in submarine school
Holtzman and Sells (1954)	Predictions of interviewees concerning success in flight training based on test data only were no better than chance
Plag (1961)	Interviewer predictions of success had validity but did not add anything over that obtained by other methods
Trankell (1959)	Validities of interview predictions were uniformly higher than statistical predictors based upon test data

OTHER INTERVIEW CONSIDERATIONS

It is one thing to ask how well one can predict job success on the basis of the interview and quite another to ask about the accuracy of the information gathered via the interview method. These are quite independent questions since consistency, not accuracy, is the primary prerequisite for validity, that is, one can *potentially* predict from any collection of data which is *not random*, no matter what its bias.

Several rather comprehensive studies have been done to determine the answer to

the accuracy question. The most comprehensive was probably that of Keating, Paterson, and Stone (1950). They examined the accuracy of information obtained from job applicants with regard to their weekly wages on previous job, duration of past employment, and the job duties in previous work. To check the information obtained in the interview on these matters they obtained data directly from the previous employers. Considering only self reports on jobs held within the past year, they found that (1) reported previous wages correlated with actual previous wages 0.90 for men and 0.93 for women, (2) for duration of employment the correlation was 0.98, and (3) for estimation of previous job duties agreement was 94 percent for men and 96 percent for women.

Somewhat poorer agreement was obtained in a study by Weiss and Dawis (1960). They checked on the accuracy of interview information obtained from 91 physically handicapped individuals by comparing against agency records and employer records. The major results are shown in Table 5.2. As the table indicates, accuracy varied tremendously as a function of the kind of information involved, with sex being always correctly reported (not unsurprising) while reports of prior financial aid were 55 percent inaccurate.

There is certainly no reason to believe that the interview as a data collection

TABLE 5.2 *Accuracy of Information Obtained from 91 Individuals*

Item	*N*	*Percent with Invalid Information*
Age	91	10
Sex	91	0
Marital status	91	15
Education	91	21
Veteran status	91	4
Nature of disability	91	10
Age at disablement	48	33
Received assistance	91	55
Job title	49	24
Job duties	49	10
Hours	49	16
Pay	46	22
Length of employment	48	29

SOURCE: D. J. Weiss and R. V. Dawis. An objective validation of factual interview data. *Journal of Applied Psychology*, 1960, *44*, 381–385.

device should be any more immune to the kinds of "response bias" discussed in Chapter 4 than are paper-and-pencil inventories. Indeed, face-to-face situations are sometimes more prone to cause pressure on the respondent to stretch the truth just a bit when replying to certain kinds of questions. No one likes to make himself look bad, but it's not quite so difficult to confess embarrassing information to a sheet of impersonal paper as it is to do so directly to an interviewer. To date, no one has directly compared the accuracy of information obtained in interviews versus the accuracy of the same kinds of information obtained via questionnaires, but it is

not unreasonable to hypothesize that any discrepancies between the two methods would most likely be related to the nature of the setting in which the question is asked.

STUDIES ON THE INTERVIEW PROCESS

One approach to a better understanding of the complex nature of the selection interview is through an analysis of the content of such interviews. Anderson (1960) analyzed the contents of over 100 interviews and found that applicants tended to speak more than interviewers; that the more the interviewer spoke the less the applicant spoke; and, perhaps most interesting of all, that those applicants who were accepted talked more than those who were rejected. Score one for verbal behavior!

In a rather interesting study, Johnson (1958) related specific trait ratings made by the interviewer to overall final assessment ratings. In other words, each interviewer rated each interviewee on each of eight different traits and then gave each interviewee a global, overall rating. While no actual trait validities were reported to indicate how well each individual trait was related to the overall evaluation, Johnson did find that over a third of those given an overall rating of excellent had only one or zero excellent trait ratings. Thus high trait scores were not a prerequisite to a high overall rating. Of course, this could simply mean that the traits used by Johnson were not ones felt to be particularly important by the interviewers.

What kinds of information or cues do interviewers tend to pay greatest attention to as they assess the applicant? Bolster and Springbett (1961) have attempted to provide at least a partial answer to that question. They simplified it a little by saying first that most of the "important" cues in the interview must be either positive or negative in their influence upon the interviewer. They then went one step further

TABLE 5.3 *Composition of Twelve Protocols of Interview Information*[a]

	Protocols											
Items	*A*	*B*	*C*	*D*	*N*	*M*	*X*	*Y*	*P*	*Q*	*R*	*S*
1 to 5	+	+	−	−	−−	+	−	++	+	++	−	−−
6 to 10	++	++	−−	−−	+	−−	++	−	−	−−	+	++

[a] + and − indicate favorable and unfavorable information. Single and double signs indicate low and high scale values.

SOURCE: B. I. Bolster and B. M. Springbett. The reaction of interviewers to favorable and unfavorable information. *Journal of Applied Psychology*, 1961, *45*, 97–103.

and raised the interesting question as to which kinds of cues had the greatest impact: Is an interviewer more influenced by some highly favorable piece of information than he would be by an equally unfavorable one? They designed a rather clever study in which "interviewers" were given prepared protocols, each of which had ten items of information about a theoretical person. Each interviewer was asked to rate 12 such protocols. However the 12 protocols differed in terms of the number of favorable and unfavorable statements used to describe that person (see Table 5.3).

The results of the study indicated clear-cut evidence that shifts in the direction of rejection are more easily accomplished than shifts in the direction of acceptance.

In other words, negative information appears to carry more weight than does information of a favorable nature when the two pieces of information are equally distant from the point of neutrality (one in the + direction, the other in the − direction). They also obtained evidence that a highly unfavorable item would produce a disproportional attitude shift on the part of the interviewer—much greater than one might expect to obtain, for example, with two moderately unfavorable items.

SUMMARY

From the point of view of practice, the interview is a widely but variously used technique. Recognizing this, the authors have refrained from a discussion of "how to interview." Rather, the interview has been considered as a selection device and the lack of clear-cut research findings has been emphasized. The interview needs to be subjected to more empirical research, and until this is accomplished there will be many unanswered questions concerning not only its validity but also the "best" interview technique to use.

APPLICATION BLANKS

Rarely does an individual apply for a job without being asked to fill out some type of application form. The amount of required information in such applications may vary from simply name, age, address, and phone number to a ten-page document covering all aspects of previous education, work history, and private life.

Application blanks generally have two functions. First, they provide pertinent information about the employee which the company will need if the individual is hired. Examples are age, sex, number of dependents, social security number, etc. Second, application blanks are designed to gather information about job applicants which the personnel officer feels are pertinent to the hiring process. In a sense, an application blank is a highly structured interview in which the questions have been standardized and determined in advance.

Generally the information requested on an application blank is concerned with what might be called *personal history*. It includes items dealing with the previous work and life history of the applicant. The major problem with application blanks is that more often than not they are designed rather haphazardly, perhaps by having one or two individuals make up the kinds of questions they feel "ought to be asked." In many instances, an application blank from some other company is used as a guide. Such a procedure can result in lengthy and inappropriate application blanks simply because one is afraid to risk asking too few, rather than too many questions.

The problem of what items need to be included on an application blank is certainly important. However, it can only be answered by considering the related question of what use will be made of the information obtained. For example, it is of little value to have an application blank ask for significant yet subtle pieces of critical information if all that is going to occur is to have the personnel office give the application blank a quick scan and use it as background information preliminary to the employment interview.

Used properly as an aid to selection, the application blank can be and sometimes is one of the better selection devices at the disposal of the personnel technician. However, just as with any other selection device, only valid informational items should be included on the form. This requires that all the items on an application blank should be checked to see if they are in any way indicative of future successful job performance. The procedure for accomplishing this type of analysis is very similar to that used in determining what items should be included in a psychological test, *viz., item analysis.* The most common method is simply to compute the correlation between an item on the application blank (e.g., number of previous jobs in the past five years) and some later measure of job success (e.g., job tenure). If a significant relationship is found to exist (and in this case we might expect to find a significant negative relationship), that item can then be used to select future job applicants.

It is thus possible to construct "keys" for each of a number of different jobs within a plant, where a key consists of a list of those items on the application blank which predict success on that particular job. Of course one must keep in mind that cross-validation of these keys is a prerequisite to their actual use in a selection situation. For an excellent discussion of how to develop scoring keys for an application blank, refer to England's *Development and Use of Weighted Application Blanks* (1961).

One advantage of the application blank over the interview and tests of personality and interest is that response bias does not usually play as great a role. Even if a person did tend to bias his answers on the application on matters to which he felt some concern about giving an honest answer, in many cases such data is subject to rather easy verification from other sources. (Of course this does not mean that such verification will always be obtained.) Intuitively, it would be expected that many questions on the application blank should be related to job success. Certainly a person's previous work history should provide some sort of clue to the probability of successful performance on the job under consideration. Similarly, past personal life history and family data ought to give some indication of the emotional and personality characteristics of the applicant, which may have some bearing on eventual job adjustment.

EXAMPLES OF APPLICATION BLANK EFFECTIVENESS

FLEISHMAN AND BERNIGER STUDY An excellent example of the use of the weighted application blank is the study by Fleishman and Berniger (1960). They examined the application-blank responses of 120 women office employees who were subsequently divided into two criterion groups depending upon how long they had remained with the company. The "long-tenure" group consisted of women who had worked for two to four years and were still on the job. The "short-tenure" women were those who had terminated employment within two years of hiring. Weights were then given to those items on the application blank which appeared to differentiate between the two groups of women, as shown in Table 5.4.

A key was then used to compute a score for each applicant using the weights shown in Table 5.4. All items which did not predict were given a weight of zero. Thus an applicant under 20 years of age got − 3 points toward her total score, while an applicant living within the city limits got + 2 points toward her score, etc.

The key was then cross-validated using a new sample of 85 job applicants. The

TABLE 5.4 *Comparison of Item Responses by Long- and Short-Tenure Office Employees*

Items on Application Blank	Percentage of Short-Tenure Group	Percentage of Long-Tenure Group	Weight Assigned to Response
Local address			
Within city	39	62	+2
Outlying suburbs	50	36	−2
Age			
Under 20	35	8	−3
21–25	38	32	−1
26–30	8	2	−1
31–35	7	10	0
35 and over	11	48	+3
Previous salary			
Under $2000	31	30	0
$2000–3000	41	38	0
$3000–4000	13	12	0
Over $4000	4	4	0
Age of children			
Preschool	12	4	−3
Public school	53	33	−3
High school or older	35	63	+3

SOURCE: Adapted from E. A. Fleishman and J. Berniger. One way to reduce office turnover. *Personnel*, American Management Association, 1960, *37*, 63–69.

correlation between scores on the keyed items and tenure of employment (long versus short) was 0.57—indicating the key was predictive of tenure of employment.

KIRCHNER AND DUNNETTE STUDY Another example of the value of the application blank with clerical workers is in the study reported by Kirchner and Dunnette (1957). Again, job tenure was the criterion, with short tenure being under 10 months and long tenure being over 18 months. They examined 40 items on the blank and found 15 to differentiate tenure groups. Using a key based upon these items, they obtained the fold-back validity results on their original validation group as shown in Table 5.5.

TABLE 5.5 *Quartiles and Means of Weighted Scores for Long-Term and Short-Term Office Employees*

Tenure Group	Original Group					Cross-Validation Group			
	Median	Q_2	Mean	SD	N	Median	Mean	SD	N
Short	10	13	10.79	4.51	33	12	12.32	3.95	40
Long	18	21	17.59	4.78	105	18	17.69	4.89	45

SOURCE: W. K. Kirchner and M. D. Dunnette. Applying the weighted application blank technique to a variety of office jobs. *Journal of Applied Psychology*, 1957, *41*, 206–208.

The long-term group received a mean score of over 17 using the key, while the mean score for the short-tenure people was 10.79. The key was then applied to a new sample of employees for the purpose of cross-validation. As Table 5.5 shows, the key retained its ability to discriminate the two criterion groups.

NAYLOR AND VINCENT STUDY Several studies have attempted to use the application blank as a predictor of job absenteeism. Naylor and Vincent (1959), for example, carried out a rather extensive study on 220 women clerical workers, ranging in age from 18 to 58, who were employed in a large midwestern manufacturing company. The variables which they obtained from the personal data sheet were marital status, age, and number of dependents. The criterion of job success was absenteeism from work over a six-month period. The three predictors and the criterion were dichotomized as follows:

Marital Status	married—single[2]
Age	32 and above—31 and below
Number of Dependents	1 or more—none
Absenteeism	4 days or more—less than 4 days

Half the subjects were put into a primary group, and the other half were set aside for purposes of cross-validation on a hold-out group. The relationship between each of the predictors and the criterion was then determined by means of x^2. Those predictors significant at the 0.01 level were then used with the second group to see if the relationship would hold up in cross-validation.

Table 5.6a shows the frequency counts and obtained x^2's for each of the three predictors against the criterion. Only one of the three showed a significant relationship with absenteeism—number of dependents. This variable was then cross-validated on the other group. The resulting relationship was again significant, this time at the 0.05 level, as shown in Table 5.6b.

TABLE 5.6a *Frequency Counts and x^2 for Each Variable (Primary Group)*

	Married	*Single*	*Old*	*Young*	*Dependents*	*No Dependents*
High absentee	36	16	24	28	25	27
Low absentee	31	27	29	29	13	45
	$x^2 = 2.87$		$x^2 = 0.16$		$x^2 = 7.99^a$	

[a] Significant at 0.01 level.

TABLE 5.6b *Frequency Count and x^2 for Dependents Variable (Hold-out Group)*

	Dependents	*No Dependents*
High absentee	34	29
Low absentee	14	33
	$x^2 = 6.40^a$	

[a] Significant at 0.05 level.

[2] Divorced and widowed women were considered single.

TABLE 5.6c *Variable Intercorrelations*

	Age	*Number of Dependents*	*Validity*
Marital status	0.253	0.539	0.212
Age		0.495	0.068
Number of dependents			0.424

SOURCE: J. C. Naylor and N. L. Vincent. Predicting female absenteeism. *Personnel Psychology*, 1959, *12*, 81–84. (Source for Tables 5.6, a, b, and c.)

Although number of dependents was the only variable to emerge as a predictor, Table 5.6c shows that marital status was correlated with absenteeism to some extent ($r = 0.212$). The magnitude of this correlation suggested that the addition of the marital status variable along with the age factor might increase the overall accuracy of prediction beyond that which could be obtained by using the number of dependents variable alone.

To test this, a multiple correlation for all three predictors against the criterion was computed. The multiple R exceeded by only 0.019 the r obtained using number of dependents alone. This was nonsignificant.

VALIDITY OF APPLICATION BLANK OVER TIME

As with other types of predictors, one must always be concerned with the stability of a predictive relationship over time. How likely is it that scoring keys, developed at one point in time, are going to successfully predict at some later time period?

Buel (1964) found that 13 out of an original 16 valid items retained their predictive efficiency even though a substantial change had taken place in the population of job applicants (the company had actually relocated its clerical offices).

Studies which specifically examined the stability of an application blank scoring key were those of Wernimont (1962) and of Dunnette, Kirchner, Erickson, and Banas (1960). Both of these studies examined the change in predictive validity of the valid items found in the earlier Kirchner and Dunnette study (1957). The Dunnette, *et al.* study reported the follow-up data on the key developed in 1954 as shown in Table 5.7. The decrement in validity over time is quite apparent.

TABLE 5.7 *Numbers of Long-Term and Short-Term 3M Employees Scoring Above and Below a Score of 15 on the Weighted Application Blank*

	Girls Hired During 1954		*Girls Hired During 1955*		*Girls Hired During 1956*	
	Short Term	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*
Score of 15 or above	5	76	12	32	27	45
Score of 14 or below	28	29	28	13	18	12
Tetrachoric correlation (r_t)	0.74		0.61		0.38	

SOURCE: Adapted from M. D. Dunnette, *et al.* Predicting turnover among female office workers. *Personnel Administration*, 1960, *23*, 45–50.

Wernimont, in a final study, found that the correlation between scores on the 1954 key and job tenure had dropped to 0.07 by 1959. He then developed a new key which, when cross-validated on a hold-out sample, gave a correlation with job tenure of 0.39. The only items which retained their prediction efficiency from 1954 to 1959 were (1) high proficiency at shorthand, (2) did not leave last job because of pregnancy, marriage, sickness, or home problems, and (3) will begin work on the new job one week or more from now. The author strongly recommended that scoring keys be revalidated at least every three to five years—a caution which applies to all other selection devices as well.

A PROBLEM IN APPLICATION BLANK VALIDATION

One difficulty in nearly all of the application blank validation studies concerns the problem of preselection. As was pointed out earlier in the chapter, application blank items are often an integral part of the interview process. Very frequently interviewers will use the application blank as a guide in conducting their interviews. Also, it is probably literally impossible to conduct a selection interview without having the applicant reveal many of the same items of information which are required on the application blank. This means that the decisions of the interviewer to accept or to reject an applicant are apt to be determined, at least to some degree, by the same items that are included on the application blank. In an examination of previous validations of application blanks, Myers and Errett (1959) were unable to find any study which took this preselection into account.

Preselection of this nature generally results in a restriction of range among those hired on the attribute in question and (if the item is truly valid) also on the criterion. Effects of this were discussed in Chapter 3, in which it was pointed out that restriction of range led to a reduced validity coefficient which is an underestimate of the true validity of that predictor. This would indicate that the validities obtained so far with the application blank are likely to be lower bound estimates of the true validity of these devices.

LIFE HISTORY DATA AND JOB SUCCESS

Aside from the kinds of information typically used in application blanks (which of course are often limited in length for obvious practical reasons), numerous investigators have recently become quite active in what might be termed *life history antecedents of worker behavior*. The interest of these researchers is much broader and certainly more theoretical than is usually found in the typical application blank validation study. The study of personal history correlates of job success is a research area based upon the logical premise that the behavior of individuals in any setting (including the work situation) should be related to the life history experiences of those individuals. The research has tended to examine a tremendous variety of different types of life history items to determine how they are interrelated and how they are related to job success, job satisfaction, etc.

STANDARD OIL STUDIES The single group most active in the area of personal history research is at Standard Oil of Indiana, where a series of studies during the past several years has attempted to examine various aspects of biographical data as predictors.

In the first study in the series (Smith, Albright, Glennon, and Owens, 1961), a personal history questionnaire was administered to a group of petroleum research scientists. The questions were both validated and cross-validated against three different criteria: (1) overall performance ratings, (2) creativity ratings, and (3) number of patents. The concurrent cross-validation indices for these three criteria were 0.61, 0.52, and 0.52 respectively—a rather impressive set of validities.

A second study (Albright and Glennon, 1961) based upon the same data found 43 different personal history items (out of an original 484) which differentiated significantly the petroleum scientists who "desired to advance into laboratory supervision" from those who desired only increased salary but who preferred to remain scientists. Thus, the degree to which a scientist has supervisory aspirations is apparently related to biographical history of that scientist.

The third study in the series (Morrison, Owens, Glennon, and Albright, 1962) was designed to determine the underlying "dimensions" of life history experience by using the technique of factor analysis to group together similar life history item "types." They factor-analyzed 75 different items which had been found to discriminate on at least one of the three criterion measures used in their first study (performance ratings, creativity ratings, and patents).

Five different factors or groupings of items were obtained:

1. *Favorable self-perception.* The items which grouped together to define this factor are:
   a. In the top 5 percent of performance in their occupation
   b. Could be highly successful supervisors if given the opportunity
   c. Work at faster pace than most people
   d. Desire to work entirely autonomously selecting both a method and a goal
   e. Like a lot of responsibility
   f. In an unpleasant situation, try to react and formulate a decision immediately
   g. Are friendly and easy-going and have many friends

All of these items are highly favorable or flattering toward the respondent. Thus people scoring high on this factor (that is, checking agreement with these statements) are those having a highly favorable self-impression. The remaining factors were:

2. *Inquisitive professional orientation*
   a. Completed their Ph.D.s
   b. Belong to one or more professional organizations
   c. Do not limit themselves to undergraduate work as an educational goal for their sons
   d. Have some close friends and a number of acquaintances
   e. Devote much time to reading of many kinds, preferring current and political topics among nonprofessional areas
   f. Desire to work entirely autonomously
   g. Have high salary aspirations
3. *Utilitarian drive*
   a. Desire extrinsic rewards, that is, from business and society
   b. Prefer urban dwelling
   c. Started dating prior to age 20
   d. Feel free to express their views and perceive themselves as influencing others in group and individual situations, do not want to work with just one other person
   e. Do not desire to work entirely autonomously; want to choose their own method, but not necessarily the goal toward which they are to strive
   f. Feel dissatisfied with themselves at times

4. *Tolerance for ambiguity*
   a. Desire to have many work projects going simultaneously
   b. Are not single
   c. Have solicited funds for charity and made speeches
   d. Have high salary aspirations
   e. Have friends with similar and dissimilar political views
5. *General adjustment*
   a. Feel that school material was adequately presented
   b. Came from happy homes where they were well treated
   c. Express their opinions readily and feel that they are effective in doing so
   d. Have high salary aspirations
   e. Can tolerate inefficiency in a job better than less controllable problems
   f. Prefer verbal to laboratory course work

The next question concerned the relationship of the three criteria measures to these five different life history dimensions. Figure 5.3 indicates the degree to which each obtained factor was related to each of the three criteria. What is interesting is that the two criteria involving ratings have very similar patterns. Thus petroleum

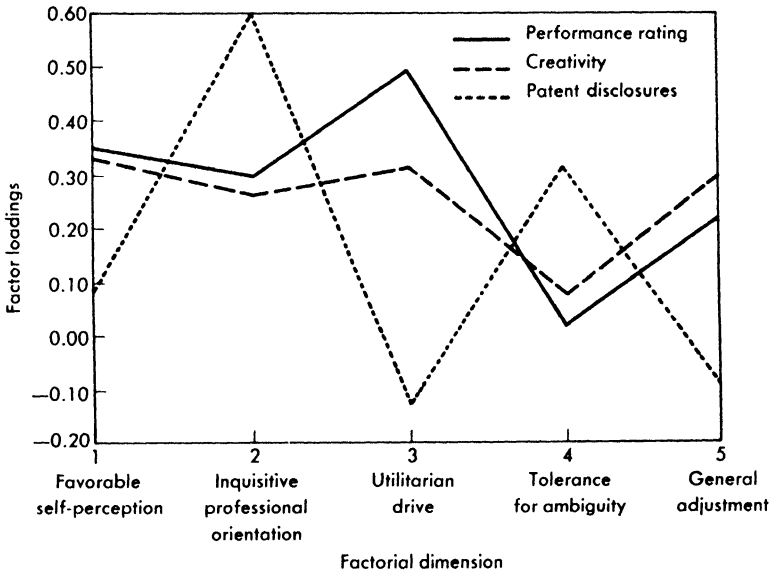


Figure 5.3. Profiles of criterion factor loadings on each of five dimensions identified. (Adapted from R. F. Morrison, *et al* Factored life history antecedents of industrial research performance. *Journal of Applied Psychology*, 1962, 46, 281–284.)

scientists who receive high ratings on performance and on creativity tend also to get high scores on the factors of self-perception, utilitarian drive, and general adjustment. Note that the pattern for patent disclosures is exactly opposite to the pattern for the ratings. Those scientists with many patent disclosures score low on self-perception, utilitarian drive, and self-adjustment, but very high on inquisitive professional orientation and tolerance for ambiguity.

As can be seen, studies such as this go far beyond the practical questions of

selection and provide a good deal of insight and understanding as to why one particular group of individuals is more likely to be successful (as defined by a particular criterion) than some other group. More of this type of research is needed.

A recent study in the series was reported by Chaney and Owens (1964). This study differed from the earlier ones in that it attempted to predict interests of college freshmen in the area of sales, research, and engineering using life history items. Responses to a 170-item, multiple-choice life history questionnaire were analyzed.

A sample of 388 university freshmen in engineering was used for the sales and research portions of the analysis, and 700 freshmen were used in appraising general engineering interest. The significant items were used to develop scoring keys for each interest criterion. When these keys were cross-validated on completely independent samples, correlations of 0.57, 0.42, and 0.51 were obtained for sales, research, and general engineering respectively.

OTHER PERSONAL HISTORY STUDIES Numerous other studies have been reported which have examined the efficiency of the personal history or biographical approach in predicting success. These have covered a variety of occupations and criteria. For example, Himelstein and Blaskovics (1960) predicted combat effectiveness ratings with biographical data; Harrell (1960) assessed the validity of biographical data in predicting success of food sales managers, although the study was subsequently criticized for poor design by Taylor and Nevis (1961). Lockwood and Parsons (1960) have related personal history data to the performance of production supervisors, and Scollay (1957) was also successful in using biographical information as a predictor of success.

PAST WORK HISTORY

There is an old adage in industrial psychology which says that "the best predictor of future work is past work performance." In spite of the adage, however, prior work performance is in most cases not utilized as systematically as it might be in the selection process. One notable exception to this has been the methods used by the United States Navy in selecting air cadets. The Navy has employed what amounts to a multiple hurdle model, using success at each stage in flight training as a predictor of degree of success in future stages. They have enjoyed a great deal of success using this technique.

An excellent example of how early performance can be effectively used to predict future job success is reported by N. L. Vincent.[3] He studied a group of 542 insurance agents, and at the end of their first three months on the job determined how successful each agent had been using the following ratio:

$$\text{Success Ratio (S.R.)} = \frac{\text{Total First Commissions Earned}}{\text{Total Finance Payments Made to Agent}}$$

The numerator in this ratio represents the amount of money the agent brought into the company during his first three months, while the denominator represents the amount of money the company has invested in the agent in the form of financing him during the same time period. Thus the larger the ratio, the more "successful" the agent in paying his own way.

[3] Personal communication from N. L. Vincent, State Farm Insurance.

The sample of agents was divided into four groups on the basis of the success ratio. The expectancy chart in Figure 5.4 shows the survival percentages through the first contract year (an additional nine months) for each of these four groups.

Clearly there exists a very dramatic relationship between a person's success ratio at the end of three months with the company and the likelihood of that person remaining with the company for the entire first year.

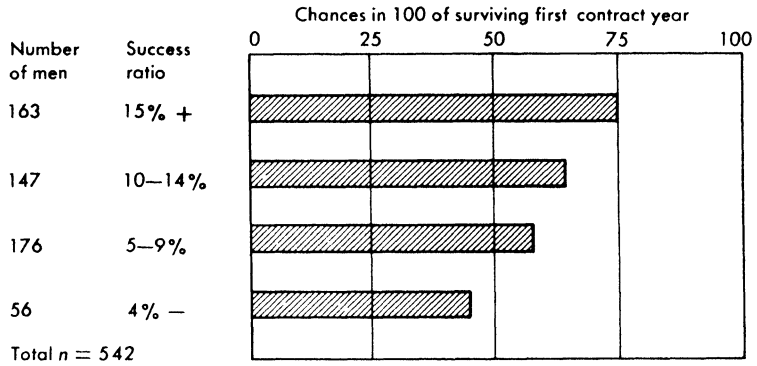


Figure 5.4. Early performance and survival for financed agents. (Data from N. L. Vincent, State Farm Insurance; personal communication.)

The use of initial job success as a means of selecting employees does have certain limitations. While it generally is conceded to be an effective predictor of success, it is also a very expensive predictor. One must pay the salary of the worker during his trial period, and in addition there are often expensive training costs which exist as corollary expenses during the period. Also, at times union agreements set very stringent restrictions upon the length of time which can be used as a trial period; once this time interval is past, the worker has the protection of the union-company agreement and no longer can be easily dismissed. Therefore it usually pays to examine the costs and efficiency of this type of selection very closely in relation to the costs and efficiency of other selection devices before becoming very deeply involved.

REFERENCES AND RECOMMENDATIONS

Many times job applicants are required to produce some sort of documentary evidence that they performed satisfactorily in one or more of their past positions. Such evidence is usually provided in the form of a letter of recommendation or letter of reference from the previous employer to the prospective employer. These letters most often follow no special format and will vary greatly in detail and information from writer to writer. At times such letters or statements are highly personalized, covering the character and morals of the applicant as well as his abilities and skills. On the other hand, some letters are very limited in scope and only cover dates of employment.

The difficulties with using references as guides in the selection of job applicants are many. One must always be concerned with the accuracy of the description

provided in the recommendation. Letters of recommendation may be very inaccurate and misleading, giving a picture of a person that may be quite incorrect. There are perhaps four major reasons for a letter of reference to be inaccurate in content.

1. *Knowledge of applicant.* Many times employers are asked to provide letters of reference to terminating employees of whom they have little or no direct knowledge. Indeed, in some organizations "form" letters of recommendation are provided to all employees who terminate on a voluntary basis. Such letters merely indicate that the employee was not fired, and certainly are totally useless to other employers as far as providing any real information about the applicant's talents or potential for success. One of the most difficult tasks faced by one of the authors arises when college seniors request a letter of recommendation for graduate study. In some instances the total student-professor contact has been less than one hour. To write any recommendation for a person based upon such limited knowledge is no more than a farce.

2. *Ability to assess applicant.* Even if the employer has had sufficient exposure or contact with the worker to become familiar with his talents, there is then the question of the degree to which the employer is capable of getting an accurate impression of the worker. That is, how capable is the employer of accurately assessing the talents of his workers?

3. *Ability to describe applicant.* Assuming that the employer has both the ability and the opportunity to *evaluate* his workers, he must still face the problem of adequately describing or communicating his impressions via the letter of recommendation to the next potential employer. Unfortunately, many people do not have the writing skill to transmit their impressions accurately to another person via letter. To the extent that the employer lacks the ability to communicate in this fashion, the accuracy of the recommendation is going to suffer.

4. *Willingness to be accurate.* The final obstacle to the accuracy of the letter of recommendation is probably the most serious of all. This obstacle is the reluctance on the part of the employer to portray the unfavorable aspects of those he is asked to recommend. Sometimes such reluctance arises out of a general tendency to be "kind" or "charitable" to the departing employee; in other, less frequent, occasions, the employer will deliberately falsify the recommendation to make the employee look very good so that another firm will hire him and the present employer will no longer be bothered with the person. In other words, the employer makes a bad employee sound good just to get rid of him. No one knows how often this occurs, but the authors suspect that such is the case in all too many instances.

VALIDITY OF RECOMMENDATION

The most definitive research on the bias and validity of the personal recommendation has been done by Mosel and Goheen (1958, 1959). They report a series of studies on civil service personnel evaluating the *Employment Recommendation Questionnaire* (ERQ). The ERQ is a civil service recommendation blank which covers (1) employment history, (2) evaluation of the applicant's personality and character, (3) evaluations of the applicant's job ability, and (4) attitude toward rehiring.

The initial study in the series dealing with the ERQ was a study of its validity. Using ERQs from 1193 civil service employees in 12 skilled occupations, Mosel and Goheen (1958) related ERQ scores to supervisory performance ratings. The

results are shown in Table 5.8. In only four trades were the correlations substantial enough to be statistically significant, and even in those cases the magnitudes were, in the absolute sense, quite small.

The second study in the series (Goheen and Mosel, 1959), attempted to compare the data collected by the ERQ with data obtained using extensive field interviews. They studied 109 applicants in three different civil service positions. Each applicant was given an ERQ score and was also given a rating based upon the field interview

TABLE 5.8 *Correlation of Mean ERQ Scores with Supervisors' Performance Ratings—Data from 1117 Employees in 12 Skilled Trades*

Trade	*N*	*r*
Carpenter	51	0.01
Equipment repairman	40	0.23
Machinist	100	0.24*
Machine operator	108	−0.10
Ordnanceman—torpedo	125	−0.01
Radio mechanic	107	0.29**
Aviation metalsmith	94	0.24*
Highlift fork operator	108	0.21*
Auto mechanic	98	0.09
Painter	70	0.07
Ordnanceman	100	0.10
Printer	116	0.11

* Significant at 0.05 level.

** Significant at 0.01 level.

SOURCE: J. N. Mosel and H. W. Goheen. The validity of the employment recommendation questionnaire in personnel selection. *Personnel Psychology*, 1958, 2, 481–490.

protocols. The correlations for three positions between these two measures were, for each position

Economist ($N = 41$)	0.22
Budget examiner ($N = 21$)	0.54
Training officer ($N = 47$)	0.45

While the latter two correlations are statistically significant at the 0.05 level, agreement was certainly far from perfect. A further analysis of the discrepancies between the two assessment methods indicated the ERQ tended not to be sensitive to disqualifying and negative aspects of the applicants—in other words, the fourth source of error listed earlier.

The final study in the series was perhaps the most interesting of all. Mosel and Goheen (1959) examined a total of 3000 ERQs from applicants in six different civil service professions and one civil service trade. Their primary objective was to determine if the leniency and validity of recommendations varied systematically with the

kind of person doing the recommending. They found that there was a general tendency for personal acquaintances to be more lenient than others. The order from most to least lenient was:

1. Personal acquaintances
2. Previous subordinates
3. Coworkers
4. Previous employers

The moral of this would seem to be that if you must have someone recommend you, select an acquaintance or a subordinate rather than a coworker or your supervisor—you will probably get a more favorable letter. If, however, you are an employer or in personnel, then establish who wrote the letter and his relationship to the applicant and estimate the probable bias accordingly.

Table 5.9 reports the data concerning the validity of the ERQs as a function of the kind of person doing the recommending.

TABLE 5.9 *Validity of ERQ as a Function of Kind of Person Doing the Recommending*

Kind of Person	*N*	*Validity*
Personnel officer	102	0.02
Supervisors	188	0.19*
Coworkers	311	0.09
Acquaintances	182	0.20*
Relatives	12	− 0.16

* Significant at 0.05 level.

SOURCE: Adapted from J. N. Mosel and H. W. Goheen. The employment recommendation questionnaire: III—Validity of different types of references. *Personnel Psychology*, 1959, *12*, 469–477.

The only people who seem to be able to produce valid recommendations are supervisors and acquaintances! Note also that personnel officers' recommendations have no validity at all, and the recommendations of relatives are *negatively* related to success.

OTHER MISCELLANEOUS PREDICTORS

While the interview, the application blank, and letters of recommendation are the major selection aids used in conjunction with tests, several other kinds of predictors have been tried in certain instances. For example, Stern and Gordon (1961) were able to predict success in a military training situation using a measure of a recruit's ability to follow simple instructions as the predictor. Peer ratings have also been used quite successfully in certain situations. A good example of their use is given by Weitz (1958) in which he was successful in using peer ratings to select supervisors.

The search for valid predictors of job success will, of course, never cease. Anything that would seem to have potential predictability (and many things which do not)

are equally as important as they are for prediction. Others (for example, Tiffin and McCormick, 1965) have defined the criterion in terms of simply *a dependent variable.* Again, such a definition implies that a functional relationship is involved—an assumption that may not necessarily be true.

Actually, any system for evaluation has the potential for becoming a criterion measure. The only major requirement is that the standard or system selected for evaluation be sensitive enough to discriminate among individuals. However, before proceeding into a discussion of the requirements of criterion measures, we first need to examine certain logical criterion constructs.

CLASSICAL VIEW OF THE CRITERION

Any discussion of the criterion should consider certain terms and concepts which are associated with the more classical view.

DISTAL VERSUS PROXIMAL CRITERIA

The criterion is a measure of the degree of accomplishment in one form or another. However, accomplishment, no matter how it is defined, is not independent of time. Thus the fact that a person has achieved a degree of accomplishment by some standard today does not automatically mean that the same person will have that degree of accomplishment by that same standard at some later date. Further, standards may and often do change over time. What might be an appropriate measure of accomplishment at one moment in time may be an entirely inappropriate measure at some later time.

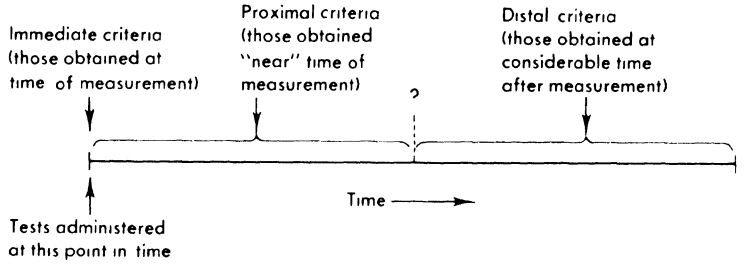


Figure 6.1. Diagram illustrating immediate, proximal, and distal criteria

Criteria which are obtained close to the time at which the prediction instruments are used are referred to as *proximal criteria,* while those which are obtained at some later date are referred to as *distal criteria* (that is, they are more distant). Those criteria which are obtained at the same time the prediction instruments are administered are called *immediate criteria.* Figure 6.1 graphically illustrates these relationships.

Judgment determines at what point in time a measure of accomplishment is no longer proximal and is instead distal. However, a good rule of thumb might be that any criteria obtained a year or more after obtaining the predictor measures can be regarded as distal. In some cases a distal criterion may include proximal criterion data,

Holtzman, W. H., and S. B. Sells. Prediction of flying success by clinical analysis of test protocols. *Journals of Abnormal and Social Psychology*, 1954, 49, 485–490.

Huse, E. Assessments of higher-level personnel: IV. The validity of assessment techniques based on systematically varied information. *Personnel Psychology*, 1962, 15, 195–205.

Johnson, K. K. Company interviewers rate job applicants. *Personnel and Guidance Journal*, 1958, 36, 422–424

Keating, E., D. G Paterson, and C. H. Stone. Validity of work histories obtained by the interview. *Journal of Applied Psychology*, 1950, 34, 6–11.

Kelly, E. L , and D W Fiske. *The prediction of performance in clinical psychology* The University of Michigan Press, Ann Arbor, 1951.

Kelly, E. L., and D. W. Fiske. The prediction of success in the V.A. training program in clinical psychology. *American Psychologist*, 1950, 5, 395–406.

Kirchner, W. K., and M. D. Dunnette. Applying the weighted application blank techniques to a variety of office jobs. *Journal of Applied Psychology*, 1957, 41, 206–208.

Lockwood, H. C , and S. O. Parsons. Relationship of personal history information to the performance of production supervisors. *Engineering and Industrial Psychology*, 1960, 2, 20–26.

McMurray, Robert N. Validating the patterned interview. *Personnel*, 1947, 23, 2–11.

Mayfield, E. C. The selection interview—a re-evaluation of published research. *Personnel Psychology*, 1964, 17, 239–260.

Meehl, P. E. *Clinical versus statistical prediction.* University of Minnesota Press, Minneapolis, 1954.

Morrison, R. F., W. A. Owens, J. R. Glennon, and L. E. Albright. Factored life history antecedents of industrial research performance. *Journal of Applied Psychology*, 1962, 46, 281–284.

Mosel, James N., and Howard W. Goheen. The validity of the employment recommendation questionnaire in personnel selection. *Personnel Psychology*, 1958, 2, 481–490.

Mosel, J. N., and H. W. Goheen The employment recommendation questionnaire: III. Validity of different types of references. *Personnel Psychology*, 1959, 12, 469–477.

Myers, J. H., and W. Errett. The problem of preselection in weighted application blank studies. *Journal of Applied Psychology*, 1959, 43, 94–95.

Naylor, J. C., and N. L. Vincent Predicting female absenteeism. *Personnel Psychology*, 1959, 12, 81–84.

Pashalian, S., and Wm. J. E. Crissy. The interview. IV. The reliability and validity of the assessment interview as a screening and selection technique in the submarine service. *United States Navy Submarine Medical Research Laboratory Report*, 1953, 12 (1, No. 216).

Plag, J. A. Some considerations of the value of the psychiatric screening interview. *Journal of Clinical Psychology*, 1961, 17, 3–8.

Prien, E. P. Assessment of higher-level personnel: V. An analysis of interviewers predictions of job performance. *Personnel Psychology*, 1962, 15, 319–334.

Raines, G. N., and J. H. Rohrer. The operational matrix of psychiatric practice: I. Consistency and variability in interview impressions of different psychiatrists. *American Journal of Psychiatry*, 1955, 111, 721–733.

Scollay, R. W. Personal history data as a predictor of success. *Personnel Psychology*, 1957, 10, 23–26.

Shaw, J. The function of the interview in determining fitness for teacher-training. *Journal of Educational Research*, 1952, 45, 667–681.

Smith, W. J., L. E. Albright, J. R. Glennon, and W. A. Owens. The prediction of research competence and creativity from personal history. *Journal of Applied Psychology*, 1961, 45, 59–62.

Spriegel, W. R., and V. A. James. Trends in recruitment and selection practices. *Personnel*, 1958, 35(3), 42–48.

Stern, F., and L. V. Gordon. Ability to follow instructions as a predictor of success in recruit training. *Journal of Applied Psychology*, 1961, 45, 22–24.

Sternberg, J J. An analytical study of a selection interview procedure. Unpublished master's thesis, Syracuse University, 1950.

Strupp, H. H., and J V Williams. Some determinants of clinical evaluations of different psychiatrists. *American Medical Association Archives of General Psychiatry*, 1960, 2, 434–440.

Symonds, P. M. Research on the interviewing process. *Journal of Educational Psychology*, 1939, 30, 346–353.

Taylor, E. K., and E. C. Nevis. *Personnel Selection*. Annual Review of Psychology, 1961, 12, 389–412.

Tiffin, J., and E. J. McCormick. *Industrial psychology*, 5th ed. Prentice-Hall, Englewood Cliffs, N.J., 1965.

Trankell, A. The psychologist as an instrument of prediction. *Journal of Applied Psychology*, 1959, 43, 170–175.

Trites, D. K. Adaptability measures as predictors of performance ratings *Journal of Applied Psychology*, 1960, 44, 349–353.

Tupes, E. C. An evaluation of personality-trait ratings obtained by using unstructured assessment interviews *Psychological Monographs*, 1950, 64 (11, Whole No. 287).

Uhrbrock, R. S. Analysis of employment interviews. *Personnel Journal*, 1933, 12, 98–101.

Ulrich, Lynn, and D. Trumbo. The selection interview since 1949 *Psychological Bulletin*, 1965, 63, 100–116.

Verplank, W. S. The control of the content of conversation· Reinforcement of statements of opinion. *Journal of Abnormal and Social Psychology*, 1955, 51, 668–676

Wagner, R. The employment interview· A critical review *Personnel Psychology*, 1949, 2, 17–46.

Weiss, D. J., and R. V. Dawis. An objective validation of factual interview data. *Journal of Applied Psychology*, 1960, 44, 381–385.

Weitz, Joseph Selecting supervisors with peer ratings. *Personnel Psychology*, 1958, 11, 23–35.

Wernimont, P. F Reevaluation of a weighted application blank for office personnel. *Journal of Applied Psychology*, 1962, 40, 417–419.

Woodworth, D. G , F. Barron, and D W. MacKinnon An analysis of life history interviewers' rating for 100 Air Force captains *United States Air Force Personnel and Training Research Center Technical Note*, 1957, No. 57–129.

Yonge, K. A. The value of the interview An orientation and a pilot study. *Journal of Applied Psychology*, 1956, 40, 25–31.

Zaccaria, M. A., J T Dailey, E C Tupes, A. R. Stafford, H. G Lawrence, and K. A. Ailsworth. Development of an interview procedure for USAF officer applicants. *United States Air Force Personnel and Training Research Center Development Report*, 1956, No. TN-56-43.

6 THE CRITERION

The criterion is basic to all measurement in industrial psychology. To overstate its importance would be literally impossible. Without adequate criteria, industrial psychology is ineffective and ceases to be a science. In other words, the magnitude of the contribution of industrial psychology is *completely* determined by the adequacy of the criterion measures evolved.

Since criteria are "man-made" and never "God-given," the industrial psychologist must not only evolve methods and systems of measuring concepts, but must also evolve methods and systems against which he can make comparisons so that he can ultimately prove that his methods, systems, and concepts are real, effective, and in fact better than other methods or systems which do not use the rigors and disciplines of industrial psychology.

Although much work on the criterion problem has been done in measuring man's performance and effectiveness on the job, the problems and needs for having criteria available on such other topics in industrial psychology as attitudes, motivation, job satisfaction, leadership, accidents, advertising effectiveness, and consumer behavior, among others, still exists.

Historically, criteria and tests have had primary attention. However, criteria is both a topic of prime importance in the body of knowledge known as Industrial Psychology and at the same time a topic against which all other topics of Industrial Psychology must be evaluated.

DEFINITION OF THE CRITERION

We have already defined the criterion in Chapter 2 as "a measure of the goodness of a worker." In more general terms, however, the criterion is an *evaluative standard* which can be used to "measure" a person's performance, attitudes, motives, etc. The criterion that is usually used concerns the degree to which a worker can be considered successful on the job, although this need not always be the case. For example, the criterion might be sales figures, readership of an advertisement, or any other measure of accomplishment. Other definitions of the criterion may be found throughout industrial psychology. Guion (1965, p. 90), for example, defines the criterion as simply "that which is to be predicted." Such a definition is quite general and pragmatic for situations of selection and placement. However, there are times when the psychologist simply wishes to evaluate without necessarily predicting (as in the case of the attention value of an advertisement). In such situations good standards for evaluation

as in the case of total sales over the first two years of employment. Obviously, total sales over the first two years will include the first month's sales, the second month's sales, etc. In such situations it would probably still be best to view such measures as distal criteria.

ACTUAL VERSUS ULTIMATE CRITERIA

A somewhat different way of viewing criteria is in terms of actual and ultimate measures of success. By definition, whatever measure of success one actually uses is the *actual criterion*. The *ultimate criterion*, on the other hand, is a theoretical and ideal criterion that usually exists only in the psychologist's mind. It is the "true" criterion of success, while our actual criterion is the measure we have been forced to adopt simply because we can do no better.

Perhaps a simple illustration will serve to clarify the concept of the ultimate criterion. Suppose we are interested in developing a test battery to select air cadets. We have to select some measure (or set of measures) which define, to our immediate satisfaction, how successful each air cadet will be. We might, for example, adopt as our criterion of success the rating by the instructor at the end of training. This rating would then become our actual criterion of success against which we would validate our test battery.

However, from a more ideal point of view we would probably say that instructor's ratings are not the ultimate measure of success of an air cadet. In fact, it is exceedingly difficult to decide and define what the ultimate measure of success is in this particular situation. From a military point of view, the ultimate criterion of success for an air cadet might be defined strictly in terms of how successful the cadet is when flying actual combat missions. In a peacetime world such a criterion can never be used, since the behavior can never be directly observed. Thus some second-best approximation must be substituted—in this case an estimate on the part of the cadet's instructor of how successful he is likely to be in such situations if they were ever to occur.

Ultimate criteria by definition also have a distal characteristic. However, immediate criteria should be considered and selected with the ultimate criterion in mind. In other words, there exists in theory a true measure of the degree of job success or "satisfactoriness" of an employee in any situation at any moment in time. The goal is to obtain an approximate estimate of this ultimate criterion by selecting one or more actual criteria which we think are appropriate.

CRITERION DEFICIENCY, RELEVANCE, AND CONTAMINATION

The three major constructs of a criterion are *deficiency*, *relevance*, and *contamination*. The following diagram, in which the circles represent variances in the same manner as was introduced in Chapter 2, is intended to define these terms graphically.

Figure 6.2 illustrates a hypothetical degree of overlap or agreement between the actual measured criterion in a given situation and the "ultimate" or "true" criterion corresponding to the same situation.

Criterion deficiency is the degree to which our criterion is lacking certain variance necessary to the ultimate criterion. Unless there is perfect overlap between the actual and the ultimate criterion, there will be some portion of the ultimate criterion *not* related to the actual criterion being used. This area may be thought of as the degree to which our actual criterion is deficient.

Criterion relevance is the degree to which the actual criterion overlaps or corre-

as in the case of total sales over the first two years of employment. Obviously, total sales over the first two years will include the first month's sales, the second month's sales, etc. In such situations it would probably still be best to view such measures as distal criteria.

ACTUAL VERSUS ULTIMATE CRITERIA

A somewhat different way of viewing criteria is in terms of actual and ultimate measures of success. By definition, whatever measure of success one actually uses is the *actual criterion*. The *ultimate criterion*, on the other hand, is a theoretical and ideal criterion that usually exists only in the psychologist's mind. It is the "true" criterion of success, while our actual criterion is the measure we have been forced to adopt simply because we can do no better.

Perhaps a simple illustration will serve to clarify the concept of the ultimate criterion. Suppose we are interested in developing a test battery to select air cadets. We have to select some measure (or set of measures) which define, to our immediate satisfaction, how successful each air cadet will be. We might, for example, adopt as our criterion of success the rating by the instructor at the end of training. This rating would then become our actual criterion of success against which we would validate our test battery.

However, from a more ideal point of view we would probably say that instructor's ratings are not the ultimate measure of success of an air cadet. In fact, it is exceedingly difficult to decide and define what the ultimate measure of success is in this particular situation. From a military point of view, the ultimate criterion of success for an air cadet might be defined strictly in terms of how successful the cadet is when flying actual combat missions. In a peacetime world such a criterion can never be used, since the behavior can never be directly observed. Thus some second-best approximation must be substituted—in this case an estimate on the part of the cadet's instructor of how successful he is likely to be in such situations if they were ever to occur.

Ultimate criteria by definition also have a distal characteristic. However, immediate criteria should be considered and selected with the ultimate criterion in mind. In other words, there exists in theory a true measure of the degree of job success or "satisfactoriness" of an employee in any situation at any moment in time. The goal is to obtain an approximate estimate of this ultimate criterion by selecting one or more actual criteria which we think are appropriate.

CRITERION DEFICIENCY, RELEVANCE, AND CONTAMINATION

The three major constructs of a criterion are *deficiency*, *relevance*, and *contamination*. The following diagram, in which the circles represent variances in the same manner as was introduced in Chapter 2, is intended to define these terms graphically.

Figure 6.2 illustrates a hypothetical degree of overlap or agreement between the actual measured criterion in a given situation and the "ultimate" or "true" criterion corresponding to the same situation.

Criterion deficiency is the degree to which our criterion is lacking certain variance necessary to the ultimate criterion. Unless there is perfect overlap between the actual and the ultimate criterion, there will be some portion of the ultimate criterion *not* related to the actual criterion being used. This area may be thought of as the degree to which our actual criterion is deficient.

Criterion relevance is the degree to which the actual criterion overlaps or corre-

sponds to the true criterion. This corresponds to the term "validity" in the earlier sections, i.e., it is the extent to which our actual criterion is a valid criterion. However, since we can rarely, if ever, actually measure the ultimate criterion, we can never get an empirical measure of criterion relevance.

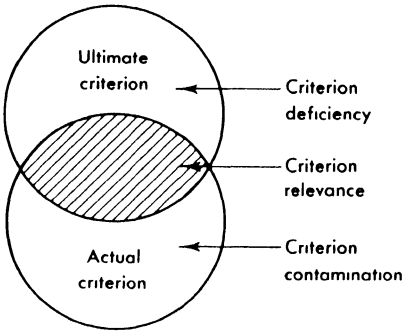


Figure 6.2. Diagram illustrating the criterion (constructs) of deficiency, relevance, and contamination.

Criterion contamination is that variance in the actual criterion which is unrelated to the ultimate criterion. It may be thought of as consisting of two distinct parts. First, there is systematic criterion contamination, which can be considered criterion *bias*. Second, there is random criterion contamination which can be considered criterion *error*. The error portion is not correlated with anything, while the bias part of the contamination is correlated with one or more other variables.

Criterion deficiency and criterion contamination are undesirable characteristics of the criterion. In fact, to the extent that they exist, our actual criterion is "distorted." Brogden and Taylor (1950) refer to three types of *criterion distortion*: (1) criterion deficiency, (2) criterion contamination, and (3) what they call "scale unit bias." Specifically, as they define distortion, it is simply giving an improper weight to a variable in constructing the criterion measure. Thus if we decide a variable should be included in the criterion (that is, we give it a nonzero weight) when in fact it does not belong, we introduce contamination error. If we fail to include an important dimension of the ultimate criterion (give it a zero weight) then we introduce deficiency error.

A MORE COMPLETE MODEL FOR PREDICTION SITUATIONS

Figure 6.2 introduces the major criterion constructs. However, in cases dealing with prediction the diagram should include a predictor as well as the actual and ultimate criterion. If we include a predictor, the scheme becomes a little more complicated, as shown in Figure 6.3.

There are nine different areas represented in the diagram, each of which is important. Seven are obvious and are lettered A–G; the remaining two are areas E + F and D + E.

A: Criterion deficiency which is *uncorrelated* with the predictor variable
B: Criterion relevance which is *uncorrelated* with the predictor variable
C: Criterion contamination which is *uncorrelated* with the predictor variable
D: Criterion deficiency which is *correlated* with the predictor variable
E: Criterion relevance which is *correlated* with the predictor variable
F: Criterion contamination which is *correlated* with the predictor variable
G: Uncorrelated predictor variance

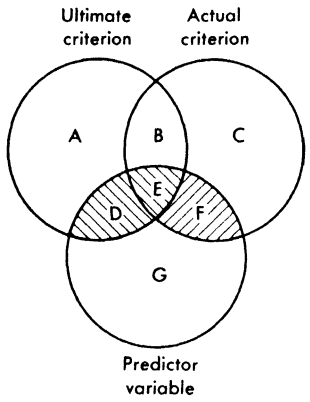


Figure 6.3 Diagram of basic criterion constructs when predictor variable is included.

Now let us consider certain combinations of the above areas.

E and F: This area represents the measured or obtained validity, since it is the amount of overlap between the predictor and the actual criterion Note that only part E of the obtained validity is really "true" validity, while part F is simply predictor correlated contamination bias, or measured false validity.

D and E: This area, if it could be measured, represents the correlation between the predictor and the ultimate criterion. Therefore it can be considered the "true" validity of that particular predictor variable. However only part E actually gets represented by the obtained validity coefficient.

Since area E plus F represents the obtained validity and area D plus E the true validity, then the relative sizes of areas D and F will determine whether the obtained validity is an overestimate or an underestimate of the true validity. If D is larger than F, then the obtained validity will be an underestimate of the true validity. On the other hand, if F is larger than D, then the obtained validity will be spuriously high. Unfortunately, in practice one has no way of ascertaining whether the empirical validity coefficient is an overestimate or an underestimate of the true validity.

CRITERION EXTENSION AND POPULATION EXTENSION

Two additional terms that need defining are *criterion extension* and *population extension*. These terms refer to the manner in which one cross-validates the validity of a test or a test battery (for example, see Mosier, 1951). Strictly speaking, the term "cross-validation" should only be used when one uses a sample of people drawn from the same population as in the original sample *and* when the same criterion measure is being used. The term criterion extension refers to validating the test against a new criterion on another sample from the same population—thus the effectiveness of this test is "extended" to a new criterion. The term *population extension* refers to those situations where the criterion remains the same, but the sample is drawn from a different population. Thus the effectiveness of the test may now be extended to a new group of people.

CRITERION EQUIVALENCE

Two criteria may be said to be *equivalent* if they are completely interchangeable without any loss in the predictive efficiency of the selection process. In other words, two criterion measures are equivalent if they correlate perfectly, assuming that one has corrected both for unreliability. Thus they are both measuring exactly the same characteristic of the individuals being assessed.

Very rarely does one find two different criteria which agree perfectly, although in some factor analytic studies (discussed in a later section) one finds clusters of nearly equivalent measures of job performance. There has been an increased emphasis upon the importance of criterion equivalence in recent years, particularly with respect to the way in which equivalence information among many measures can help provide greater insight into the overall criterion problem.

There are several ways of defining criterion equivalence besides perfect correlation, as has been pointed out by Wherry, Ross, and Wolins (1954, p. ii). They suggest six definitions for equivalence:

1. If two criteria are equivalent, they correlate perfectly with each other when random error is excluded
2. Correlations of equivalent criteria with a given set of predictors are identical if the criteria are equally reliable and proportional if the criteria are of unequal reliability.
3. Regression weights for a given set of predictors of two perfectly equivalent criteria are identical if the two criteria are perfectly reliable and proportional if the criteria differ only in reliability.
4. If the two criteria are equivalent, their loadings on the factors underlying any combination of possible predictors are identical or proportional
5. Two equivalent criteria contain the same basic elements from any job analysis system.
6. Equivalent criteria are predicted equally well by a battery selected to predict either of them, if errors of fitting are eliminated.

Nearly all of these definitions of equivalence are statistical. The only exception to this is definition 5, which defines equivalence in terms of the similarity of data obtained from a common job analysis method.

FURTHER COMMENTS ON CRITERION CONTAMINATION AND BIAS

Earlier the term *criterion contamination* was defined as that portion of the actual criterion which did not overlap with the ultimate criterion. Further, it was pointed out that criterion contamination consists of two parts, error and bias. Error, by definition, is usually considered random variation and cannot, except by chance, correlate with anything, including the predictor variable. However, the bias portion of criterion contamination is systematic variation, and it *can* correlate with the predictor. This can be seen from our model, where predictor-correlated contamination (which must therefore be predictor-correlated bias) contributes to a spuriously high obtained validity.

What sorts of bias are likely to enter into criterion contamination and what sorts are also apt to be correlated with the predictor variable? Perhaps Brogden and Taylor's (1950) formal definition of criterion bias will help answer this question. They define bias in terms of "any variable, except errors of measurement and sam-

pling error, producing a deviation of obtained criterion scores from a hypothetical 'true' criterion score."[1] Figure 6.4 illustrates this point.

Such biasing factors are more frequent than one might suspect and should be closely guarded against. They can occur because of a variety of circumstances and as a function of a number of variables.

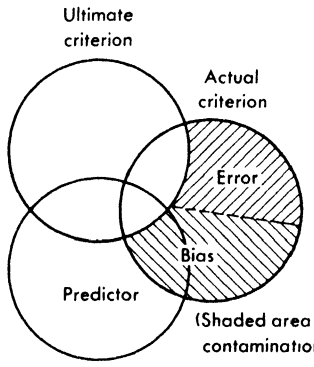


Figure 6.4. How the bias portion of criterion contamination contributes to a spuriously high validity coefficient.

OPPORTUNITY BIAS This kind of criterion bias refers to those situations where factors beyond the control of the worker considerably influence the amount produced. Examples of opportunity bias are numerous. In sales work, the particular territory or sales counter can affect the sales volume of an employee—perhaps even more so than his own ability. In factory work, opportunity bias can be attributed to such simple things as the quality of the tools and the condition of the machinery a worker has, whether he must work the day shift or the night shift, where his work station is located, how good his lighting is, and a variety of other variables.

In short, unless the work environments and the jobs are identical, comparing work performance records among individuals has its limitations. If work performance is correlated with predictor devices under such circumstances, one has no way of really knowing whether the resulting correlation is a correlation with true job performance or is primarily a correlation with, say, quality of one's work station. While opportunity bias does not necessarily correlate with the predictor, even if it is predictor-free, it will have an influence on the obtained validity.

Other more subtle examples of opportunity bias can often be found by closer examination of the predictor validities of such items as age and experience. In many jobs people who are older and who have had more experience get better performance scores simply due to circumstances of union seniority, in that these people get the better work stations, better sales counters, etc.

GROUP CHARACTERISTIC BIAS Another important source of criterion bias concerns the characteristics of the group to which a person belongs. For example, if a person belongs to a group which artificially limits the productivity of its members, there is going to be a clear biasing of criterion data. Similarly, if a person belongs to a group which has been preselected on some a priori basis, and if that group is then com-

[1] The more advanced reader will recognize that such a definition allows for the actual validity to increase, decrease, or remain the same due to such bias, since it does not specify the way in which the biasing variable produces a deviation from the true score. Thus if the deviation produced was always in the form of a linear transformation, no change in the validity would occur.

pared to a group not so selected, any selection variable related to the a priori variable is apt to show spurious validity. Brogden and Taylor (1950) give an example of the situation where an executive has decreed that all office boys hired must have a high school education. If, for purposes of test validation, a group of present employees is compared against a sample of job applicants, it is almost a certainty that any predictor likely to correlate with amount of education will demonstrate validity!

The problem of age and job tenure is another illustration of opportunity bias. When the job is one in which productivity tends to increase with additional experience and where a wide range of experience does exist (that is, there are both experienced and inexperienced employees), it is inevitable that the criterion (productivity) and experience will be correlated. Thus the criterion is biased by the characteristic of experience, and any predictor likely to correlate with experience is apt to end up with a spuriously high validity.

BIAS IN RATINGS One of the more frequently employed criteria in industry—ratings of competence by supervisors—is also subject to bias. While the chapter on performance appraisal discusses the rating problem in much greater detail, it should be pointed out here that all the usual sources of bias in production records are also likely to bias ratings of individual competence. Can supervisors, in making their ratings of competence, make adjustments for unequal opportunity among those being rated? If they can, then their ratings are going to lead to a less biased criterion than the actual production record would be.

The most often cited biasing effect in ratings is the phenomenon known as the "halo effect." Briefly, this refers to the situation where a rater considers a person to be outstanding on all traits merely because he (the person being evaluated) possesses one outstanding characteristic. For example, if a person has performed outstandingly on one aspect of job performance, a halo error would be made if we assumed the person must also be good on all other aspects of his work. Halo errors often occur when a rater is asked to rate individuals with pleasant personality characteristics ("If a person is nice, he must also be capable") and pleasant features ("If a girl is pretty, she must also be talented"). Of course, tests which correlate highly with ratings having a halo bias may simply be correlating with the halo variable rather than with "true" job competence. Chapter 7 will examine rating error and bias problems in greater detail.

KNOWLEDGE OF PREDICTOR BIAS When criterion data are obtained using ratings, the most serious error which can occur is for the rater to have prior knowledge of the predictor scores. This is called *knowledge of predictor bias*. If a rater knows the scores the people achieved on the predictor variable, it is entirely possible that he might let his criterion judgments be influenced by this knowledge. Such an error can cause an increase of an entirely spurious nature in the obtained validity. Unfortunately, not only is this type of error most serious in criterion ratings (since by definition it is a bias that will always be predictor correlated), but also it is one of the most frequently found errors. Examples are perhaps easiest to find in school settings, where teachers make judgments about the academic talent of their students after having seen their entrance test scores. One cannot overemphasize the strict need always to keep predictor information isolated from individuals who are providing criterion data. Thus, supervisors should never be permitted access to an

applicant's test battery scores. Once such contamination bias has occurred, any subsequent attempts to validate the test empirically can readily result in overinflated correlation coefficients.

REQUIREMENTS OF CRITERIA

Weitz (1961) has made a strong plea for developing what he calls "criteria for criteria" in psychological research. He points out that there are really three main criterion questions: *time, type,* and *level.*

Time: When should one obtain his criterion data? At what moment in time is our criterion data not "ripe"? For example, in training, if one wants to know the relative superiority of training method A to training method B, should one measure the effects immediately after training, a year later, or ten years later?

Type: Of all the performance measures available, which one or ones should be selected? How dependent are our conclusions likely to be simply due to the criterion selected?

Level: What performance level (given that we have decided upon a time and a type) is going to be considered acceptable? How many units must a person produce before he is considered to be a good worker? The question of criterion level, of course, will only occur in those situations when it is necessary to dichotomize the underlying variable into two categories of success or failure.

More specifically, Bellows (1954) lists a number of characteristics which are necessary and/or desirable in any criterion:

1. Reliable
2. Realistic
3. Representative
4. Related to other criteria
5. Acceptable to job analyst
6. Acceptable to management
7. Consistent from one situation to another
8. Predictable

To this list perhaps several other items need to be added, such as:

9. Inexpensive
10. Understandable
11. Measurable
12. Relevant
13. Uncontaminated and bias-free
14. Discriminating

Obviously, some of these 14 criterion requirements are more important than others. Some, such as reliability, relevance, and freedom from bias, are of paramount importance in any prediction situation. Others, such as discriminative ability, cost, understandability, etc., vary in importance depending upon the particular situation. Still others, such as relation to other criterion measures and being acceptable to management and job analysts, can be argued. However, all fourteen should at least be considered in any situation where a new criterion is being selected.

CRITERION DEVELOPMENT

Several authors have outlined what they consider to be the appropriate steps in the development of a criterion for a particular situation. Guion (1961), for example, has listed with wry humor what he considers to be the usual steps to criterion development and what he considers to be the desirable steps to criterion development:

Usual Sequence:

1. The psychologist has a hunch (or insight!) that a problem exists and that he can help solve it.
2. He reads a vague, ambiguous description of the job.
3. From these faint stimuli, he formulates a fuzzy concept of an ultimate criterion.
4. Being a practical psychologist, he may then formulate a combination of several variables which will give him—as nearly as he can guess—a single, composite measure of "satisfactoriness."
5. He judges the relevance of this measure: the extent to which it is neither deficient or contaminated.
6. He may judge the relative importance of each of the elements in his composite and assign some varying amount of weight to each.
7. He then finds that the ideas required for his carefully built composite are not available in the company files, nor is there any immediate prospect of having such records reliably kept.
8. Therefore, he will then select "the best available criterion." Typically, this will be a rating, and the criterion problem, if not solved, can at least be overlooked for the rest of the research.

Desired Sequence:

1. Analyze the job and/or the organizational needs by new, yet-to-be-developed techniques.
2. Develop measures of actual behavior relative to the behavior expected, as identified in job and need analysis. These measures are to supplement measures of the consequences of work—the so-called objective criteria commonly tried at present.
3. Identify the criterion dimensions underlying such measures by factor analysis or cluster analysis or pattern analysis.
4. Develop reliable measures, each with high construct validity, of the elements so identified.
5. For each independent variable (predictor), determine its predictive validity for each one of the foregoing criterion measures, taking them one at a time.

Nagle (1953) has also set forth a sequence which one should attempt to follow in developing a criterion measure. His sequence consists of only four steps:

1. Define the activity
2. Analyze the activity
   a. Purposes of the activity
   b. Types of behavior called for
   c. Standards of performance in the activity
   d. Relative importance of the various behaviors
3. Define success
   a. Find the elements of success
   b. Weight the elements of success
4. Develop sub-criteria to measure each element of success
   a. Relevancy of each sub-criterion to its success element
   b. Reliability of each sub-criterion
   c. Combination of the sub-criteria

A comparison of the list of Guion and Nagle indicates a slightly different philosophy about what a criterion ought to be. For example, Guion is quite specific about how the elements of job success are to be defined (see step 3), while Nagle tends to be less specific (see his step 3). Nagle also tends to prefer the use of a single-weighted composite criterion, while Guion favors evaluating (predicting) each dimension of success separately. This is an important distinction and is discussed in more detail in the following section.

COMPOSITE VERSUS MULTIPLE CRITERIA

For years many industrial psychologists have labored under the notion that the objective is to find "the criterion" in a selection situation in the same way that the knights of King Arthur's Round Table were charged with finding "the Holy Grail." Both have had about equal and limited success.

COMPOSITE CRITERIA

The notion of "the criterion" is one which implies that there is, in any selection situation, a single *best* criterion measure. Such a criterion is usually referred to as a *composite* criterion. Early in the history of selection, the "best criterion" was a tag placed upon that single measure of job success which the psychologist felt was most satisfactory according to whatever set of criteria he happened to have at hand. This view gradually disappeared with the more sophisticated concept of the composite criterion. This point of view holds that true job success (our ultimate criterion, if you will) is multidimensional in nature. Thus, it would logically follow that any actual criterion used by the psychologist must also be multidimensional or it will lack relevance. This actual criterion became referred to as a composite criterion measure since it was composed of several individual aspects or measures of job success. The major advocates of the composite criterion have been Toops (1944) and Nagle (1953).

If one accepts the premise of a single composite measure of success as the correct approach to criterion development, one must of necessity also face the further problem of how to create this global entity from a group of individual measures. In other words, how should one combine a collection of various measures of success into one composite evaluative index. There are a number of alternative procedures which have been used to combine subcriteria—each predicted upon a somewhat different philosophy. Nagle (1953) has outlined these various alternatives.

1. *Weighting subcriteria using "expert" judgments.* This method is the method preferred by Nagle and was first proposed by Toops (1944). Using this procedure involves having a number of experts examine the separate criteria and assign weights to each in accordance with their own perception of the respective importance of the individual measures. To illustrate, suppose one had five measures of job success and the task was to combine these into a composite measure of job performance.

Let c_1 = number of units produced per unit time
c_2 = average quality of a produced unit
c_3 = number of accidents per unit time
c_4 = rating of immediate supervisor
c_5 = number of times tardy

The basic equation for the composite criterion becomes (assuming a linear model)

$$C = w_1c_1 + w_2c_2 + w_3c_3 + w_4c_4 + w_5c_5$$

and the task of each expert is to decide what values should be assigned w_1 through w_5. For example, we might have three different sets of weights given us by five different judges, as shown in Table 6.1.

TABLE 6.1 *Weights Assigned to Each Individual Criterion by Five Different Experts*

	w_1	w_2	w_3	w_4	w_5
E_1	0.8	0.2	0.0	0.0	0.0
E_2	0.6	0.2	−0.1	0.1	0.0
E_3	0.6	0.1	−0.1	0.1	−0.1
E_4	0.3	0.3	−0.2	0.0	−0.2
E_5	0.2	0.2	−0.1	0.3	−0.2
Sum:	2.5	1.0	−0.5	0.5	−0.5
Avg.:	0.5	0.2	−0.1	0.1	−0.1

The basic equation, if we now use the average of these five experts as our best estimate, becomes

$$C = 0.5c_1 + 0.2c_2 - 0.1c_3 + 0.1c_4 - 0.1c_5$$

Notice that two subcriteria are negatively weighted in the equation as might be expected, since number of accidents and being tardy are generally considered as undesirable aspects of job behavior.

The development of a composite criterion by this method is equivalent to doing so on the basis of *judged criterion relevancy*. Both Nagle and Toops take the position that relevancy of subcriteria is the *only* basis on which to form a composite. The present authors would agree with this position, but would take issue with the more basic question of whether a composite criterion is the most desirable approach.

One problem with judged relevancy that must also be faced concerns the question of who will play the role of "expert." Selecting "experts" is a problem in using this procedure; often there are few, if any, individuals available who can be considered knowledgeable and unbiased enough to be given the task of assigning weights to subcriteria. For these reasons this method must often be abandoned, since the very concept of relevancy only makes sense if the judges are intimately familiar with the job in question, its particular mission and goals, and the mission and goals of the entire organization.

2. *Weighting subcriteria proportional to their reliabilities.* This procedure for obtaining a composite criterion gives large weights to those individual performance measures which are most reliable. It is predicated upon the point of view that unreliable criteria, no matter what their relevancy, are of little value in actual practice. While such is true, one must also point out that no matter how reliable a subcriterion may be, it is of no practical value unless it has some degree of relevance. Therefore to weight strictly on the basis of reliability does not seem to be a defensible procedure. However, given subcriteria of approximately equivalent judged relevance

but substantially different reliabilities, it would certainly seem prudent to select the more reliable of the two.

3. *Weighting subcriteria proportional to their average correlation with other variables.* This procedure is based on a purely statistical process and is not a particularly defensible technique. One must, of course, ask what "other variables" are involved. If they are a set of predictor-correlated variables, then the procedure has some similarity to procedure number 4. If they are a set of variables chosen at random, then the method tends to weight most highly those variables which (1) are most reliable, and (2) are most related to any underlying *general factor* in the data (as defined by factor analysis).

4. *Weighting subcriteria on the basis of their predictability.* This method has been referred to as the "most predictable criterion" procedure (Hotelling, 1955). One deliberately weights the subcriteria to achieve the maximum correlation between the predictor and the composite criterion.

Justification for this combining procedure is very hard to manufacture. While it certainly guarantees that one will end up with a correlation of substantial magnitude, the question must be raised of how realistic this correlation is as an estimate of the true validity. Obviously, this procedure makes no attempt to use relevance as a basis for combining. The end result may merely be a high correlation with a criterion of little or no relevance whatsoever, thus deceiving both management and psychologist together. As Nagle (1953) points out, the predictors one uses should never have an influence on the composition or the weighting of the subcriteria (with the possible exception of subcriteria which correlate zero with the predictor).

5. *Selecting and weighting subcriteria by factor analysis.* In an earlier chapter factor analysis was explained as a method which grouped variables into homogenous clusters on the basis of their intercorrelation. As such, it has been a method which some have used to determine what the basic dimensions of the composite criterion should be. While it can be and has been effectively used for this purpose, it does not provide a method for combining dimensions into a composite. Therefore, even though one might be able to find a variable that is a pure measure of each factor, there still remains the problem of how to weight these variables into a composite measure.[2]

6. *Weighting subcriteria to maximize the difference between individuals.* This method attempts to increase the range of the criterion scores as much as possible, thereby making discrimination on the criterion dimension easier to accomplish. Statistically the method involves giving larger weights to those subcriteria which are highly intercorrelated, or to weighting any underlying general factor in the subcriteria.[3] The method probably only has value, however, when job success is composed of a single underlying factor. Since such cases are rare (the authors cannot think of an example, but the reader may be more creative) the method seems to have little to recommend it.

7. *Equal weighting of subcriteria.* One way of avoiding the problem of deciding which subcriteria should receive the greatest weight in the composite is to

[2] One method that has been suggested is to weight the dimensions relative to the proportion of systematic variance attributable to each factor. Since this is highly dependent upon the original choice of variable, this does not seem a desirable solution.

[3] The variance of a set of composite scores is related to the variance of the separate variables as $\sigma_C^2 = \sigma_A^2 + \sigma_B^2 + 2r_{AB}\sigma_A\sigma_B$. Thus increasing σ_A or σ_B (by using a large weight) is going to increase σ_C^2 more in those cases where r_{AB} is large.

weight all subcriteria equally. This, of course, assumes that all individual aspects of job success are equally important. While such an assumption is undoubtedly always going to be incorrect, one could argue that in the long run, since we can only guess at the correct weighting anyway (and are thus usually going to be off in our estimate of weights), we will make less error if we simply give all variables an equal weighting. There is a good deal of merit to such an argument—particularly when the "experts" available for making relevancy judgments are not of the desired caliber.

8. *Weighting on the basis of dollars cost.* Brogden and Taylor (1950) have argued for what they call the "dollar criterion" as a way of handling the weighting problem when combining various aspects of job success into a composite criterion measure. Their procedure is deceptively simple and turns the problem over to cost accountants. They suggest, first of all, that the ultimate criterion for any business firm must be one of dollars and cents. Thus *all* subcriteria should be transformed into dollar values. For example, if the amount of profit to the company on each unit produced were allocated appropriately (say on a percentage basis according to judged contribution) across all those participating in its construction, one might find that there is five cents per unit profit allocated to worker A for every unit produced. Thus, a worker who produces 100 units each day is worth $5 in daily profit, while another worker who produces 500 units per day is worth $25.

Let us further suppose that our cost accounting department tells us that every broken unit costs the company $10.00 in repairs, time lost, etc. Thus if worker A above breaks an average of two units each day, he costs the company $20.00, leaving a profit of $5.00, the same as worker B who breaks no units (obviously a very slow but careful worker). This method suggests a way of combining breakage and units produced into a single composite criterion. Brogden and Taylor argue that the conversion to dollars gives the only really correct weighting for combining subcriteria.

The proposal of a dollar criterion has much to recommend it. If the reader will take a moment to refer back to the list of requirements for a criterion, he will find that a dollar criterion would meet most of them. There are, however, two major difficulties in employing a dollar criterion. First is the tremendous difficulty of converting all measures of job success to a dollar base. How does one decide what proportion of the total profit on each unit should be allocated to each worker? The task faced by the cost accounting department in trying to set up such a criterion system would be impossible. The second major difficulty is that some measures of job success are just not expressible in monetary terms. For instance, a company might feel that one requirement for being a good executive is to participate in community affairs and to create a good image in the community. How can such things as degree of community participation be converted to a scale of dollar value to the company? Obviously they cannot be so converted. The advantage of the dollar criterion approach to a composite criterion measure is therefore probably restricted to those situations where the job is rather basic and the subcriteria are all easily related to dollar-and-cents return to the company.

MULTIPLE CRITERIA

Although the concept of a workable composite criterion has a great deal of intuitive appeal, perhaps because of its logic and its seeming simplicity, there appears to be an increasing disenchantment taking place with its use in modern selection

research. A number of rather eminent industrial psychologists have been leading the attack on the composite criterion, suggesting instead that a procedure of multiple criteria be adopted without attempting to combine these individual measures into any global composite measure of success.

Ghiselli (1956) was probably the first person to rebel against a criterion composite. He strongly argued for treating separately (predicting) each dimension of job success. To quote Ghiselli (1956, p. 2):

> If the proposition is accepted that criteria are multidimensional with the dimensions being independent, or at least relatively so, then the situation is not an easy one. There is no way to combine the independent scores of an individual into a single value that will describe him uniquely. Rather, it will be necessary to locate his position in the multidimensional criterion space. This can be accomplished in either of two ways: each criterion dimension can be predicted separately and the individual's position in the space estimated, or the space can be divided into parts and that portion of the space in which the individual is most likely to fall could be estimated by the discriminate function.

To Ghiselli, the problem of multidimensionality was even more severe than others acknowledged. He pointed out that not only was success usually a multidimensional concept, but that there were at least three different types of criterion multidimensionality which needed consideration. He called them "static dimensionality," "dynamic dimensionality," and "individual dimensionality."

STATIC CRITERION DIMENSIONS If we consider the criterion development problem at any single point in time, we usually will find that job success is multidimensional, that is, there are a variety of different ways of defining success on that job at that moment. These dimensions are called the *static dimensions* and are the kind of criterion dimensions most usually considered in dealing with the criterion problem.

DYNAMIC CRITERION DIMENSIONS Nearly all discussions of criterion dimensionality fail to deal adequately with the fact that the performance of individuals on the various dimensions will change over time as they learn and develop on the job. The fact that such changes do indeed occur raises the further question of possible changes in the intercorrelations between criterion dimensions. If the correlations between dimensions change as a function of practice or experience, this would indicate that the *basic* dimensions are also changing. Thus, any prediction of a weighting scheme for the dimensions of success set up early in the game might need drastic revision as time passes.

Several empirical studies have clearly demonstrated the dynamic nature of criterion dimensions, most notably those of Ghiselli and Haire (1960), Bass (1962), and Fleishman and Fruchter (1960). They found the factorial composition of performance to vary with practice. A detailed discussion of the Fleishman and Fruchter research on this topic can be found in Chapter 8 ("Training and Learning"), since it is more appropriately treated in that context. The Ghiselli and Haire study showed that the intercorrelations between weekly production records remained rather stable over time. However, they also found that the best predictors of productivity changed over time. In fact, they found quite different predictors for (1) early productivity, (2) late productivity, and (3) rate of increase in productivity.

INDIVIDUAL CRITERION DIMENSIONS Ghiselli's third type of criterion dimensionality is somewhat more ambiguous than the first two. He suggests that several workers

on the same job may be considered to be equally good, and yet the nature of their contributions to their organization might be quite different. He therefore argues that often workers on the same job might need to be evaluated in terms of different criterion dimensions.

It is hard to see how this type of dimensionality differs from static dimensionality. In fact, at times Ghiselli gets strikingly close to returning to the composite criterion concept, since the term "equally good" implies a common standard. For example, he remarks that "this is not saying that an individual is to be considered good if he is high on any one criterion variable. Rather the notion is that while certain criterion variables are appropriate in describing the performance of some workers, they just are not pertinent in describing the performance of other workers on exactly the same job" (Ghiselli, 1956, p. 3).

Again, in this statement there seems to be some indication of an all important "goodness" criterion lurking in the background.

Guion (1961, 1965) has also strongly suggested a move away from the notion of the composite criterion. His list of steps for criterion development given earlier gives evidence of his preference for obtaining validities for each separate measure of job success (Guion, 1961, p. 145).

> A broad and useful definition of a criterion is behavior, or consequences of behavior, that one wishes to predict. The fallacy of the single criterion lies in its assumption that everything that is to be predicted is related to everything else to be predicted—that there is a general factor in all criteria accounting for virtually all of the important variance in behavior at work and its various consequences of value. . . .
>
> Where criterion elements or dimensions are shown to be related, then there may be some point in combining them into a general composite. Where they are clearly independent, however, then prediction should also be independent.

Another vociferous advocate of abandoning the composite measure of success is Dunnette. Some of his more forceful comments (1963, p. 252) will serve to illustrate his position.

> The point of all this is to suggest that much selection and validation research has gone astray because of an overzealous worshiping of the criterion with an accompanying will-o'-the-wisp searching for a best single measure of job success. The result has been an oversimplification of the complexities involved in test validation and the prediction of employee success. Investigators have been reluctant to consider the many facets of success and the concomitant investigation of the prediction of many success measures and instead persist in an unfruitful effort to predict the criterion. Thus, I say: junk the criterion! Let us cease searching for single or composite measures of job success and proceed to undertake research which accepts the world of success dimensionality as it really exists.

He further states (p. 253):

> I suggest, therefore, that we cease talking about the criterion problem and that the notion of an ultimate criterion be disregarded. In so doing, a change in research emphasis should occur which will focus on defining the meaning of scores on any given test in terms of a variety of other (both test and nontest) behaviors rather than only in terms of some complex single or composite measure of job success. As a consequence, validation studies will become less restrictive and simple-minded, and will result in broader knowledge about the full meanings and uses of our predictor batteries.

SUMMARY OF THE CONTROVERSY

The problem of composite versus multiple criteria has a variety of different facets, and both major points of view have certain merit. However, the main disenchantment that has arisen with the composite criterion in recent years seems to stem from either or both of two major sources: the weighting question and the need for progress. By examining these two sources, the controversy can be placed in better perspective.

WEIGHTING PROBLEM The first major objection to the composite criterion reduces to the problem of weighting subcriteria into an overall measure of job success. Those who oppose the composite argue that one can only legitimately combine measures when they are measures of the same thing (for example, dollars), and that to try to combine into a single measure subcriteria which are not compatible is a deceiving and illegitimate process. An example given by Ghiselli is that of two college professors. One professor might be considered good because he is an excellent teacher. The other professor might be considered good because his research is highly significant. Since these are two rather different dimensions of job success (the correlation between them is probably low positive), combining them into a composite measure of goodness necessitates (1) determining their relative importance (weight) and (2) developing a common scale unit for each dimension. Both tasks, of course, would be exceedingly difficult. Therefore, we should not attempt to combine but should set up ways of predicting each dimension separately.

On the other hand, as long as *both* dimensions are related in some fashion to our concept of overall goodness of a college professor, the composite criterion can be legitimately defended, even if one admits to the difficulty of its practical implementation.

In a sense, the problem of weighting independent dimensions into a composite on the criterion side is very similar to the problem of weighting independent predictors into a composite on the prediction side. In Chapter 3 we discussed the arguments for and against the multiple regression model and the multiple cutoff model. Those who are opposed to the multiple regression prediction model have pointed out that the different predictors are often not additive and therefore cannot and should not be substituted for each other—instead, a system of cutoff scores should be employed wherein each predictor is used separately.

Those who disapprove of the composite criterion argue not only that success is a many-faceted thing in most jobs, but also that these separate aspects are *not additive* and, therefore, cannot be legitimately combined into a global measure of success. Going back to the example of college professors, consider a situation when the Dean has declared, "I want my professors to be *either* top-notch teachers *or* top-notch researchers, but I don't want anyone trying to be both!" Obviously he refers to separate yardsticks, each with its own cutoff score. Under the Dean's system, all that is necessary to be considered *successful* is for a person to be *above cutoff* on one criterion dimension. This is the analog of the predictor case where all that is necessary to be considered *unsuccessful* is for a person to be *below cutoff* on one predictor variable. Of course, the composite criterion advocate could reply in rebuttal that the Dean has taken the single job of college professor and turned it into two distinctly different jobs—college professor (teaching) and college professor (research)—and that *obviously* one needs different criteria for different jobs.

There is not a definite answer to the weighting problem at the present time. To the authors, however, the *logic* of an ultimate composite criterion for any specific job does seem hard to deny. For example, one can usually ask supervisors to rate or to rank their subordinates according to some global criterion such as "worth to the company" and get fairly respectable agreement between supervisors in terms of the order in which they rank the employees. "Worth to the company" must certainly be a composite criterion, and the various supervisors must be weighting a variety of subcriteria in arriving at their decisions. Yet the fact that they are often able to do so reliably must imply that (1) the subcriteria are indeed additive into a composite, and (2) the judges must have some agreement, even if they are not able to overtly express it in words, concerning the weights that are to be attached to the various subcriteria.

TABLE 6.2 *Sample Profile of a Simulated Job Incumbent for Rater Group A*

Profile 1

Trait	*Very Poor* 1	2	*Poor* 3	4	*Average* 5	6	*Good* 7	8	*Very Good* 9
1. Utilizes opinions of subordinates					5				
2. Maintains discipline				4					
3. Ability to communicate						6			
4. Knows theory of aircraft systems					5				
5. Ability to teach others							7		
6. Knows tech. orders-manuals-sops						6			
7. Willing to delegate work						6			
8. Careful with property and equipment		2							
9. Perseverance on the job		2							
10. Careful and thorough in inspections					5				
11. Proficient in trouble shooting						6			
12. Monitors work of subordinates					5				
13. Safe work habits			3						
14. Willingness to learn				4					
15. Willingness to train his men					5				
16. Plans and organizes his work					5				
17. Supports his men				4					
18. Ability to perform under pressure						6			
19. Willingness to make decisions								8	
20. Informs his men						6			
21. Fairness and impartiality				4					
22. Job ingenuity					5				
23. Punctuality and dependability		2							
	1	2	3	4	5	6	7	8	9

SOURCE: J. C. Naylor and R. J. Wherry. Feasibility of distinguishing supervisors' policies in evaluation of subordinates using ratings of simulated job incumbents. USAF PRL Technical Documents Report, 64–25, October 1964.

Naylor-Wherry Study. Both implications in the preceding sentence were substantiated in an extensive study by Naylor and Wherry (1964). Working with four different Air Force career specialties, they asked the following two questions:

(1) Can supervisors reliably rank order their subordinates in terms of a global criterion "worth to the Air Force" when given a large amount of information, in the form of a profile, about each subordinate?

(2) Will the supervisors tend to place similar weights on the different profile elements in making their judgment of overall worth?

Fifty supervisors in each of the four different career specialties were each asked to examine 250 profiles of men in their specialty (but of lower rank) similar to that shown in Table 6.2. After examining each profile, they were to indicate that individual's worth to the Air Force using the nine-point scale of "Worth to the Air Force."

After obtaining the judgments, it was then possible to compute the correlations for each rater between each profile characteristic and the rater's criterion judgments. For example, the correlations for three such raters are given in Table 6.3.

TABLE 6.3 *Correlations Between Profile Characteristics and Judged Worth to the Air Force for Selected Raters*

	Profile Characteristics																						
Rater	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*
1	53	45	48	54	64	61	62	44	70	65	73	68	53	61	43	50	52	68	58	52	48	51	39
3	51	44	44	51	62	57	58	41	67	69	68	67	46	57	41	52	44	70	58	52	47	44	45
11	52	45	48	50	58	64	59	35	71	73	67	70	43	59	38	50	45	70	58	52	44	44	41

SOURCE: J. C. Naylor and R. J. Wherry. Feasibility of distinguishing supervisors' policies in evaluation of subordinates using ratings of simulated job incumbents. USAF PRL Technical Documents, 64–25, October 1964.

Notice the high degree of similarity between the trait-criterion correlations for each rater. Quite clearly the supervisors seemed to feel pretty much the same about the way in which these 23 subcriteria were related (or should be combined) into a global or composite criterion of job worth. The three raters shown in Table 6.3 are simply illustrative of the findings by Naylor and Wherry: They found there was a great deal of agreement among supervisors *within a specialty* as to the weighting of the subcriteria. As might be expected, there were substantial differences in weights *between specialties*, that is, supervisors in different job categories tended to have different notions of how the composite criterion was to be constructed out of the subcriteria.

It is also interesting that the 23 profile traits used in the study were not all correlated. Thus in combining or weighting the subcriteria, the supervisors were combining independent aspects of job success and were able to do so reliably. The study would therefore seem to indicate clearly that the weighting problem is perhaps not the major objection to the use of the composite criterion since it is possible reliably to combine independent aspects of job success into a composite and to obtain the relative weights used in the combining process. If combining independent subcriteria into a composite had not made sense to the supervisors, their resulting judgments would have been quite unreliable.[4]

[4] It should be pointed out that the technique used in the Naylor-Wherry study was one called judgment analysis (JAN), a method initially suggested for this use by Bottenberg and Christal (1961).

LACK OF PROGRESS IN SELECTION The problem of weighting, however, has not been the only source of dissatisfaction with the use of a composite criterion. A second major objection is that there has been little progress in the efficiency of selection techniques over the years. This problem was mentioned in Chapter 2, in which it was pointed out that validities are no higher today, on the average, than they were a decade earlier. In Chapter 2 it was suggested that one reason for this may lie in the type of predictors which the psychologist has used. However, another equally likely explanation may lie in the way in which we have handled the criterion development problem.

Wallace (1965) has suggested that it is the failure to do any creative research on the criterion problem that has led to this difficulty, particularly the tendency to be concerned with a statistical relationship between a poorly defined composite criterion and a set of predictors. To be so concerned is simply to perpetuate the current state of affairs in prediction. He suggests that only by looking individually at all kinds of subcriteria (that is, relating them to predictors and to each other) to gain a better understanding, *regardless* of the relevance of these subcriteria, will we see progress in developing better foundations to selection and placement. Wallace (1965, p. 417) states:

> In the effort to provide such foundations, I believe we must expand our thinking about criteria and validity beyond the point of utility alone. We must reexamine our relations to and communications with psychologists in the test-and-measurement, experimental, and clinical fields. We must question our current fertility in the generation of psychological thought. We must somehow convince our sponsors and ourselves that it is as important to gain insight into why our procedures do or do not work as it is to produce a tried-and-true predictive gimmick. Doing this involves many steps, but the first and most important is to recognize that the answer to the question "Criteria for What?" must always include —for understanding!

In conclusion, the more advantageous position is probably that of multiple rather than composite criteria—if for no other reasons than that more information about the criterion situation is probably attained via the approach and that obtaining relevance weights is often difficult because of too few or unknowledgeable judges. However, the logic of the composite criterion is unassailable.

THEORY AND CRITERIA IN INDUSTRIAL PSYCHOLOGY

It is interesting that although it does not make sense to talk about a theory of criteria per se, it is probably impossible to develop any type of behavioral theory in industrial psychology without a criterion. In fact, theory in industrial psychology may only be possible if one *ignores* the criterion problem—otherwise we would never get around to being able to theorize.

As an example, suppose one is concerned with the need-hierarchy theory of Maslow as applied to motivation of managerial personnel. The theory itself does not concern itself with how to measure the different needs—it simply defines what is meant by each particular type of need. The researcher attempting to substantiate or disprove the theory must decide how he is going to measure each of these needs. These arbitrary measures then become his criteria. Obviously, his research finds are going to be important for the theory only to the extent that his criteria are relevant, that is, they truly represent the constructs of the theorist.

One might wonder at what would result if theorists were restricted to theorizing only in terms of empirical or actual criteria rather than in terms of the original constructs. Indeed, we might suggest that perhaps this has already become the case all too often in industrial psychology. Do industrial psychologists worry so much about the criterion that they tend to be reluctant to theorize in terms of constructs? In the viewpoint of the authors, the answer would have to be yes. An overconcern with the criterion problem can only lead to a *restriction* on theory development. The theorist must not have to worry about how to measure—let that problem come later. Only when the industrial psychologist becomes willing to release his secure hold on a measurable criterion while he does his thinking will we be likely to see any substantial advances in theoretical developments.

REFERENCES

Bass, B. M. Further evidence on the dynamic character of criteria. *Personnel Psychology*, 1962, *15*, 93–97.

Bellows, R. M. *Psychology of personnel in business and industry*, 2nd ed. Prentice-Hall, Englewood Cliffs, N.J., 1954.

Bottenberg, R. A., and R. E. Christal. An iterative technique for clustering criteria which retains optimum predictive efficiency. WADD-TN-61-30, Personnel Laboratory, Lackland Air Force Base, Texas, 1961, 16.

Brogden, H. E., and E. K. Taylor. The dollar criterion—applying the cost accounting concept to criterion construction. *Personnel Psychology*, 1950, 3, 133–154.

Campbell, D. T., and D. W. Fiske. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81.

Dunnette, M. D. A note on *the* criterion. *Journal of Applied Psychology*, 1963, *47*, 251–254.

Fleishman, E. A., and B. Fruchter. Factor structure and predictability of successive stages of learning Morse code. *Journal of Applied Psychology*, 1960, *44*, 97–101.

Ghiselli, E. E. Dimensional problems of criteria. *Journal of Applied Psychology*, 1956, *40*, 1–4.

Ghiselli, E. E., and M. Haire. The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology*, 1960, *13*, 225–231.

Guion, R. M. Criterion measurement and personnel judgments. *Personnel Psychology*, 1961, *14*, 141–149.

Guion, R. M. *Personnel testing*. McGraw-Hill, New York, 1965.

Hotelling, H. The most predictable criterion. *Journal of Educational Psychology*, 1935, *26*, 139–142.

Mosier, C. I. Problems and designs of cross-validation. *Educational and Psychological Measurement*, 1951, *11*, 5–11.

Nagle, B. F. Criterion development. *Personnel Psychology*, 1953, *6*, 271–289.

Naylor, J. C., and R. J. Wherry. Feasibility of distinguishing supervisors' policies in evaluation of subordinates using ratings of simulated job incumbents. USAF PRL Technical Documents Report, No. 64–25, October 1964.

Tiffin, J., and E. J. McCormick. *Industrial psychology*, 5th ed. Prentice-Hall, Englewood Cliffs, N.J., 1965.

Toops, H. A. The criterion. *Educational and Psychological Measurement*, 1944, *4*, 271–297.

Wallace, S. R. Criteria for what? *American Psychologist*, 1965, *20*, 411–418.

Weitz, J. Criteria for criteria. *American Psychologist*, 1961, *16*, 228–231.

Wherry, R. J., P. F. Ross, and L. Wolins. Analysis of methods for determining equivalence of criteria. USA PRB *Technical Research Note No. 30*, Behavioral Science Research Laboratory, Office Chief of Research and Development, Department of the Army, October 1954.

7 PERFORMANCE APPRAISAL

The preceding chapter presented the criterion problem from a conceptual point of view. We now turn to the more specific problem of why and how criteria are used in industry. We will examine the respective strengths and weaknesses of various criteria as well as the research that has been helpful in adding to our knowledge in this important area. Criteria are necessary if we are to know how well people perform their jobs. Since performance more or less implies evaluation, a contribution of industrial psychology is in determining objective, scientific methods and techniques conducive to more valid evaluations of performance on the job.

TYPICAL INDUSTRIAL CRITERIA

Many measures are used by the industrial psychologist to appraise job performance. Which of these are used often depends upon the specific set of circumstances. Table 7.1 identifies eleven of the most commonly used criteria. In addition to being aware that criteria selection varies, it is also important to know that performance appraisal can take place at different stages of job experience. There are primarily three different kinds of evaluative circumstances, or situations, in which one might be interested in evaluating a job incumbent. For example, one can evaluate performance during training while the individual is learning the skill or task. In addition, one can evaluate performance while the individual is actually on the job. Further, one can create artificial job environments for purposes of potential performance evaluation. Such artificial work situations are generally referred to as *simulated* job or task settings. They have become exceedingly popular in recent years and are used quite widely by the military services. They are becoming increasingly popular in industry since using simulated tasks is less costly and involved than hiring and waiting for job performance measures. Another advantage of the simulated job situation is that generally one can introduce better controls and thereby reduce criterion contamination problems. Recent examples of simulated tasks at the executive level in industry are the "in-basket" test and computerized management games. Both will be discussed presently.

TABLE 7.1 *Different Performance Measures and Different Evaluative Situations in Industrial Settings*[a]

Performance Measure	*Evaluative Situation*		
	Training Performance	*Actual Job Performance*	*Simulated Job Performance*
Rate of work	time to learn	amount per unit time	amount per unit time
Quality of work	ratings	ratings	ratings
Accidents and breakage	accident rate	accident rate	accident rate
Dollars earned	earnings	earnings	simulated earnings
Job knowledge	ratings or test	ratings or test	
Job tenure	time to learn	length of time	
Absenteeism	number of days	number of days	
Rate of advancement	improvement during training	salary history—promotion history	simulated earnings
Supervisory judgments	ratings	ratings	ratings
Peer judgments	ratings	ratings	ratings
Self-judgments	ratings	ratings	ratings

[a] One example listed in each category.

As an example, consider the general measure of accidents listed in Table 7.1. By accidents we refer to a variety of incidents which result in damage or lost time to product, equipment, or personnel. The use of this measure as a criterion in either training or actual job performance would probably involve a tally of the actual number of such incidents observed over a period of time with, perhaps, a conversion to a dollar base for comparison purposes. In the case of the simulated job settings, however, it is often possible to talk in terms of the number of theoretical accidents rather than real accidents. It is obviously better to have a "theoretical" accident than a real one. An excellent illustration is in flight simulation in which a pilot may have an "accident" which only occurs in theory rather than reality—thus saving large sums of money and lives while, at the same time, accurately assessing a pilot's flight skill. Our purpose, however, is not to argue for simulated criteria situations although they seem to be quite ideally suited for many types of performance evaluation.

SOME EXAMPLES

The measures listed in Table 7.1 fall into several categories. For example, some of them are production measures, others are personal data measures (for example, tenure and absenteeism), while others are judgmental in nature (supervisor and peer ratings and self-ratings).

PRODUCTION DATA

Of these major categories, one might logically expect production data to be the most frequently used criterion data in industrial settings. Interestingly enough, this does not appear to be the case. Schultz and Siegel (1961a) point out that production

records have been used far less extensively than one would expect, and they conclude that such criteria are probably most useful for routine, manual jobs.

Some of the best data on production criteria has been provided by Rothe and his coworkers in a series of studies of output rates among such diverse occupations as butter wrapper (1946a, 1946b), machine operator (1947, 1959, 1961), chocolate dipper (1951), and coil winder (1958). Their data offer ample proof that job performance as measured by production data is subject to substantial time variations. Correlations on the reliability of production data ranged from 0.05 to 0.85, with the magnitude being influenced by the length of the time period between measures. Other reliability information on production data has been provided by Ghiselli and Haire (1960); they found the correlation between weekly time periods to be considerably higher for adjacent weeks than for weeks widely separated.

PERSONAL DATA

Often it is easier to acquire information about a person's job behavior that is not in terms of a production figure. Data concerning the number of times a person is absent from his job, the number of times he is late, the number of grievances he has submitted, and the number of accidents and/or work stoppages he has had are all measures which have been used to define the difference between good and poor employees. In using such measures there is often the implicit assumption that they are related to production performance in some manner—that is, that tardy workers are also those who will produce less. Actually, this is an extremely risky assumption. In fact, as we will point out in a later section, the relationship of personal data measures to production measures is most often very low. Thus, if these measures are used to define job success it is important that they be considered relevant criteria in their own right, rather than simply a substitute for a more direct measure of job productivity.

Examples of the use of personal data as criteria in industry are quite numerous. To illustrate, Baumgartle and Sobol (1959) demonstrated a relationship between job absenteeism and personal history and organizational characteristics. Similarly, Argyle, Gardner, and Cioffi (1958) used both absenteeism and turnover measures in addition to productivity as criteria for the effect of different supervisory methods.

JUDGMENTAL DATA

Although there is no published information, one would probably be fairly safe in deciding that criteria involving judgments is the most frequently used type of criteria. Out of curiosity we selected a random sample of 50 articles from the *Journal of Applied Psychology* for the period 1960–1965 and examined them to see what criteria were used. Table 7.2 is a summary of this survey. The results indicate a tendency for judgmental criteria to be used more often than either personnel or production data. (The numbers for frequency of each type add to more than 50 since some studies used multiple criteria.) Twelve of the fifty articles sampled did not use *any* criteria in the usual or accepted meaning of the term.

Because of the frequent use in industry of judgmental criteria, we shall devote our major attention in this chapter to a description of the various rating techniques and problems as they apply to industry.

TABLE 7.2 *Number of Times Different Criteria Were Involved in a Selected Sample of Research Studies*

Type	*Frequency*	*Percent*[a]
No criteria	12	24
Production criteria	13	26
Personal criteria	9	18
Judgmental criteria	23	46
	57	

[a] Percentages exceed 100 because some studies used multiple criteria.

JUDGMENTAL CRITERIA

One way of determining how good or how poor a particular employee might be is simply to ask somebody. We might ask his supervisor and obtain a *supervisory judgment* of competence, or we might ask his fellow workers and obtain a *peer judgment*. We might even ask the worker himself and obtain what is known as a *self-judgment*. While this procedure may seem almost *too* simple a way to obtain criterion information—particularly the self-judgment method—an experience of one of the authors may serve to suggest that one should not overlook the obvious.

A number of years ago a major university was having a great deal of difficulty with its scholarship students. Every year a certain percentage of the freshman class was given scholarships based upon high school performance, a battery of entrance tests, and teacher recommendations. However, every year about 40 percent of these students would *fail* in the first semester. In desperation, the university began a large-scale research project in an attempt to improve prediction of success. A wide variety of data were collected. Students were given large numbers of tests, ratings were obtained from peers, teachers, and principals, etc. However, when all the data were analyzed, nothing seemed to predict success. That is, nothing but the simple question "How well do you expect to do in your first term in college?" The answers to this gave a high positive correlation with observed first term performance. The moral of this story would seem to be, if you want to know how well a person is going to do, why not ask him?

Now, the discerning reader will already be asking the critical question "How do you know you will get an honest response?" Suppose the high school student thinks that the way in which he responds to this question will play a role in determining whether or not he will receive a scholarship. Will he still admit that he does not think he is going to do very well in his studies? In all likelihood, not—he has too much at stake to be completely straightforward in his response. Instead, he will probably tend to respond in what he feels is the "most desirable" fashion.

Thus we have before us the classic problem of the value of subjective information. If honestly given it is a very rich source of data. How can one solicit such information and be relatively certain that the data obtained are "true" rather than biased? Perhaps one can never be completely certain, although a knowledge of the kinds of errors most apt to occur can be helpful.

ERRORS AND BIAS IN PERFORMANCE APPRAISAL

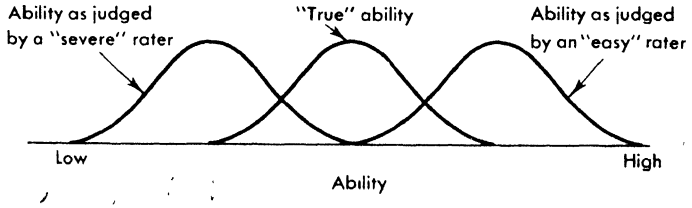
Since human judgment is all too often subject to the influence of bias, prejudice, and other subjective and extraneous influences, the problem of obtaining objective and accurate assessments uncolored by these influences is tremendously difficult.

Guilford (1954) has classified the different kinds of constant errors which are apt to occur in the rating process and suggests certain precautions which can be taken. Constant errors are those which result from some systematic bias on the part of the rater and are usually somewhat easier to cope with than those rater errors which are nonsystematic or random.

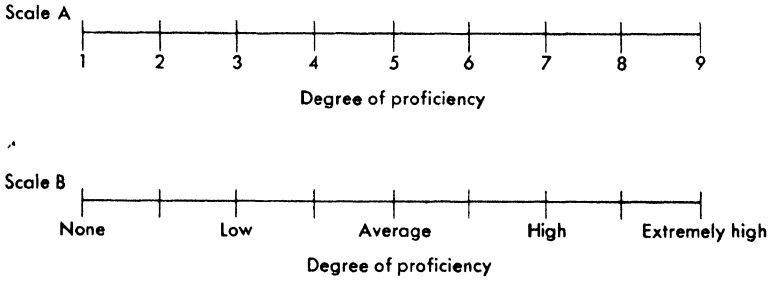
ERRORS OF LENIENCY

Every rater tends to carry around his own built-in set of standards or frame of reference which he uses in making performance appraisal judgments. One of the major problems is to equate different raters for differences in their standards so that their ratings can be compared. Some raters might best be described as being generally "easy" or lenient, while others may be classed as being "hard" or severe in their judgments or ratings. When a rater is overly severe he is said to be making an error of *negative leniency*, while easy raters make the error of *positive leniency*.

The following diagram illustrates the distributions of ratings one might get from two different raters—one who is overly lenient and the other who is overly severe.



There are several ways in which rating systems can be designed to help avoid the leniency error. The first method is to define clearly the categories or "anchor points" used on the scale. For example, if we were requesting supervisors to rank their subordinates on "total proficiency," we could use either of the two scales shown below.



Scale A has no anchor or reference points which the rater can use as a guide; he must establish his own. The risk is that different raters will define the points differently. With Scale B a number of the points have been defined for the rater, with the probability of reducing the problem. These anchor points do not completely solve

the leniency problem, since one rater's "low" might be another rater's "average." However, the verbal scale introduces the possibility of a common frame of reference.

An anchoring procedure which is potentially more effective than the one given in Scale B is know as the "Key-Man" method. It is discussed in detail in a following section and involves using actual individuals as anchor points.

One of the most widely used methods of combating the leniency error is to employ the *forced-distribution* technique. This is a procedure which requires the rater to place a certain proportion of his responses into different categories. Such a distribution is shown in Figure 7.1.

Requiring all raters to adhere to a standard distribution in terms of the number of persons that must be assigned to each category is an effective method of avoiding errors of leniency. It does, however, often cause raters to become slightly disgruntled at being forced into so strict a response pattern. It also takes a great deal more time since care must be taken to have the right number in each category.

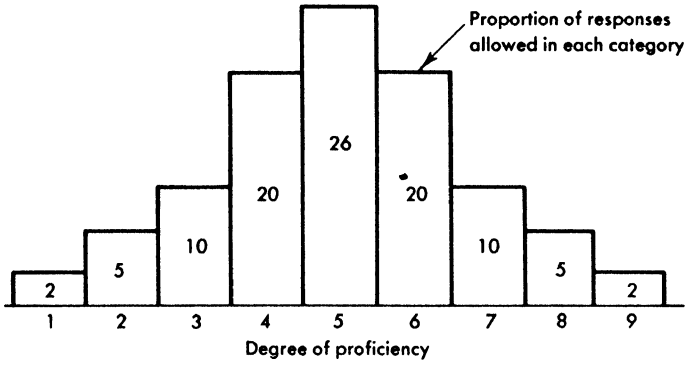


Figure 7.1. Forced distribution.

THE HALO EFFECT

Although this error was mentioned in the preceding chapter, perhaps a few additional comments are necessary here. The halo error is *a tendency to let our assessment of an individual on one trait influence our evaluation of that person on other specific traits.* Therefore, if we felt that Worker X was a top-notch employee in one respect, we might tend to rate him very high on all traits, even though he may be rather mediocre on some.

This is a *very* common type of error and is also one that is very difficult to correct. Symonds (1925) has suggested that it is most likely to occur with the following traits:

1. Traits not easily observed
2. Unfamiliar traits
3. Traits not easily defined
4. Traits involving interpersonal reactions
5. Character traits

LOGICAL RATING ERRORS

This error is quite similar to the halo error. In the halo error the rater unwittingly lets his judgment of certain traits be influenced by his assessment of the individual on other traits. A logical error involves a rater who tends to give a person a high score

on one specific trait simply because he feels the individual possesses a lot of a second specific trait and *he feels the two traits are logically related*. When a rater tends to overestimate the true relationship between traits, he probably commits this rating error.

CONTRAST AND SIMILARITY ERRORS

The contrast error is a rather interesting type of bias. It refers to a general tendency on the part of a rater to judge others in a manner opposite from the way in which he perceives himself. If he perceives himself as being very honest, for example, his tendency would be to rate others slightly lower than normal on the "honesty" dimension. The opposite of a contrast error, which might be called a similarity error, is for the rater to rate other people in the same way he perceives himself. An illustration of this would be a rater, feeling himself very honest, who attributes high honesty to all other individuals as well. In other words, raters should be trained to rate others and not themselves.

CENTRAL TENDENCY ERRORS

Some judges or raters are often very reluctant to make extreme judgments about other individuals. This reluctance results in their tending not to use the extreme scale scores on the rating instrument. This, in turn, results in a substantial change of shape in the distribution of scores for that rater, as can be seen in Figure 7.2. Notice that

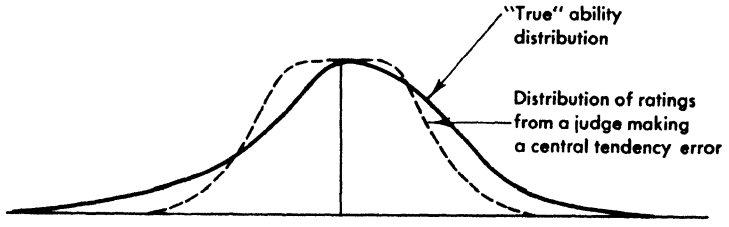


Figure 7.2. The effect of a central tendency error upon the shape of a distribution of ratings.

the dispersion (variability) of the judgments is much less for the rater making a central tendency error. This kind of error thus results in a restriction of range of the criterion scores—an artifact which can have an effect on subsequent validity coefficients (see Chapter 2). One of the better ways to avoid this error is by using the forced-distribution system previously discussed.

PROXIMITY ERRORS

The last error we shall consider usually comes about from the way in which the various items have been placed or ordered on the rating form. Sometimes referred to as an "order-effect," this error illustrates the influence that surrounding items have on the rating one gives a person on a particular item. If the preceding item was a trait on which the individual was given a very favorable rating, the rater may tend to let the favorable response "set" carry over to the next item on the list. The result will be a rating on the next item slightly higher than might otherwise be given. There is always the possibility of the reverse reaction occurring. If the preceding items have

been on traits in which the worker was generally rated high, when the rater gets to a trait in which the worker truly deserves only a "moderate" ranking, he may give a rather unfavorable score simply because of the contrast effect of the preceding items.

The most common procedure for minimizing proximity error is to have several different forms of the rating scale, with the items in a different order on each form. Unfortunately, when the number of items becomes large, systematic control over item order becomes extremely complex and difficult in most practical situations. Common practice in avoiding this error is usually limited to constructing only two forms, with the items assigned to each form using a table of random numbers.

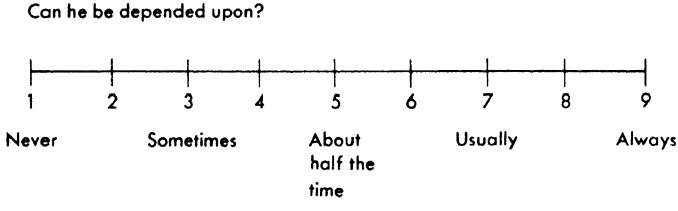
JUDGMENTAL METHODS OF APPRAISING PERFORMANCE

The industrial psychologist has a choice of methods to establish a procedure for appraising job performance involving judgments on the part of individuals. These methods may be grouped into classes depending upon the kind of comparison required of the individual doing the appraising. The three most commonly used are *rating systems, ranking systems,* and *paired comparisons.*

RATING SYSTEMS

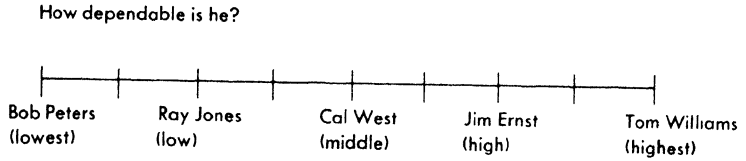
The mechanics of any rating scale system are quite simple. The task of the judge is to make a judgment concerning the degree to which the individual possesses, or is described by, a particular characteristic. While it may take many different forms, its one distinguishing characteristic is that the judge may give two individuals the same score, a feature not found in the other two methods.

NUMERICAL RATING SCALES This form is probably the most popular and common type of rating scale. Such scales are often called "graphic scales." An example is given below:



The task of the judge is to give each worker a score from 1 to 9 which best describes his "degree of dependability." The criterion score for a worker is usually the average score of all judges. In the situation where a number of different questions comprise the criterion (for example, degree of dependability, general level of work knowledge, relations with other workers, etc.) the problem becomes more complicated. Often the total criterion used is simply the total score of all questions. However, when multiple questions are used to evaluate workers, one immediately must face the composite versus multiple criterion problem discussed in the preceding chapter. An approach becoming more popular in such situations is to factor analyze the questions to determine underlying independent criterion dimensions and then to assign "factor scores" on each of these dimensions which are then used as the criteria.

STANDARD SCALES Another type of rating scale involves the use of a set of standards or examples for comparison purposes. These are used in place of, or along with, the verbal anchor points found with the graphic scale. A standard scale designed to evaluate dependability might look like this:



The rater can then compare the man he is rating with men on the job who are designated as scale referent points. This method is also sometimes referred to as a *man-to-man* comparison scale.

The major difficulty with standard scales exists in finding appropriate standards. This usually requires some careful prescaling of a group of present employees. There is also the very touchy problem of using the names of these men without any embarrassment. Thus, one must be very careful that the standard scale is not allowed to be made public. Another difficulty is that the intervals between the men on the scale are often not of equal magnitude.

The advantage of the method is quite clear. As long as a group of judges can agree on the qualifications of a small group of workers who are well known to all judges, a ready-made comparison scale is provided which assures that all the judges will have a common referent in rating the rest of the men.

CUMULATED POINTS SCALE (CHECK LIST) Many rating scales require the judge to check or indicate which of a number of statements, adjectives, or attributes are descriptive of the person being rated. For example, the rating form shown in Figure 7.3 is representative of an adjective check list.

Check only those adjectives which describe the person being rated.

_____friendly	_____tenacious	_____selfish
_____eager	_____willing	_____radical
_____withdrawn	_____cruel	_____greedy
_____aggressive	_____stingy	_____stubborn
_____spoiled	_____defiant	_____helpful
_____happy	_____conservative	_____quiet

Figure 7.3. A typical rating descriptive check list.

Each person is assigned a total score based upon the particular items checked. In the example shown, a person might receive a +1 for every favorable adjective and a −1 for every unfavorable adjective. In more elaborate check lists, the different items may be assigned differential weights. For example, Uhrbrock (1961) has scaled 2000 statements concerning worker behavior. Each statement has been given a mean scale value of its perceived favorableness in describing a worker. Statements can then be selected for use on a check list which in turn is used by supervisors to rate their employees. The method of scoring could be based upon either the total (algebraic sum) score for those items checked or, perhaps, the median (middle) scale value of all the items checked. Some examples of Uhrbrock's statements, along with their scale values, are shown in Table 7.3.

TABLE 7.3 *Some Examples of Worker Descriptions and Their Scale Values*

Item	*Mean*	*S.D.*
1. Is dynamic leader who stimulates enthusiasm	109.38	2.42
2. Does superior work	108.75	3.31
3. Is exceptionally reliable	108.75	3.31
4. Is extremely accurate and efficient	108.75	3.31
5. Has remarkable mastery of all phases of work	108.75	3.31
6. Supervises with rare skill	108.13	3.90
7. Is an expert on his job	107.50	4.33
8. Earns highest respect from others	107.50	4.33
9. Shows top potential for advancement	107.50	4.33
10. Does exceptionally good work	106.88	5.83
11. Performs tasks exceptionally well	106.25	5.99
12. Consistently has exceptional output	106.25	5.99
13. Has exceptional skill in motivating others	105.63	6.09
14. Quality of work is exceptional	105.63	7.88
15. Quality of work is superior	105.63	6.09
16. Is an outstanding worker	105.63	6.09
17. Is outstandingly accurate	105.63	6.09
18. Is highly ingenious	105.63	6.09
19. Has high integrity	105.00	6.12
20. Stimulates his employees	105.00	6.12
21. Has unusually high output	105.00	6.12
22. Inspires people to do their best work	105.00	6.12
23. Works well under pressure	105.00	7.07
24. Is highly efficient	105.00	7.91
25. Is unusually creative	105.00	7.91
26. Maintains stability and poise during periods of sustained pressure	105.00	7.91
27. Will make exceptional progress with company	105.00	8.66
28. Shows exceptionally dependable performance	104.38	6.09
29. Has superior ability to foresee consequences of action	104.38	7.04
30. Works harmoniously with others	104.38	7.88
31. Can be relied upon to find a solution to a new problem	104.38	7.88
32. Has excellent judgment	104.38	8.64
33. Is clearly superior mentally	104.38	8.64
34. Quality of work is at unusually high level	103.75	4.84
35. Is a real find	103.75	5.99
36. Has unusual sense of responsibility	103.75	5.99
37. Is extremely successful in working with others	103.75	5.99
38. Knows job thoroughly	103.75	6.96
39. May be counted on in a crisis	103.75	6.96
40. Shows unusual skill in control of quality	103.75	6.96
41. Makes sound decisions promptly on difficult problems	103.75	6.96
42. Has excellent knowledge of all phases of job	103.75	8.57
43. Operation is of highest quality	103.75	10.53
44. Has immense capacity for responsibility	103.75	10.53
45. Is excellent prospect for advancement to position of greater responsibility	103.75	10.53

SOURCE: R. S. Uhrbrock. 2000 scaled items. *Personnel Psychology*, 1961, *14*, 375–420.

CRITICAL INCIDENT CHECK LIST The critical-incident technique is a procedure developed by Flanagan (1954). As normally used, it can probably best be described as a check list rating procedure. However, it is also sufficiently different in its development that it deserves separate mention. The method involves three distinct steps:

1. Collecting critical incidents
2. Scaling the incidents
3. Constructing the check list scale

Step one is what makes this method unique. The term *critical incident* refers to all those on-the-job behaviors of people which the supervisor would consider to be "noteworthy." Noteworthy, in turn, can refer to instances of either outstandingly "good" or "bad" on-the-job behavior. Each supervisor, in other words, is asked to provide as many instances of outstanding actual job behavior as he can recall. These job incidents are then presented (step two) to a group of experts who assign scale values as to the degree of desirability of this incident.

In step three a check list is constructed which includes only those statements or incidents felt to be relevant to defining "good" and "poor" workers. These check lists can then be used by a supervisor to rate his workers. In rating a worker, he simply places a check mark alongside each incident that he has observed that worker "commit" or exhibit. The score for a worker is usually taken as the mean or median scale value of all incidents checked for him (total score can also be used).

The critical incident check list has the great advantage that it is based upon actual job behaviors or incidents. A supervisor is not placed in the position of being forced to judge traits or to rank his workers, etc. He only has to respond in terms of whether or not he has or has not observed a particular incident on the part of each worker. Kirchner and Dunnette (1957) have reported favorable results in a study following somewhat the procedure outlined above.

Before leaving the critical-incident method, it should be pointed out that it has more generality and application than simply being a source for performance appraisal check list items. Job-based behavior incidents have been used very successfully in research in the areas of leadership (Fleishman, 1955), organizational values (Shartle, Brumback, and Rizzo, 1964), and creativity (Buel, 1960).

RANKING SYSTEMS

A characteristic of a rating system is that it permits two or more individuals to have the same rating or scale value. A ranking system by its nature discourages this. Simple ranking requires that the judge order the individuals from highest to lowest. Thus, a particular group of supervisors might rank ten employees as shown below:

	Supervisor A	*Supervisor B*	*Supervisor C*
1.	Axel	Cerny	Axel
2.	Bond	Axel	Bond
3.	Cerny	Dixon	Dixon
4.	Dixon	Bond	Cerny
5.	Engle	Engle	Frye
6.	Frye	Green	Engle
7.	Green	Frye	Houghton
8.	Houghton	Jones	Green
9.	Inman	Inman	Inman
10.	Jones	Houghton	Jones

Each man can then be assigned an average rank.

	A	B	C	*Average Rank*
Axel	1	2	1	1.3
Bond	2	4	2	2.6
Cerny	3	1	4	2.6
Dixon	4	3	3	3.3
Engle	5	5	6	5.3
Frye	6	7	5	6.0
Green	7	6	8	7.0
Houghton	8	10	7	8.3
Inman	9	9	9	9.0
Jones	10	8	10	9.3

Ranking systems have the inherent advantage in that they are extremely simple to explain and are usually easily accepted by the persons assigned as judges. The procedure "makes sense" to them. It also has the advantage of permitting a rater to rank fairly large numbers of individuals without great difficulty. A rule of thumb recognizes that satisfactory and reliable results with a ranking system can be achieved with *N*s as large as 50 to 60.

Guilford (1954) has pointed out another advantage of ranking. Since the judge is forced to make *man-to-man* comparisons rather than *absolute* comparisons, he must compare individuals when deciding to place one above the other.

As previously mentioned, the method does force the judge to discriminate between the individuals being ranked *regardless of whether he feels they are truly different!* Suppose the ten men given in our earlier example were rather equal in their ability. The judges would then probably have an exceedingly difficult time trying to rank order the men, and the rankings of a judge would be highly deceptive as to the relative merits of the ten men in question. One way to check on this possibility, of course, is to use several judges. If the judges *agree* in terms of their rankings, the possibility of there being no "real" or true differences among the men can be discounted. In the earlier example, the judges appear to be reasonably consistent in judging who is a good and who is a poor worker. Thus, their choices must be predicated on some real rather than random reasons for ordering the men in question. This is one excellent reason for using multiple judges whenever possible and for computing the agreement among judges with some index such as a correlation coefficient.

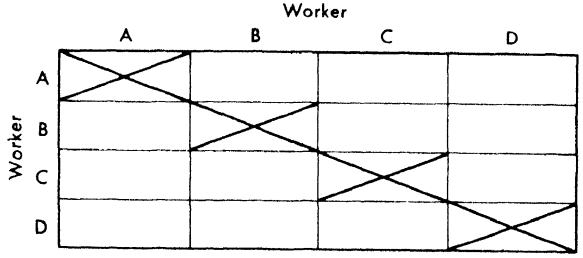
PAIRED COMPARISONS

Although the rank order systems implicitly require that a man be compared to others on the list during the ranking process, this is not systematically built into the method. A procedure which does systematically force the rater to compare each man with every other man is known as the method of paired comparison. For example, consider the situation where four employees are being evaluated by a supervisor. If we form all possible pairs of men, we shall have $N(N-1)/2$ such pairs formed; in our case, $4(3/2) = 6$ pairs.

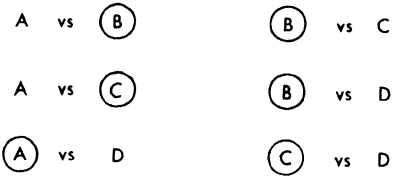
Workers	*Possible Pairs*	
A (Agnew)	A with B	B with C
B (Baker)	A with C	B with D
C (Cooper)	A with D	C with D
D (Davis)		

The usual procedure, then, is to present each of the pairs (one at a time) to the supervisor. The supervisor indicates (usually with a check mark) which of the two men he feels to be best. A paired comparison matrix can then be constructed like that shown below, in which the following rule is used: *Whenever a supervisor prefers a column to a row, place a mark in that cell.*

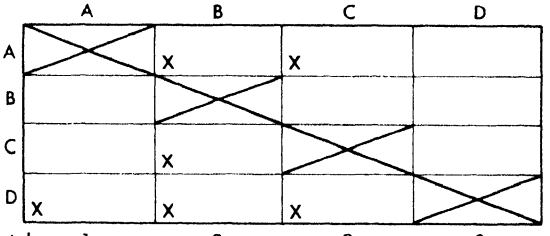
Sample Paired Comparisons Matrix



Suppose supervisor X makes the following judgments about the six pairs of workers:



If we place tally marks in our matrix using the above rule we have the following:



	A	B	C	D
Total	1	3	2	0
Proportion of time preferred over other persons	0.33	1.00	0.67	0.00

Thus, this supervisor rates B as the best worker, followed in order by C, A, and D. In actual practice, the judgments of a number of judges are included in the matrix to provide a more stable estimate of the order of merit of the men being evaluated. In addition, it is common practice to convert the proportions within each cell to standard scores (Z scores) and then find an average Z score for each column.[1] This average Z score is then the *scale value* for the person who is represented by that column in the matrix. The advantage of this latter procedure is that these scale values may be taken as representing not only the rank order of worker ability, but also as an indication of the degree to which two workers differ in ability.

[1] A Z score is a "normalized score" in that it has been adjusted for differences in rater means and rater variances.

A simple illustration based upon the responses of ten hypothetical supervisors to the four men rated above may be helpful. First let us consider the following tally matrix

	A	B	C	D
A		XXXX XXXX X	XXXX XXXX	XXX
B	X		XXX	X
C	XX	XXXXX XX		X
D	XXX XXXX	XXXX XXXXX	XXXX XXXXX	

Tally matrix

Now let us convert the tally matrix to a proportion matrix, where we define a proportion as $^{n}/_{N}$, where n equals the number of tally marks in a cell and N equals the number of judges or number of possible tally marks. It is customary to enter proportions of 0.50 in all the diagonal cells, even though we do not actually ever present a person paired with himself to the judges.

	A	B	C	D
A	(0.50)	0.90	0.80	0.30
B	0.10	(0.50)	0.30	0.10
C	0.20	0.70	(0.50)	0.10
D	0.70	0.90	0.90	(0.50)

Proportion matrix

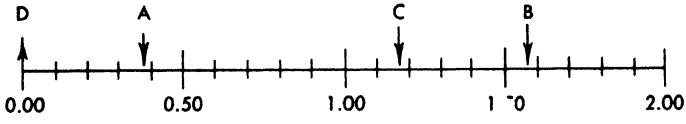
The next step is to convert the proportions to Z scores. (A table for doing this can be found in any elementary book on statistics.)

	A	B	C	D
A	0.00	1.29	0.84	−0.53
B	−1.29	0.00	−0.53	−1.29
C	−0.84	0.53	0.00	−1.29
D	0.53	1.29	1.29	0.00
Total	−1.60	3.11	1.60	−3.11
Mean	−0.40	0.78	0.40	−0.78
Mean + 0.78	0.38	1.56	1.18	0.00

Z score matrix

= Individual scale value

If we add a constant of 0.78 to all scale values to remove negative signs, the individuals can then be represented along a continuum of merit:



Notice that the distance between C and A is considerably larger than the distance between D and A or B and C. If the assumptions of this method have been met, these relative distances may be taken as being indicative of the relative differences in merit among the individuals being rated—an appealing characteristic of this performance appraisal method. For a more detailed coverage of paired comparison, refer to Edwards (1957) or Guilford (1954).

One of the drawbacks of the method of paired comparisons is that it is not particularly useful when the number of individuals to be rated becomes large. Earlier, it was pointed out that a supervisor could probably successfully rank order 50 or 60 workers. If we tried to use the paired comparisons procedure, we would have

$$50\,(49)/2 = 1225 \text{ pairs}$$

Even if a supervisor could work at the rate of six pairs each minute, it would require more than three hours to accomplish the rating. The larger the number of persons to be rated, the more astronomical becomes the time requirement.

FORCED-CHOICE SYSTEMS

One of the most popular methods for obtaining appraisals of performance is through the technique known as "forced choice." This rating method, which was an outgrowth of the usual problems encountered in the development of personality measurement instruments, was quickly adapted for use in performance appraisal. Historically, the method seems to have originated with Horst during the early 1940s, and was also used by Wherry at about the same time.

The method has been used in a wide variety of situations and for a wide variety of purposes. In an excellent critique of the forced-choice method, Zavala (1965, p. 117) summarizes as follows:

> The FC technique has been used to rate professional personnel such as Air Force officers and servicemen (Gough, 1958; Wherry, 1959), highway patrolmen and police officers (Peres, 1959; Stander, 1960), engineers (Lepkowski, 1963), teachers (Leftwich, 1962; Tolle and Murray, 1958), and physicians (Newman and Howell, 1961; Newman, Howell, and Harris, 1957). The FC concept has also been used to develop tests for rating or measuring interests (Bendig, 1958; Peel, 1961), leadership and supervision (Izard and Rosenberg, 1958; Wollack, 1959), and employee attitudes (Miller and Gekoski, 1959). Gordon (1951), Denton (1954), and Hatch (1962) have put FC techniques to use in personality measurement and to obtain measures of empathy. The FC approach has also been used in problems related to signal detectability and auditory thresholds (Lukaszewski and Elliott, 1962; Swets, 1959). Blackwell (1952) used the FC method for the psychophysical measurement of sensory functions. Webster (1959) developed a figure-preference test using the FC technique.

CONSTRUCTION OF FORCED-CHOICE RATING SCALE Guilford (1954, p. 275) has outlined very clearly the appropriate steps to be followed in developing a forced-choice instrument—so clearly that they are worth reiterating here. He suggests the following eight steps:

> 1. Descriptions are obtained concerning persons who are recognized as being at the highest and lowest extremes of the performance continuum for the particular group to be rated.

2. Descriptions are analyzed into simple behavior qualities, stated in very short sentences or phrases or by trait names, which may be called elements are used to construct items.

3. Two values are determined empirically for each element: a discrimination value and a preference value. The discrimination value is an index of validity, and the preference value is an index of the degree to which the quality is valued by people like the raters who will use the instrument.

4. In forming an item, elements are paired. Two statements or terms with about the same high preference value are paired, one of which is valid and the other not. Both should have "face validity" for the rater, i.e., the rater should think that they are both favorable for superior performance in the group rated. Two statements or terms with about equally low preference value are also paired, one being valid and the other not.

5. Two pairs of statements, one pair with high preference value and one with low preference value, are combined in a tetrad to form an item. The reason for this kind of combination is that although the average rater will not object to picking one of two favorable descriptions for a person whom he knows, he sometimes balks at picking one of two unfavorable descriptions. Sometimes a fifth, neutral, description is added to form a pentad, but this is less common. An example of a tetrad follows:

careless
serious-minded
energetic
snobbish

The traits "serious-minded" and "energetic" would have been found to have equal preference value because they were applied about equally often as favorable traits in describing the type of personnel for whom the scale was developed. The trait "serious-minded," however, was found to be valid, since it was applied to the high criterion group significantly more often than to the low criterion group. The traits "careless" and "snobbish" were found equally unpopular, but "careless" discriminates the low from the high criterion group.

6. The instruction to the rater is prepared. The rater is to react to each tetrad as an item, saying which one of the four best fits the ratee and which one of the four is least appropriate.

7. An experimental form of the instrument is tried out in a sample for which there is an outside criterion, for the purpose of validating the responses when the descriptions are set up in this form. Discriminating responses are determined, and, if desired, differential weights are assigned.

8. A scoring key is devised, based on the results in Step 7. Ordinarily, a valid favorable trait marked as most descriptive of the ratee receives a positive weight, also a valid, unfavorable trait judged as least descriptive.

RATIONALE Of course, from the above description the logic of the name "forced" choice" is immediately apparent, in that the rater is continually being forced to select from among equally favorable (or unfavorable) alternatives. This supposedly prevents the rater from deliberately checking only the most favorable traits when rating those workers to whom he might wish to show some degree of favoritism. If he chooses his responses strictly on the basis of how favorable they appear to be, then he will in theory have no reason for selecting one trait over the other in any pair since they have been equated on this basis. Thus, his choice would be a random one and, by chance, he will end up selecting the discriminating trait half of the time.

Since each tetrad has one favorable pair (with a positive discriminator in it) and one unfavorable pair (with a negative discriminator in it), a random selection in each pair would mean that the rater would tend to choose a favorable discriminator about half the time and an unfavorable discriminator about half the time. Thus, a random or chance score would be zero (assuming we give a +1 for every positive discriminator and a −1 for every negative discriminator checked). To the extent that the rater is truly attempting to give an accurate assessment of the worker, he will tend to select traits which discriminate in the positive pair and not to select traits which discriminate in the negative pair (assuming he is rating a good worker). If he is rating a poor worker, his pattern would be reversed in that he would tend to select the discriminating trait more often than chance among the negative pair and select the nondiscriminating trait more often than chance in the positive pair. Good workers should, thus, end up with high positive scores and poor workers with high negative scores.

FORCED-CHOICE INDICES The key to a successful forced-choice scale lies in the several types of indices which are obtained for each trait prior to constructing the final version of the rating scale. The different indices are as follows:

A. Discriminative indices
B. Equating indices
   1. Preference index
   2. Favorability index
   3. Importance index
   4 Desirability index

Discriminative Indices. The discriminative index is simply a measure of the degree to which a trait is found to differentiate between good and poor workers. All the indices normally employed in item analysis are potentially appropriate for use as measures of trait discriminality since one is exclusively concerned with how valid each trait is in predicting how good an employee might be. For a discussion of item analysis methods, consult any basic text in psychological testing.

Equating Indices The rationale of the forced-choice system requires that traits be paired in such a manner that they are equally "attractive" to the rater. In other words, one trait must not possess more of any characteristic which would lead to its being picked over the other trait by a rater having a desire to bias his responses. Selection of traits in any pair is to be based solely upon their discriminating power rather than in terms of any population stereotype about the traits themselves.

We are immediately led to the problem of attempting to determine what kinds of trait characteristics are apt to influence the response of a rater wishing to bias his answers. Four such characteristics might be considered as potentially important.

1. *Preference for a trait.* This can best be defined as the general tendency for raters to use the trait with all people, regardless of whether they are good or poor workers. Sisson (1948) defined this characteristic as "the extent to which people, in general, tend to use it in describing other people." In a sense, trait preference is a measure of the commonness of a trait in the descriptive vocabulary of raters. Equating pairs of traits in terms of their general preference would certainly seem a logical step if one were trying to remove extraneous item characteristics which were likely to influence the choice of a rater.

2. *Favorability of a trait.* Traits clearly differ in terms of their apparent favorableness when used as a descriptor of people. As has been mentioned, the forced-

choice format usually pairs two favorable traits with two unfavorable traits. It is also important that the two favorable traits appear equally favorable and the two unfavorable traits equally unfavorable, or the rater may be tempted to select the most favorable in each pair in an attempt to make the ratee look as good as possible.

Trait favorability is probably the most commonly used index for equating forced-choice pairs. It is an index that is relatively easy to obtain by having a number of judges rate each trait in terms of its favorableness to the person it is going to be used to describe and using the mean scale value as the favorability index.

3. *Importance index.* This index has been mentioned by Zavala (1965) and is a measure of "the importance of the statement as a qualification for the position in question" (Zavala, 1965, p. 118). This index recognizes that a person with a bias tendency is more apt to use a specific situation as a frame of reference than a more general reference. That is, a rater who wishes to make a ratee look good may be more apt to pick that specific trait which he feels is most important for the job in question than he is to pick that trait which is the most favorable general descriptor. The rater's choice of which trait is the best to use is, therefore, likely to be job specific. In order to successfully equate traits for their importance, it is necessary that importance scale values be obtained (usually in the same general fashion that one obtains favorability scale values) for each different rating situation —a task that may, at times, be difficult to accomplish.

4. *Desirability index.* In Chapter 4 the notion of social desirability was discussed as an influence in determining the responses of individuals to personality inventories. The effect of this trait characteristic is probably equally important in rating situations. Differentiating between favorability and social desirability is not always an easy thing to do. Certainly, traits which are judged favorable are also apt to be considered socially desirable. The distinction might be illustrated by considering the favorability of a trait like intelligence which might be fairly high, while its social desirability might tend to be somewhat lower due to a general cultural reaction against "eggheads," etc.

FRAMES OF REFERENCE Not only are there a variety of different indices which can be used to equate pairs of traits in a forced-choice format, but there are also several instructional sets or frames of reference which can be given to the judges when obtaining scale values for equating purposes. As a result, different instructions are apt to be quite critical. For example, if one is interested in getting importance indices, he might give the judges either of the following sets of instructions: (1) Gives each trait a score based upon how important *you* feel that trait is for successful job performance, or (2) give each trait a score based upon how important you think *the rater eventually using the scale* will feel the trait is for successful job performance.

In the first instance, one obtains scale values based upon the judges' own frame of reference. In the second instance, the judges are asked to place themselves in the "shoes" of the person who will eventually be asked to use the performance appraisal instrument and to make judgments as if they were that person. The importance scale values obtained under these two sets of conditions might turn out to be quite different.

THE "IDEAL-MAN" STRATEGY A related problem to the frame of reference question is the rater strategy referred to as the "ideal-man" strategy. One very typical way of biasing responses to a forced-choice rating scale is for the rater to select his best

worker and mentally substitute his characteristics for those of the worker whom he is actually rating and whom he wishes to have score well.

If he has no worker good enough to be used as a guide in rating, the rater may draw a mental image of an ideal worker and use this as his frame of reference in making his ratings. This type of deliberate biasing is *extremely* difficult to cope with since, in a sense, the rater does not pay any attention to the kinds of trait characteristics which the forced-choice scale tends to protect against. That is, the rater tends to do an "honest" job of rating, except the "man" he is rating when he completes the scale is quite a different person than the one he is supposed to be rating. When a rater is clever enough to adopt this rating strategy as a method of raising the performance scores of his friends, etc., there is little that can be done to avoid the resulting bias, even with forced choice.

RESEARCH ON FORCED-CHOICE METHOD

The method of forced choice has been the subject of a great deal of research since its introduction in the 1940s. As with many newer methods, it has turned out to be somewhat less of a cure-all than many had hoped, but it still appears to be one of the most effective ways of reducing rater bias available to the industrial psychologist. A brief overview of some of the findings dealing with various aspects of the forced-choice method will give an idea of the general status of the method.

FORCED-CHOICE ITEM FORMAT

In what has come to be considered one of the classic studies on the forced-choice method, Highland and Berkshire (1951) compared six different types of item formats in terms of (1) their odd-even reliability, (2) their susceptibility to bias when raters were instructed to assure a high score, (3) their validity against a criterion composed of instructors' rank orders of the students, and (4) their general popularity as determined by the raters.

The six different kinds of item formats used were:

1. Two statements per item, either both favorable or both unfavorable. Rater was asked to select most descriptive of the pair of statements.
2. Three statements per item, either all favorable or all unfavorable. Rater was asked to select the most and the least descriptive statement.
3. Four statements per item, all favorable. Rater was asked to select the two most descriptive statements.
4. Four statements per item, all favorable. Rater was asked to select the most descriptive and the least descriptive.
5. Four statements per item, two favorable and two unfavorable. Rater was asked to select both the most and the least descriptive statements.
6. Five statements per item, two favorable, two unfavorable, and one neutral. Rater was asked to select both the most and the least descriptive statements.

The results of the study led Highland and Berkshire to the following conclusions:

Reliability: All formats led to high reliability coefficients, although formats 5 and 6 might be considered as giving the best showing on this criterion.

Validity: Format 4 was generally found to be the most valid, with format 3 running a strong second. This indicates that the use of only favorable alternatives seem to affect validity.

Rater preference: The preference of the raters for the six formats as (from most to least preferred) 3, 1, 6, 5, 4, and 2.

Susceptibility to bias: The forms differed considerably in terms of the degree to which they were resistant to deliberate attempts at biasing. The order of bias resistance (from most to least) was 3, 2, 1, 4, 5, and 6.

Highland and Berkshire suggest that format 3 is the best of the six studied when the four criteria listed above are considered.

COMPARATIVE VALIDITY OF FORCED CHOICE

The Highland and Berkshire study examined different formats of forced choice among themselves. An equally important question concerns the validity and reliability of the method when compared to other performance appraisal procedures. In reviewing the studies which have purported to examine the advantage of forced choice in terms of validity, Zavala (1965) makes the following points:

1. Too many studies of the validity of forced choice have tended to use other forms of rating as a criterion. This is probably more of a reliability measure than a validity measure.
2. Most comparative studies seem to show a slight superiority for forced-choice over conventional rating methods.
3. The longer the rating scales, the more apt the forced-choice method is to be superior.
4. The validity of a forced-choice scale is also apt to be a function of the kind of equating index used in constructing the items.

SUSCEPTIBILITY TO BIAS

Although the question of validity is probably the most important one which can be used to evaluate the forced-choice method, the question of the degree to which the method is subject to bias or deliberate distortion is also critical. Indeed, the method was specifically designed to reduce response bias, since response bias tends to reduce validity.

The evidence on the effectiveness of the forced-choice procedure as a bias reducer is somewhat equivocal. There appears to be little doubt that faking remains possible under a forced-choice format, as indicated by Sisson (1948), Howe (1960), and Howe and Silverstein (1960). However, there does seem to be substantial evidence that the degree of fakability is reduced by forced-choice items. Karr (1959), Taylor and Wherry (1951), and Izard and Rosenberg (1958) offer three examples of research studies which indicate a resistance to faking by forced choice that exceeds that of other types of scales.

Waters (1965) has recently suggested that one of the major difficulties with research on the fakability of forced choice has been that the "set" to fake the test which has been established has usually differed from the "set" under which the equating indices were originally obtained. Thus, the indices are never really completely appropriate for the actual appraisal situation. He suggests that there are three different response sets under which equating indices may be obtained (regardless of which index is being used).

1. *Honest appraisal:* Answering as the respondent truly believes the statement applies to the person being evaluated
2. *Social acceptability:* Answering so as to appear acceptable to oneself and to others in general
3. *Success appearance:* Answering so as to make the person appear to possess the qualities necessary or desirable in relation to a particular job or activity

It is further suggested by Waters that fakability studies may be classified into three general classes, depending upon how the response set and the group composition are varied. To quote from Waters (1965, p. 189), we have

> 1. *Fakability studies:* Sets and groups at least similar in the situations in which the attractiveness indices are obtained and the scale is administered.
> 2. *Generalization studies:* Either sets or groups, but not both, changed from the situation in which the attractiveness indices are obtained to the situation in which the scale is administered.
> 3. *Extension studies:* Both sets and groups changed from the situation in which the attractiveness indices are obtained to the situation in which the scale is administered.

Fakability studies are not to be confused with the latter types of studies. They are the only type which directly provide a test of how adequate the equating indices are in doing their job. The latter two simply evaluate the generality of the index to other situations. Unfortunately, according to Waters, most fakability studies to date have been either generalization or extension studies; there is a strong need for some research which directly attacks the fakability question.

NORMAN STUDY

One of the more interesting studies on the problem of forced-choice faking and detection is the study by Norman (1963). He was concerned with the dynamics of item popularity and item discrimination indices under normal conditions and under faking conditions. In addition, he was interested in determining to what extent faking, if it did occur, could be controlled and/or detected.

The major findings of the research relative to the behavior of the indices of popularity and discrimination are given below:

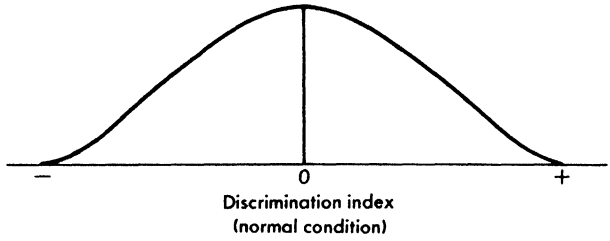
1. Discrimination indices under normal and fake conditions were correlated approximately zero.
2. Popularity indices under normal and fake conditions were correlated low (0.24 and 0.23).
3. Reliability of discrimination indices under fake conditions was zero.
4. Reliability of discrimination indices under normal conditions was moderately high.
5. Reliability of popularity indices was nearly perfect under both normal and fake conditions (0.97 and 0.98).

The last of these results (finding number 5) implies that the *difference* in popularity between normal and fake conditions must also be very reliable (that is, finding number 2, which says there are substantial differences, can be modified to imply that such differences are reliable and consistent). However, this was not evaluated directly in the study.

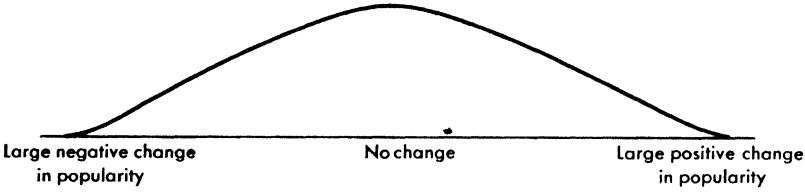
Norman then suggests a procedure for developing a forced-choice instrument which will have the same mean under fake conditions as it will under normal

conditions, have a smaller variance under fake conditions, and have a highly sensitive detection scale for identifying fakers. To accomplish this, he outlines the following steps:

1. Select items with high discrimination indices under normal conditions (select more than will eventually be required).



2. Rank order the selected items in terms of the size of their differences in popularity under the normal and the fake conditions (see number 2 in above list).



3. Select items from both sides of zero (as many as desired), making certain that the algebraic sum of popularity differences is equal to zero when completed.

Step 3 will ensure (since we know popularity *differences* are reliable) that the mean test score of people under the normal conditions will equal that under the fake conditions. This occurs since the mean of the distribution of test scores is equal to the sum of the item popularities. Thus, *mean* scores under normal and fake conditions are equated. Tests constructed with this procedure will also tend to have smaller variances of their test score distributions when taken under a set to fake.[2] This is an outcome which tends to *reduce* the possibility that people who are apt to fake will get high enough scores to be accepted.

To illustrate, consider the diagram shown below (Figure 7.4) in which we have the distribution of test scores taken under normal conditions. (7.4a), the distribution of test scores taken under a set to fake (7.4b), and a composite distribution made up of fakers and nonfakers (7.4c).

Examination of Figure 7.4c provides an indication of the controlling effect this method exerts on deliberate fakers. In the usual performance evaluation setting, we are interested in rewarding those people who score high on the forced-choice form. Since the variance of the fake distribution is reduced, one can select the top scoring people (those to the right of the cutoff line in 7.4c) without the risk of getting too many fakers included among the selected or rewarded group.

Figure 7.5 shows five different empirical demonstrations of this restriction in variance obtained in the Norman study.

[2] The reason for the smaller variance under the fake condition is that, in general, the item popularities become more extreme since item popularity is the same as item difficulty, and since item difficulty determines item variance, the item variances tend to be less under fake conditions.

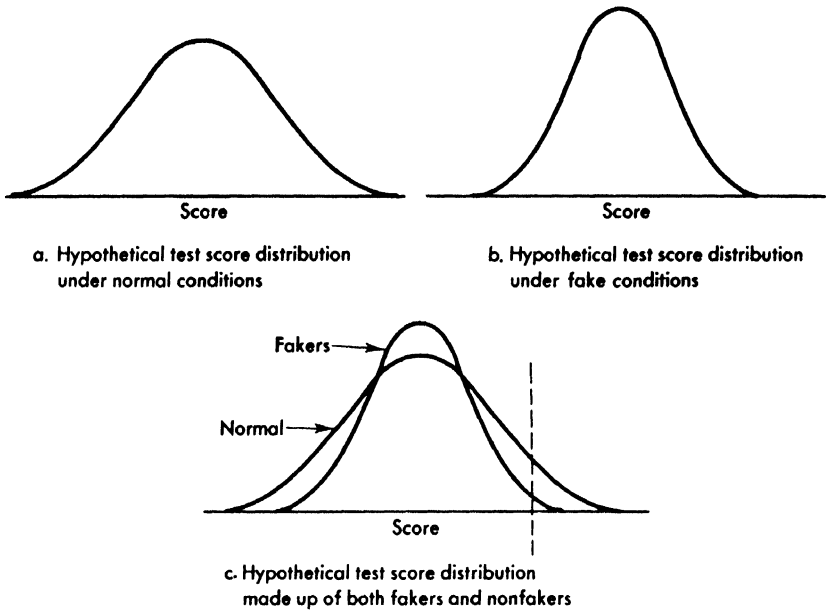


Figure 7.4. Effects of faking upon the distribution of forced-choice test scores based upon Norman's rationale.

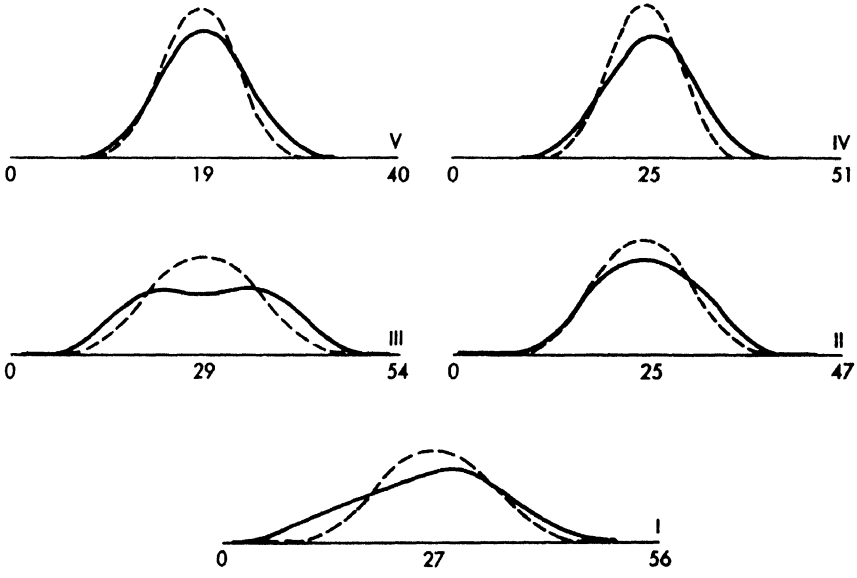


Figure 7.5. Forced-choice score distribution under normal conditions (solid line) and fake conditions (dotted line) illustrating restriction of variance under fake conditions. (Adapted from W. T. Norman. Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology*, 1963, 47, 236.)

A final characteristic of the Norman method is that it is very easy to construct a "detection" scale for determining who is faking and who is not. One simply includes on the scale a number of items which meet the following requirements:

1. They must have near zero validities for the criterion variables.
2. They must have displayed a large shift in popularity from the normal to the fake condition.
3. Their popularity under the normal condition must either be very high or very low.

One then constructs a key which scores the *infrequent* response under the normal condition for each item (that is, the frequent response under the fake condition), since one response tends to indicate faking and the other response indicates normal responding. Another way of describing these detection items is that their popularity

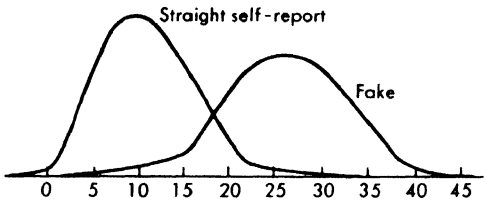


Figure 7.6. Distribution of persons on the detection key under straight self-report and faking conditions (Adapted from W. T. Norman. Personality measurement, faking, and detection An assessment method for use in personnel selection. *Journal of Applied Psychology*, 1963, 47, 236.)

correlates with tendency to fake. Figure 7.6 shows the distributions of scores on the detector scale obtained by Norman under normal and under fake conditions. Notice the large shift of scores when people were asked to fake. A cutoff point of about 20 would tend to identify most of the fakers without falsely accusing very many of the normals.

RATER CHARACTERISTICS

So far we have ignored the actual raters themselves and their importance to the performance appraisal process. While it has been indicated that raters are subject to numerous kinds of "errors" in making judgments and that rating scales should be designed to minimize these errors, the characteristics of the individual raters have not been examined as to their effect on the rating process.

RATER RELIABILITY AND PREDICTABILITY OF RATINGS

The problem of rater characteristics has received a good deal of attention in recent years. Buckner (1959) has explored the relationship between how well raters agree and the degree to which one can predict ratings with some other variable. Stated another way, Buckner asked the question "Do highly reliable ratings imply

that I have a criterion that is going to be easier to predict?" The answer turned out to be *no*. Buckner divided his ratees into four groups, depending upon how well the judges agreed in assigning their ratings.

Group 1: Judges were in high agreement on these ratees
Group 2: Judges were in moderate agreement on these ratees
Group 3: Judges were in low agreement on these ratees
Group 4: Judges were in zero agreement on these ratees

He then proceeded to validate two tests by correlating test results against rating —this was done for each group separately. He found no systematic relationship between the size of his obtained validities as a function of the group with which he was working.

Windle and Dingman (1960) criticized Buckner's interpretation and did a second study in which they found the results given in Table 7.4.

TABLE 7.4 *Correlations Between Intelligence Test Scores and Criterion Ratings Differing in Reliability*

Criterion Ratings	*N*	*Validity*
Reliable ratings from reliable raters	243	+0.37
Reliable ratings from unreliable raters	119	+0.23
Unreliable ratings from reliable raters	56	+0.38
Unreliable ratings from unreliable raters	131	+0.17
Total:	549	+0.31

SOURCE: Adapted from C. D. Windle and H. F. Dingman. Interrater agreement and predictive validity. *Journal of Applied Psychology*, 1960, *44*, 203–204.

Notice the relative magnitudes of the validity coefficients shown in Table 7.4. The more reliable the raters, the higher the validities. Similarly, the more reliable the ratings, the higher the validities. In the discussion in Chapter 2 on the relationship of reliability and validity, it may be recalled that the reliability of a variable sets *limits* on the size of the validity coefficients which can be obtained. Therefore, one might logically expect results more like those of Windle and Dingman than those of Buckner. However, one must also keep in mind that high reliability is only a necessary condition for high validities—it is not a sufficient condition.

Wiley, in a series of studies, has examined the consistency of rater judgments over time with respect to how well they agree with prior judgments by the same rater (Wiley, 1963; Wiley and Jenkins, 1963) and how well they agree with a group composite rating (Wiley and Jenkins, 1964). In general, raters were found to be consistent in their ratings up to a time period of ten months. In addition, he found that those raters who agreed very closely with the group composite in an initial rating task also were the raters who agreed most closely with the group composite on a different rating task one month later. He suggests that this knowledge can be used to select raters who are truly representative of the average consensus of a larger group of raters.

Unfortunately, the problem of whether or not raters who agree with the group composite are the best raters to solicit has not itself been determined. However, there is logic to the position that if it is possible to get a composite rating using a

small number of raters that will approximate the composite obtained using a larger number, one can certainly save time and money.

Numerous other rater characteristics have been shown to play a role in performance appraisals obtained by ratings. Christal and Madden (1960) have shown that an important consideration is the degree to which a rater is familiar with the occupation he is rating, a finding supported by additional studies by Madden (1960a, 1961). Similarly, Wiley, Harber, and Giorgia (1959a, 1959b) have demonstrated the influence of generalized rater tendencies to be noticeable in their effect upon ratings.

RATING-SCALE FORMAT AND PERFORMANCE APPRAISAL JUDGMENT

Madden has reported a number of studies dealing with the influence of the rating scale itself. In one study (Madden, 1960b) he found that rating reliability and ease of rating were not influenced by the use or lack of use of examples in defining the rating scale, but that whether the scale was defined as opposed to not being defined *did* affect reliability and ease of use. In a subsequent study Madden and Bourdon (1964) studied the effect of seven different rating-scale formats upon the ratings of 15 different occupations on 9 different job factors. The results, although somewhat complicated in nature, clearly indicated that the rating assigned to an occupation was dependent upon both the job factor involved and the *particular rating-scale format being used.*

DIMENSION OF CRITERIA

The industrial psychologist has a wide selection of potential criteria available when he undertakes the task of appraising the performance of workers. It is also clear, however, that each of these measures may be appraising a somewhat different aspect of job performance. The problems of criterion relevance and single versus composite criterion measures are of paramount importance.

One approach to simplifying the problem is to determine how the different performance appraisal measures relate to each other. Are they, in general, all very similar in what they are measuring, or do they tend to group themselves into smaller subclusters of criteria which have a lot in common within a cluster but little in common between clusters? Information of this sort could be very useful in helping decide which criteria should be selected for use in a particular situation. Data of this sort also have the potential for providing a better understanding of the underlying dimensions of work behavior.

Generally, such studies employ the method of factor analysis upon a large number of criterion or performance appraisal variables. Because the authors consider such criterion grouping studies to be rather important anchor points in the domain of performance appraisal, several such studies will be described.

TURNER STUDY: DIMENSIONS OF FOREMAN PERFORMANCE

Turner (1960) examined, using factor analysis, 20 different measures of foreman performance in an attempt to identify underlying dimensions (factors) of job performance. Nine of the measures were ratings and the remaining eleven were

TABLE 7.5 *Reliabilities and Rotated Factor Loadings of Criterion Measures*

	Plant X							Plant Y						
	Reliability		Factor Loadings[a]					Reliability		Factor Loadings[a]				
Criterion Measure	*n*	r_{tt}	1	2	3	4	5	*n*	r_{tt}	1	2	3	4	5
Grievances	99	0.28	21	42	−34	−23	23	104	0.30	05	28	07	07	26
Turnover	102	0.37	−17	29	−06	36	−06	104	0.28	−10	50	−07	40	−19
Absences	102	0.74	01	46	11	00	05	104	0.60	11	56	−02	−02	06
Suggestions	102	0.14	−06	19	−28	04	05	104	0.45	05	−12	−44	34	−10
Hospital passes	102	0 26	18	30	−22	−10	−19	90	0.57	09	−28	22	12	39
Disciplines	102	0.32	−09	34	−14	−02	33	104	0.39	−15	57	01	−05	−05
Absentee flexibility	90	0.36	−09	14	10	−44	−02	91	0.74	02	−15	−44	−20	−01
Scrap	99	0.88	02	12	61	−05	−01	91	0.76	−14	−10	64	01	−05
Expense (tools)	102	0 79	−04	45	15	04	−14	91	0.87	02	35	35	−06	−27
Expense (processing supplies)	102	0.88	10	09	04	−20	−27	91	0.79	−12	−15	−41	−23	−08
Efficiency	102	0.86	10	02	−28	25	−02	104	0 68	05	−08	−02	47	−03
Quantity rating	100	0.88	90	−01	−14	−15	−04	97	0 82	69	02	−10	−28	28
Quality rating	100	0.87	87	08	04	06	22	97	0 81	72	08	−07	−22	−03
Cooperation rating	100	0.82	71	04	−08	−09	−11	97	0 82	49	−15	10	11	24
Cost control rating	100	0 83	80	05	14	06	21	97	0 91	85	00	−04	04	06
Organization rating	100	0.85	87	11	05	−19	08	97	0.86	89	02	−08	06	06
Employee relations rating	100	0 87	77	20	−05	03	−27	97	0.90	72	29	−04	13	20
Safety rating	100	0 82	60	11	24	25	−11	97	0.77	53	−06	−25	06	−13
Housekeeping rating	100	0.88	54	15	29	25	21	97	0.76	62	−07	04	05	−26
Overall performance rating	100	0.92	95	00	00	04	01	97	0.87	95	−02	02	01	01

[a] Decimals omitted.

SOURCE: W. W. Turner. Dimensions of foreman performance: A factor analysis of criterion measures. *Journal of Applied Psychology*, 1960, *44*, 216–223.

objective performance measures. The results of his factor analyses (done separately in each of two different plants) are given in Table 7.5. The first two dimensions turned out to be highly similar in both plants, while the third and fourth factors were less alike.

Factor 1 was defined as "Job Performance Reputation." Looking at Table 7.5 it is easy to see that the large loadings on factor 1 came primarily from rating criteria in both plants. This indicates two very important conclusions about ratings, at least in this context. First, they seem to measure similar things even though different traits were supposedly being evaluated. Such a finding may be indicative of a general halo bias on the part of the raters. Second, and more important, *the analysis indicates that the ratings are independent of the other measures of performance.* That is, there appears to be little relationship between the ratings given to the foremen and their performance on the various objective criteria used in the study.

Factor 2 turned out to be an "Employee Relations" dimension. Criteria loading high on this factor were those that indicated a good relationship between foremen and their subordinates, such as frequency of absences, of disciplines, and a tendency to stay within budget allowances. Factors 3 and 4 are less easily defined, although Turner called factor 3 a "Scrap versus Organization of Production Operations" dimension. He did not attempt to name factor 4.

DUGAN STUDY: DIMENSIONS OF SALES PERFORMANCE

A similar investigation of the interrelationships among performance measures was made by Dugan (1960) in the insurance field. He obtained 19 different measures for agents in 41 different states. He then grouped these measures, using factor analysis, into five different clusters. These are shown in Table 7.6.

The five clusters can be identified in terms of the high loading variables which serve to define them. Dugan named them as follows:

Factor I: *Absolute Size.* The size of a state as measured by total car registration, policies in force, potential, total sales, total number of agents, number of career agents, and total number of jobs in the state.

Factor II: *Potential per Agent.* This factor appeared to measure the degree of potential available to each agent in the state and was identified by sales to potential, potential per job, and registrations per career agent.

Factor III: *Overall Effectiveness.* This factor gave an overall measure of the effectiveness of the state sales organization. It was defined by sales per agent and sales per job and also identified by percentage of agents who were career agents, production of part-time agents, percentage of career agents to total jobs, sales per policies in force, increase to written business, and sales to potential.

Factor IV: *Manpower Utilization.* This factor was identified by the percentage of agents who were full time, production of the part-time agents, percentage of the agents to the total jobs in the state, and sales per agent.

Factor V: *Rate of Growth.* This factor was identified by sales per policies in force, ratio of increase to written business, and policies per job (negatively).

Table 7.6 *Rotated Factor Matrix for 19 Variables*

Variables	Factors				
	I Absolute Size	*II* Potential per Agent	*III* Overall Effectiveness	*IV* Manpower Utilization	*V* Rate of Growth
Registrations	95**	23	03	−04	10
Policies in force	97**	−15	00	10	−03
Potential	92**	29	07	−07	17
Total sales	94**	−11	20	07	14
Total agents	93**	−07	−21	−08	07
Number of career agents	95**	−10	11	15	15
Total number of jobs	97**	−08	−08	00	10
Sales to potential	−08	−86**	35*	00	−18
Potential to jobs	07	98**	04	−08	10
Registration per career agent	−14	86**	−13	−47**	−01
Percentage of agents with 1200 points	25	−06	63**	63**	07
Production of part-time agents	33*	−01	43**	56**	12
Percentage of career agents to total jobs	21	−15	54**	68**	07
Sales per job	−18	−19	91**	04	−15
Sales per agent	10	−10	89**	32*	04
Sales per policy	−19	30*	65**	−14	55**
Increase to written business	18	05	33*	09	50**
Policies per job	02	−44**	−07	15	−68**
Loss ratio	−13	12	−03	−08	−17

* $r = 30 = 0.05$ level.
** $r = 39 = 0.01$ level.
SOURCE: Adapted from R. D. Dugan. Evaluating territorial sales efforts. *Journal of Applied Psychology,* 1960, *44,* 109.

TABLE 7.7 *Factor Analysis of Contributions of One Sample of Scientists (only loadings of 0.20 or greater are listed)*

Contribution	*Categories (or Factors)*															*No. of Significant Loadings*
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
Articles	58							−26	22							3
Reports, memos	37															1
Cooperation	−36							64								2
Patent rate	29			21		24										3
Prod. predicted	27								40							2
Educ. (years)	27				23											2
Write up incompl.	−26															1
Papers completed	21															1
Reports—2 yrs.		62												23		2
Pub. est.		23						•								1
Monitor		22				23								59		3
Drive, resource		21						63								2
Free-pub. bind.		21								−30						2
Persistence		20						60								2
Organ. reports			70													1
Relev., reports			62													1
Eleq., reports			54													1
Signif., reports			42	51												2
Integrity			23					48								2
Orig., reports				72							−21					2
Suggestions				21								−24				2
Societies					53											1
Activities					22											1
No. supervised					21					34	40		24			4
Months in center					−20					23					48	3

Prod. rank. (peer)	61			29						2
Prod. rank. (supervisor)	58		20	29				25		4
Consultant	39									1
Lab nomin.	35			20	49					3
New problems	27		23							2
50% retention	23			51						2
Math, cogn., etc.	20		54							2
Sci. char. cklist		84								1
Creat. cklist		79								1
Prod. cklist		66		20						2
Flexibility			67							1
Indep. discovery			65							1
Desire for facts			54							1
Creation			51							1
Inform. ability			50							1
Likability				39						1
Trustworthy				−22						1
Center nomin.				20	67	21				3
Papers read					34	38				2
Awards					25	54				2
Noncomply, alter.					−20					1
Promotion rate							54		−23	2
Work years							−23		80	2
Pay								89		1
Primary activity								81		1
Proj. coop.									−28	1
Age									82	1

NOTE: Decimal points omitted.

SOURCE: C. W. Taylor, *et al.* Explorations in the measurement and prediction of contributions of one sample of scientists. ASD-TR-61-96, Personnel Laboratory, Lackland Air Force Base, Texas, 1961, 7.

TAYLOR STUDY: DIMENSIONS OF SCIENTIFIC PERFORMANCE

A third study using the factor analytic methodology has been reported by Taylor, Smith, Ghiselin, and Ellison (1961). They, however, were interested primarily in the dimension of job performance to be found with scientific tasks. They interviewed scientists at two Air Force research centers concerning the nature of scientific productivity and the characteristics of effective scientists. Using these interview suggestions as a guide, they proceeded to collect data on 52 different measures of job performance from 215 physical scientists working in 15 different laboratories. The data were factor analyzed to obtain information about the various independent dimensions of scientific performance. Table 7.7 is a summary of the factor analysis.

The analysis resulted in 15 relatively distinct dimensions of scientific contribution. The authors interpreted or defined each of the dimensions as follows:

Factor	*Name or Definition*
A	Productivity in written scientific work (general)
B	Short-term written productivity
C	Rated excellence of research reports
D	Originality of work and thought of written reports
E	Professional society membership
F	Rated work output (supervisors and coworkers)
G	Rated creativity and productivity (supervisors)
H	Rated global assessment (supervisors)
I	General likability in organization
J	General visibility in organization
K	Official organizational recognition
L	"Organizational-man" tendencies
M	Current organizational status
N	Contract monitoring load
O	Length of work experience

A number of interesting facts emerge from the factor analysis. Again we find that ratings tend to group together into clusters which are not related to more objective measures of job performance. Also, both pay and promotion rate are quite separate from nearly *all* the measures used in the study. In fact, the only serious predictors of salary were the *number of people supervised* and the *primary activity* in which a scientist was engaged. Nothing else was relevant.

The authors then took these factors and tried to predict them using different types of tests. These data have previously been discussed (see Chapter 4) and will not be elaborated here.

Taylor, Price, Richards, and Jacobsen (1964) have subsequently done much the same sort of thing with criteria in medical schools. Using interviews, questionnaires, compendium listings, medical college files, and peer evaluations, 80 criterion measures of *on-the-job* performance of 102 college of medicine faculty members were factor analyzed. Twenty-five different factors were obtained in the analysis, indicating this area to be an extremely complex one with regard to the criterion problem and performance appraisal.

SEASHORE STUDY: GENERALITY OF CRITERION INTERRELATIONSHIPS

The three studies just discussed are excellent examples of the current trend in criterion research. A fourth example is an excellent and important study by Seashore, Indik, and Georgopoulos (1960). In this study the authors asked not only what were the interrelationships among several criteria, but also whether or not these relationships remained stable across situations—a rather critical question. They had three hypotheses which they specifically wanted to evaluate:

Hypothesis 1: The intercorrelations among a set of job performance measures for a given homogeneous population of individuals or organizations should be consistent (in terms of positive or negative association) and relatively high.

Hypothesis 2: The pattern of intercorrelations among a set of job performance variables, allowing for error of measurement, should be similar in size and sign at various levels of organizational analysis.

Hypothesis 3: The relationships among job performance criteria for the individuals within any one organization are, within sampling error, representative of the relationships which hold across a homogeneous set of such organizations.

Five criteria were used in the study. Tables 7.8 and 7.9 present the findings of the research. From Table 7.8 it can be seen that when criterion intercorrelations were computed among the 27 different stations (using station means), four of the ten correlations were significant. When they were computed based upon the data from all 975 individuals located at these 27 stations, six of the ten were significant. These findings indicate that Hypothesis 1 does not seem to be substantiated. Similarly, comparing the patterns of correlations obtained in population A with that of population B indicates that Hypothesis 2 *does* seem to be reasonable.

To test the third hypothesis, the intercorrelations among the five variables were computed separately at each of the 27 stations. Line 1 of Table 7.9 gives the range of these correlations for each pairing, and line 3 gives the mean of the 27 correlations for each pair. If the third hypothesis were true, one would expect the correlations between any pair of variables to be the same for all stations, that is, their range and variability should be very small. This was only true for two pairs, productivity versus accidents and absences versus errors. Thus the third hypothesis does not seem to be supported by the data.

The conclusion one can attach to the overall study is that (1) there is no support for the "single criterion dimension" notion (Hypothesis 1), and thus combining criteria is dangerous, (2) the basic dimensions of job performance can be affected by the level of organizational analysis at which one is working, and (3) there is also likely to be considerable variation in interrelationship among criteria measures from one organization to another even though the organizations are ostensibly quite similar.

OTHER CLUSTERING RESEARCH

Investigations other than those discussed in the previous sections have been reported which have been concerned with the problem of criterion grouping. Of these, the research of Schultz and Siegle (1961a, 1961b, 1964) is probably the most interesting in that they have been concerned with applying clustering methods other than standard factor analyses. In one study (Schultz and Siegle, 1964) they

TABLE 7.8 *Intercorrelations Among Five Job Performance Criterion Variables in 27 Operating Stations of a Delivery Service Firm*

Criteria	Population A (Station)				Population B (Individuals)				Population C (Individuals Within Stations)			
	2	3	4	5	2	3	4	5	2	3	4	5
1. Effectiveness	74*	25	−10	−46*	28*	−02	−08	−32*	36*	−05	−17*	−42*
2. Productivity	—	30	08	−39*	—	.12*	−01	−26*	—	−01	02	−23*
3. Chargeable accidents		—	02	−65*		—	03	−18*		—	09	04
4. Unexcused absences			—	−11			—	15*			—	17*
5. Errors (nondeliveries)				—				—				—

* Significant at 0.05 level.

SOURCE: S. E. Seashore, *et al.* Relationships among criteria of job performance. *Journal of Applied Psychology,* 1960, *44,* 198.

TABLE 7.9 *Summary of 270 Intercorrelations Among Five Job Performance Criterion Variables for 975 Nonsupervisory Employees Within Each of 27 Operating Stations of a Delivery Service Firm*

	Effectiveness Versus				*Productivity Versus*			*Accidents Versus*		*Absences Versus*
	Productivity	*Accidents*	*Absenses*	*Errors*	*Accidents*	*Absences*	*Errors*	*Absences*	*Errors*	*Errors*
1. Obtained range of correlations	−0.558 +0.830	−0.441 +0.443	−0.444 +0.359	−0.737 +0.232	−0.257 +0.503	−0.342 +0.373	−0.888 +0.282	−0.234 +0.962	−0.502 +0.498	−0.117 +0.509
2. Number of + correlations − that are 0	26 1 0	11 16 0	5 14 8	2 25 0	11 16 0	8 9 10	4 22 1	4 13 10	14 13 0	11 6 10
3. Weighted average within group correlation	+0.357	−0.046	−0.169	−0.421	−0.014	+0.019	−0.224	+0.088	+0.038	+0.168
4. Probability that the obtained correlations are homogeneous	Very low	Low	Very low	Very low	High	Low	Very low	Very low	Low	High

SOURCE: S. E. Seashore, *et al.* Relationships among criteria of job performance. *Journal of Applied Psychology*, 1960, *44*, 199.

applied advanced multidimensional scaling procedures to 18 different job tasks and found that these tasks could be reduced to four basic job dimensions. Their methodology is very promising and is particularly appealing because it is an extremely simple method in many respects. It is likely that many future clustering studies will employ methods of this newer type.

The research discussed in the preceding sections clearly points out that not only is the criterion-performance appraisal problem very complex, but the nature of its complexity tends to vary from occupation to occupation.

Both of the Taylor *et al.* studies were done with rather specialized and high-level people. Their first study was with physical scientists and their second examined medical school faculties. It is no accident that research in criterion development and performance appraisal is now being done with occupational groups of a professional and semiprofessional nature. It is with these kinds of individuals that a simple measure of success becomes almost impossible to find and/or even to define. One should become more knowledgeable about the criterion complex and be concerned with theory constructions, and then design experimental studies. It is clear that progress will not occur until concentrated efforts are exerted along such lines.

EXECUTIVE PERFORMANCE APPRAISAL

The executive category is a most difficult group for which to develop criteria. The problems inherent in executive performance appraisal are many and varied. Even in the same organization the kinds of activities two different executives are likely to engage in may differ considerably—thus, the criterion or criteria may have to be made "position specific." Even when there is apparent similarity among executive positions to encode them as a group, there still remains the difficult problem of developing criteria of performance. As a result the psychologist must rely upon some form of rating more out of necessity than out of choice. In addition, in most such situations the actual number of executives available for use in criterion development is usually very limited. All these problems tend to make the task of objective executive performance appraisal an exceedingly complex and difficult one. In fact, there is probably no area of criterion development more critically in need of new techniques and data than the executive performance area.

IN-BASKET TEST

A novel method of executive performance appraisal is known as the "in-basket" technique. First suggested by Frederiksen in 1957 (Frederiksen, Saunders, and Ward, 1957), the "in-basket" is a situational test which attempts to simulate important aspects of the job of an executive. In its more typical form it consists of the letters, memoranda, telephone calls, and other materials which have supposedly accumulated in the in-basket of a hypothetical executive.

The person being evaluated or appraised using the in-basket test is given appropriate background information concerning the business organization as a frame of reference. He is typically instructed that he is to consider himself the new incumbent in the executive position (various explanations for this can be given) and that he is to deal with the material in the in-basket. The amount of background information

is always detailed enough so that the incumbent can reasonably be expected to take action on the majority of problems presented to him as he proceeds through the series of subtasks. As Frederiksen (1962, p. 1) has explained,

> He is to bring to the new job his own background of knowledge and experience, his own personality, and he is to deal with the problems as though he were really the incumbent of the administrative position. He is not to say what he *would* do; he is actually to write letters and memoranda, prepare agenda for meetings, make notes and reminders for himself, as though he were actually on the job.

Thus the situation for the incumbent is relatively unstructured in terms of the response he is permitted to make. However, the in-basket contains a specified set of problem situations which become a type of standardized performance test used to approximate actual on-the-job executive behavior. In fact, one might also view it as a rather elaborate performance test of personality.

Research using the in-basket procedure can be classified into such general categories as:

1. Determining dimensions of executive performance (a research device)
2. Selecting potential executives (a selection device)
3. Evaluating potential executives (a criterion measure)
4. Training potential executives (a training measure)

Of these four functions, the first is really a necessary prerequisite to the latter three. That is, it is extremely important that one obtain information about the basic characteristics of the in-basket itself before attempting to utilize it for the other listed objectives. This is particularly true in the case of a complicated performance appraisal tool like the in-basket.

DIMENSIONS OF IN-BASKET PERFORMANCE Frederiksen (1962) conducted a rather extensive study intended to ascertain what basic behavioral dimensions of executive performance were involved in in-basket performance. He gave the test to 355 individuals who represented five different "types" of people (1 = college undergraduate, 2 = graduate students of business, 3 = government administrators, 4 = business executives, 5 = Army officers). He scored each person on 70 different categories of in-basket behavior. Of these, the most reliable 40 measures were factor analyzed. The major dimensions he obtained were:

1. *Preparing for Action.* People high on this factor tend to defer final decisions and actions on problems and instead spend time preparing to make the decisions.
2. *Amount of Work* People high on this have more work output than those who are low.
3. *Seeking Guidance.* People high on this factor appear to be anxious to please their superiors and are continually seeking the advice of others.

The study also indicated substantial differences in patterns of behavior among the five different subgroups studied. For example, undergraduates had very high verbal productivity, government administrators had a great lack of concern with outsiders, business executives showed extreme courtesy, and Army officers tended to exert strong control over subordinates.

In another study Frederiksen (1961) has reported the reliability of performance in an in-basket situation across a variety of in-basket protocols. He found (1) no

differences in the reliabilities of the scores developed from each of the various in-baskets used, and (2) the reliability of performance within an in-basket tended to be much greater than consistency of performance across in-baskets. He attributed the latter finding to the notion that each in-basket creates a kind of "set" or attitude which prevails across all items but which may differ from in-basket to in-basket.

As a method for studying complex performance, the in-basket seems to have a good deal of promise. But as a method of criterion development and performance appraisal, it is of course subject to the same criticisms as any other type of simulated task activity: how representative is the in-basket of real-life work situations? Are the situations themselves realistic and, perhaps even more important, does the person taking the test respond in a simulated situation in the same manner as he would respond to the same item in a real-life situation? On the other hand, simulated task activities do save time and money. The in-basket needs much more research before it can be used with assurance as either a criterion or a predictor of executive performance.

OTHER EXECUTIVE PERFORMANCE APPRAISAL EFFORTS

A technique for appraising the performance of research managers which is based upon an Operations Research (OR) Model has been reported by Lamouria and Harrell (1963). They suggest that OR methods can result in a more objective criterion. However, their research has been severely criticized by Marks (1965), who eloquently points out many of the difficulties involved when one tries to employ models from other disciplines as possible solutions to the criterion problem.

Another approach is that of Forehand (1963), who suggests that when performance is very complex the best procedure is to focus upon partial criteria rather than global assessment. He further suggests that the complex criterion problem has two stages:

1. Assessing particular behaviors
2. Determining the relevance of the selected behaviors to the overall success of the executive

Forehand used several different ratings of "innovative" behavior as his partial criterion measure. However, he found that it was generally unrelated to other, more direct assessments of general executive effectiveness, although a forced-choice measure of innovative behavior did, under some conditions, correlate with the effective ratings.

PERFORMANCE APPRAISAL INTERVIEWS

By this time it should be more than apparent that the problems of performance appraisal are both numerous and diversified in nature. Several recent studies are worthy of mention that have provided additional insight into some of the complications which exist in performance appraisal situations involving the interview.

PERFORMANCE APPRAISAL AND FEEDBACK

In some circumstances performance appraisals are used as a basis for counseling employees and attempting to improve their subsequent performance. Meyer and Walker (1961) studied the effectiveness of performance appraisals in such a setting and found the most critical aspect to be the *way* in which the information is presented to the employee rather than the actual information itself. Thus, skill of performance appraisal feedback is apparently a very critical variable when such information is used to inform employees about themselves and their work.

EFFECTS OF THREAT IN PERFORMANCE APPRAISAL

In what was a very intriguing piece of research, Kay, Meyer, and French (1965) studied the influence of threat upon the effectiveness of performance appraisals. They analyzed 92 management-subordinate appraisal interviews, obtaining reactions of subordinates before and after their appraisal interviews. Measures of improvement in performance were obtained on each subordinate 12 weeks after the appraisal. The results of the study (p. 311) indicated that:

> . . . a manager's attempts to assist a subordinate by pointing up improvement needs were likely to be perceived by the subordinate as threatening to his self-esteem and to result in defensive behavior. The greater the threat, the less favorable the attitude toward the appraisal system and the less the subsequent constructive improvement in job performance realized.

PSYCHOLOGICAL DISTANCE AND PERFORMANCE APPRAISAL

Rothaus, Morton, and Hanson (1965) used role-playing to investigate the effect of "psychological distance" upon the rating attitudes of supervisors and subordinates in a performance appraisal interview.

Nursing administrators were assigned the roles of supervisor and subordinate and were assigned into one of two groups, observer and participant. Participants were given appropriate role instructions for the appraisal interviews, and observers were told to assume that they were peers of the participants and to observe the interaction. Observers did not participate in any way themselves. Some of the interviews were conducted in private; that is, no observers were present. Everyone made initial trait ratings of the hypothetical subordinate to be appraised in the interview. Following the role-playing, the participant supervisors and subordinates rated their reactions.

The results indicated that in the role-playing situation supervisors of both observer and participant groups were more negative in their initial trait ratings than were subordinates. In addition, observers were more negative in their ratings than participants.

The authors concluded that as the role a person assumes places him at a greater psychological distance from a person to be appraised, his ratings tend to become more negative and more critical—a finding which has important implications for the entire performance appraisal process.

REFERENCES

Argyle, M., G. Gardner, and F. Cioffi. Supervisory methods related to productivity, absenteeism, and labor turnover. *Human Relations*, 1958, *11*, 23–40.

Baumgartel, H., and R. Sobol. Background and organizational factors in absenteeism. *Personnel Psychology*, 1959, *12*, 431–443.

Buckner, D. N. The predictability of ratings as a function of interrater agreement. *Journal of Applied Psychology*, 1959, *43*, 60–64.

Buel, W. The validity of behavioral scale items for the assessment of individual creativity. *Journal of Applied Psychology*, 1960, *44*, 407–412.

Christal, R. E., and J. M. Madden. Effect of degree of familiarity in job evaluation. USAF WADD *Personnel Laboratory Technical Note*, 1960 (Nov.), No. 60-263.

Dugan, R. D. Evaluating territorial sales efforts *Journal of Applied Psychology*, 1960, *44*, 107–110.

Edwards, A. E. *Techniques of attitude scale construction*. Appleton-Century-Crofts, New York, 1957.

Flanagan, J. C. The critical incident technique. *Psychological Bulletin*, 1954, *51*, 327–358.

Fleishman, E. A. Leadership climate, human relations training, and supervisory behavior. *Personnel Psychology*, 1955, *6*, 205–222.

Forehand, G. A. Assessments of innovative behavior. Partial criteria for the assessment of executive performance. *Journal of Applied Psychology*, 1963, *47* (3), 206–213.

Frederiksen, N. Consistency of performance in simulated situations. Office of Naval Research and Educational Testing Service Research Bulletin 61-22. Educational Testing Service, Princeton, N.J., 1961.

Frederiksen, N. Factors in in-basket performance. *Psychological Monographs*, 1962, *76*, No. 22 (Whole No. 541).

Frederiksen, N., D. R. Saunders, and B. Ward. The in-basket test. *Psychological Monographs*, 1957, *71*, No. 9 (Whole No. 438).

Ghiselli, E. E., and M. Haire. The validation of selection tests in the light of the dynamic character of criteria *Personnel Psychology*, 1960, *13*, 225–231.

Guilford, J. P. *Psychometric methods*, 2nd ed. McGraw-Hill, New York, 1954, 274–278.

Highland, R. W., and J. R. Berkshire. A methodological study of forced-choice performance ratings. *Research Bulletin*, 51–9, Lackland Air Force Base, San Antonio, Texas, 1951.

Howe, E. S. Further comparisons of two short-form derivatives of the Taylor Manifest Anxiety Scale. *Psychological Reports*, 1960, *6*, 21–22.

Howe, E. S , and A. B Silverstein. Comparison of two short-form derivatives of the Taylor manifest anxiety scale. *Psychological Reports*, 1960, *6*, 9–10.

Izard, B. R., and S. Rosenberg. Effectiveness of a forced-choice leadership test under varied experimental conditions. *Educational and Psychological Measurement*, 1958, *18*, 57–62.

Karr, C. A comparison of EPPS scores obtained from the standard forced-choice procedure and a rating scale procedure. *Dissertation Abstracts*, 1959, *19*, 3382–3387.

Kay, E., H. H. Meyer, and J. R. P. French, Sr. Effects of threat in a performance appraisal interview. *Journal of Applied Psychology*, 1965, *49*, 311–317

Kirchner, W. K., and M. D. Dunnette. Using critical incidents to measure job proficiency factors. *Personnel*, 1957, *34*, 54–59.

Lamouria, L. H., and T. W. Harrell. An approach to an objective criterion for research managers. *Journal of Applied Psychology*, December 1963, *47* (6), 353–357.

Madden, J. M. A comparison of three methods of rating scale construction. USAF WADD *Personnel Laboratory Technical Note*, November 1960 (b), No. 60-262.

Madden, J. M. Familiarity effects in evaluative judgments. USAF WADD *Personnel Laboratory Technical Note*, November 1960 (a), No. 60-261.

Madden, J. M. A further note on the familiarity effect in job evaluation. USAF ASD *Personnel Laboratory Technical Note*, June 1961, No. 61-47.

Madden, J. M., and R. D. Bourdon. Effects of variations in rating scale format on judgment. *Journal of Applied Psychology*, 1964, *48* (3), 147–151.

Marks, M. R. "An objective criterion for research managers:" A critique. *Journal of Applied Psychology*, 1965, *49*, 91–92.

Meyer, H. H., and W. B. Walker. Need for achievement and risk performance as they relate to attitudes toward reward systems and performance appraisal in an industrial setting. *Journal of Applied Psychology*, 1961, *45*, 251–256.

Norman, W. T. Personality measurement, faking, and detection: an assessment method for use in personnel selection. *Journal of Applied Psychology*, 1963, *47*, 225–241.

Rothaus, P., R. B. Morton, and P. G. Hanson. Performance appraisal and psychological distance. *Journal of Applied Psychology*, 1965, *49*, 48–54.

Rothe, H. F. Output rates among butter wrappers: I. Work curves and their stability. *Journal of Applied Psychology*, 1946 (a), *30*, 199–211.

Rothe, H. F. Output rates among butter wrappers. II. Frequence distributions and a hypothesis regarding the "restriction of output." *Journal of Applied Psychology*, 1946 (b), *30*, 320–327.

Rothe, H. F. Output rates among chocolate dippers. *Journal of Applied Psychology*, 1951, *35*, 94–97.

Rothe, H. F. Output rates among machine operators: I. Distributions and their reliability. *Journal of Applied Psychology*, 1947, *31*, 384–389.

Rothe, H. F., and C. T. Nye. Output rates among coil winders. *Journal of Applied Psychology*, 1958, *42*, 182–186.

Rothe, H. F., and C. T. Nye. Output rates among machine operators: II. Consistency related to methods of pay. *Journal of Applied Psychology*, 1959, *43*, 417–420.

Rothe, H. F., and C. T. Nye. Output rates among machine operators. III. A nonincentive situation in two levels of business activity. *Journal of Applied Psychology*, 1961, *45*, 50–54.

Schultz, D. G., and A. I. Siegel. The analysis of job performance by multidimensional scaling techniques. *Journal of Applied Psychology*, 1964, *48*, 329–335.

Schultz, D. G., and A. I. Siegel. Generalized Thurstone and Guttman scales for measuring technical skills in job performance. *Journal of Applied Psychology*, 1961 (a), *45*, 137–142.

Schultz, D. G., and A. I. Siegel. Post-training performance criterion development and application. *Applied Psychological Services*, Wayne, Pa., 1961 (b).

Seashore, S. E., B. P. Indik, and B. S. Georgeopoulos. Relationships among criteria of job performance. *Journal of Applied Psychology*, 1960, *44*, 195–202.

Shartle, C. L., G. B. Brumback, and J. R. Rizzo. An approach to dimensions of value. *Journal of Applied Psychology*, 1964, *57*, 101–111.

Sisson, E. D. Forced choice, the new army rating. *Personnel Psychology*, 1948, *1*, 365–381.

Symonds, P. M. Notes on rating. *Journal of Applied Psychology*, 1925, *9*, 188–195.

Taylor, C. W., P. B. Price, J. M. Richards, and T. L. Jacobson. An investigation of the criterion problem for a medical school faculty. *Journal of Applied Psychology*, 1964, *48*, 294–301.

Taylor, C. W., W. R. Smith, B. Ghiselin, and R. Ellison. Explorations in the measurement and prediction of contributions of one sample of scientists. ASD-TR-61-96, Personnel Laboratory, Lackland Air Force Base, Texas, 1961.

Taylor, E. K., and R. J. Wherry. A study of leniency in two rating systems. *Personnel Psychology*, 1951, *4*, 39–47.

Turner, W. W. Dimensions of foreman performance: A factor analysis of criterion measures. *Journal of Applied Psychology*, 1960, *44*, 216–223.

Uhrbrock, R. S. 2000 scaled items. *Personnel Psychology*, 1961, *14*, 375–420.

Waters, L. K. A note on the "fakability" of forced-choice scales. *Personnel Psychology*, 1965, *2*, 187–192.

Wiley, L. Influence of time on rater bias measurements when estimating task requirements. *Journal of Industrial Psychology*, 1963, *1* (2), 39–43.

Wiley, L., H. B. Harber, and J. M. Giorgia. Evidence for a generalized rating tendency. *Engineering and Industrial Psychology*, 1959 (a), *1*, 55–61.

Wiley, L., H. B. Harber, and J. M. Giorgia. Rater tendencies in estimating qualifications required by Air Force tasks. USAF WADC *Personnel Laboratory Technical Note*, 1959 (b), (Sept.), No. 59-195.

Wiley, L., and W. S. Jenkins. Method for measuring bias in raters who estimate job qualifications. *Journal of Industrial Psychology*, 1963, *1* (1), 16–22.

Wiley, L., and W. S. Jenkins. Selecting competent raters. *Journal of Applied Psychology*, 1964, *48*, 215–217.

Windle, C. D., and H. F. Dingman. Interrater agreement and predictive validity. *Journal of Applied Psychology*, 1960, *44*, 203–204.

Zavala, A. Development of the forced-choice rating scale technique. *Psychological Bulletin*, 1965, *63*, 117–124.

TRAINING AND LEARNING 8

Training is a process that develops and improves skills related to performance. Effective training programs can result in increased production, reduced labor turnover, and greater employee satisfaction. They should include all employees, from factory workers to executives, and apply not only to inexperienced workers but also to experienced workers new to the company. A training program should also include those who are promoted to higher level jobs and the periodic retraining of present employees by means of "refresher" courses.

Despite the fact that job training should be one of industry's responsibilities, there are many companies which seem to do their best to avoid it. Industry in general has attempted to get around the problem of training in two ways: (1) It insists on hiring skilled or experienced workers on the assumption that training will be unnecessary, or (2) it hires inexperienced people and turns them over to experienced employees for training, under the mistaken belief that a worker who is experienced can train an inexperienced person even though he has no knowledge of training techniques.

Although learning can take place without effective aids, it is uneconomical because it takes longer and does not permit the ineffectively trained worker to reach his maximum output. The National Institute of Industrial Psychology carried out a series of experiments on two equated groups of workers. The essential difference between the groups was that one was hired and trained for the job it was to perform, and the other was put to work at the job in the usual manner. An investigation of about ten different jobs led to the conclusion that the trained group was from 10 to 40 percent more efficient than the untrained group.

Psychologists have gathered sufficient information in the laboratory, the classroom, the armed forces, and to a certain extent in industry to permit the knowledge about training to be classified into a series of general principles. While these principles should never be considered to represent absolute "truths" concerning training, it is generally safe to assume that using them as guidelines will result in more satisfactory training programs than will ignoring them.

THE LEARNING PROCESS

LEARNING CURVES

Regardless of the task involved, a clear similarity in the learning curve from task to task is usually evident. The curve is often characterized by a rapid increase in the earlier learning period and a tapering off toward the end of the process. The

pioneer studies in this area conducted by W. L. Bryan and N. Harter (1899) prior to 1900 resulted in the establishment of a learning curve for telegraphy; this is reproduced in Figure 8.1.

Familiarity with the typical learning curve allows for comparison between it and any specific learning process of a particular task. Marked deflections of the curve, if they are frequent, may indicate that something is wrong with the training method. Since the quality of teaching is vitally important, wide variations in what is learned are often the result of either efficient or inefficient teaching. Anyone who has ever been to school knows the difference between a good teacher and a poor one; it cannot be assumed that everyone can teach equally well.

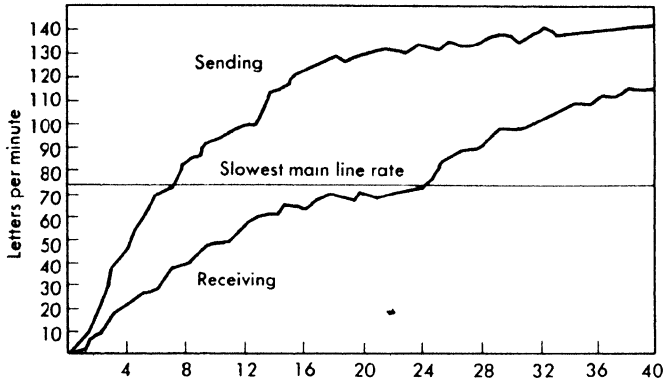


Figure 8.1. A learning curve for telegraphy. (From W. L. Bryan and N. Harter. Studies in the telegraphic language. *Psychological Review*, 1899, 6, 346–376.)

Often the concept "plateau" appears in connection with the learning curve. Sometimes there is a flatness in the learning curve which is eventually followed by a spurt. This flattening indicates a period of no apparent progress and is referred to as a plateau. There are many reasons for the appearance of such a plateau. It may be the result of lack of motivation, inefficient performance methods, or, very often, ineffective teaching or poor training. However, a plateau is *not* an integral part of the learning process; hence one should not be concerned by its absence.

A NOTE OF CAUTION ON LEARNING CURVES

While examination of learning curves can indeed provide an excellent means of comparing various training methods, various teachers, etc., it is important that one only compare learning curves which are based upon similar kinds of learning measures. Bahrick, Fitts, and Briggs (1957) have pointed out that the shape of the learning curve can be influenced tremendously by the kind of measure that one is using (time scores, quantity scores, etc.).

FACTORIAL COMPLEXITY OF LEARNING CURVES

Another difficulty with overall learning curves is that they really only represent a partial picture of the actual learning process. At best, they are a gross index of the change in skill level being experienced by an individual or by a group of individuals. This was first strikingly demonstrated by Fleishman and Hempel (1955). Using

the factor analysis methodology, they examined the relationship between the basic ability dimensions of the task and amount of practice. Figure 8.2 presents their findings.

Notice from Figure 8.2 that the importance of certain ability dimensions increases as the skill level increases. Thus with additional practice, reaction time and rate of movement account for more and more variance in performance. Conversely,

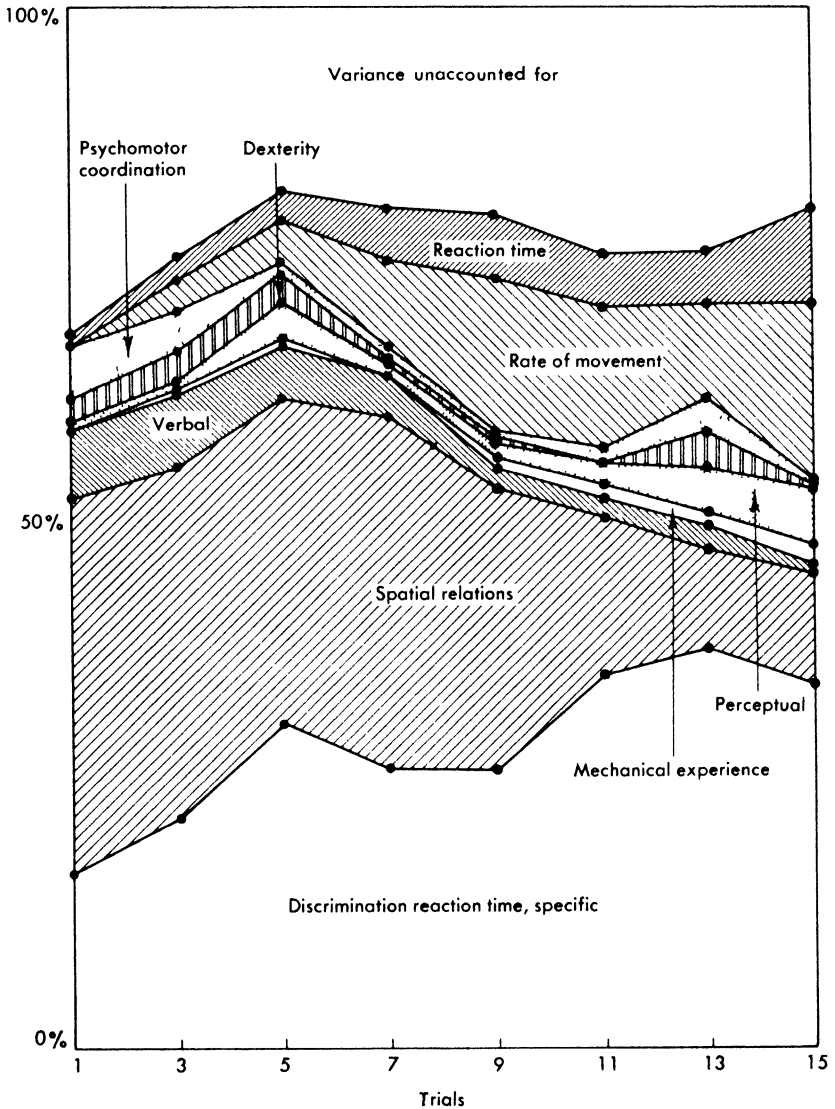


Figure 8.2. Percentage of variance represented by each factor at different stages of practice on the Discrimination Reaction Time Task. Percentage of variance is represented by the shaded area for each factor. (From E. A. Fleishman and W. E. Hempel, Jr. The relationship between abilities and improvement with practice in a visual discrimination task. *Journal of Experimental Psychology*, 1955, 49, 301–312.)

spatial relations ability accounts for less and less performance variance as skill increases. The deduction to be made is that individuals are indeed learning different "things" at different stages in the learning, and that the learning curve for any individual measure of task performance may not be truly representative of the skill acquisition process.

A second and equally dramatic demonstration of the same type of phenomenon has been reported by Fuchs (1962). Fuchs did his study using a complex "tracking" task requiring skills similar to those involved in flying an airplane. Any type of complex motor response can be separated into various components such as position, rate, acceleration (rate of change), and jerk (rate of rate of change). Fuchs was interested in testing the hypothesis that subjects learn these different components of a motor task at different periods in the overall learning process. Figure 8.3 shows the results of his study which indeed supported the hypothesis.

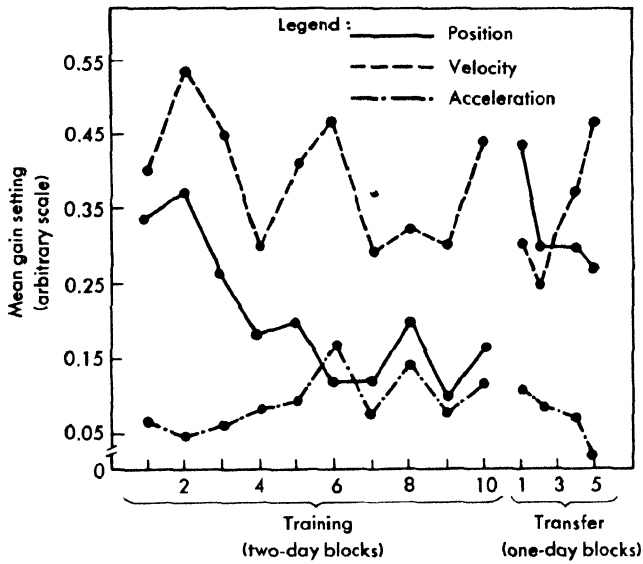


Figure 8.3. Mean weights (analog gain settings) assigned to stimulus components during training and transfer. (From A. H. Fuchs. The progression-regression hypotheses in perceptual-motor skill learning. *Journal of Experimental Psychology*, 1962, 63, 177–182.)

The more the subjects progressed in their training, the less their motor performance consisted of a position element and the more important became the acceleration component. Thus this study supports the contention that *what* is being learned tends to change during training—particularly with complex tasks—and that individual measures of skill acquisition may tend to give unrealistic or unrepresentative pictures of the extent to which "learning" is really taking place. In other words, learning must be defined in terms of the specific learning situation and the components that contribute to the learning process which ultimately results in a skill or performance ability.

PRINCIPLES OF LEARNING

It is customary when discussing the problems of training to begin by outlining a series of so-called "learning principles" which are conducive to gaining maximum efficiency in a learning situation. At this point we shall defer to this custom and present, in simple form, some traditional learning principles. However, the authors are of the opinion that blind adherence to these principles can often cause more harm than good, and that each should be interpreted and applied carefully in full consideration of the particular task being learned and the context in which the learning itself takes place.

KNOWLEDGE OF RESULTS

It is generally conceded that knowledge regarding one's own performance is a necessary condition for learning. The explanation for this fact is generally attributed to either the *informational* characteristic or the *reinforcing* characteristic of the knowledge of results. The informational characteristic may be best viewed as the *specificity* of the information received, while the reinforcing characteristic may be thought of as the extent to which the knowledge about one's own performance is capable of *motivating* the individual involved.

The term "knowledge of results," which is often simply referred to as KR, is sometimes used interchangeably with the term *feedback*. This latter term is of somewhat more recent vintage, having come to psychology from the engineering sciences. Feedback has been categorized into various types, depending upon the particular kind or source of information involved.[1]

EXTRINSIC VERSUS INTRINSIC FEEDBACK Knowledge about task performance can come from cues internal to the organism, such as muscle tension, general body equilibrium, etc. Such feedback is called *intrinsic* feedback. An example would be the knowledge about flight performance a pilot might receive when flying blind simply due to "G" forces upon his body.[2] *Extrinsic* feedback, on the other hand, refers to cues about performance which come from sources *outside* the organism.

PRIMARY VERSUS SECONDARY FEEDBACK Any knowledge of results which can be considered as being an integral part of the task itself is referred to as *primary* feedback. Information about performance from a source outside the task is called either *secondary* or *supplemental* feedback. For example, if one were target shooting, the holes in the target would provide primary feedback about performance; the comments and facial expressions of the instructor might be an excellent source of secondary feedback. Both primary and secondary feedback would be classed as extrinsic types of feedback.

AUGMENTED VERSUS SUMMARY FEEDBACK Both of these two classes of KR are types of secondary feedback. The primary differentiation is time. *Augmented* feedback is used to describe information about performance that occurs almost concurrently with the actual performance; there is a minimal time lag between the "doing" and the secondary feedback. Because of this short time lag, augmented feedback is

[1] The authors are indebted to Dr. George E. Briggs for his comments on the classification of feedback types.

[2] "G" stands for gravitational pull. A force of two Gs is equivalent to a force twice the strength of normal gravity.

usually very precise information and accordingly ranks high on the information-giving quality of KR. *Summary* feedback, as the name implies, involves a delayed overall summary of task performance. It is thus usually much more global in nature than is augmented feedback.

To illustrate, if an instructor were to comment on each test item individually as soon as it was completed by a trainee, this would be considered augmented feedback. If, on the other hand, he were to wait until all items had been completed and then discuss total test performance, this would be considered summary feedback.

SPECIFIC VERSUS GENERAL FEEDBACK These are types of primary feedback and are analogous to augmented and summary secondary feedback. *Specific* feedback refers to very precise primary feedback, while *general* feedback refers to rather global primary feedback.

Perhaps the relationships between these different types of feedback measures can be illustrated best by the "Feedback Tree" shown in Figure 8.4.

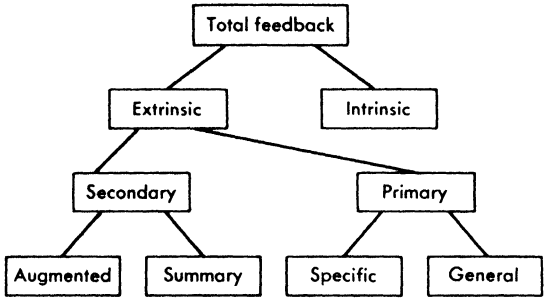


Figure 8.4. Diagrammatic "tree" showing the relationships between the various types of feedback.

SOME GENERAL FINDINGS CONCERNING KR Annett (1961) has cogently summarized the research dealing with the KR variable as follows:

1. There is little evidence to support the position that learning can occur without KR, although under certain conditions simple exposure and familiarization with the learning situation and materials can facilitate later learning.
2. "Positive" KR information seems to be a more effective procedure than "negative" KR.
3. The degree of specificity of KR and its relationship to the learning process does not appear to be linear. Indeed, there is some indication that it may take the shape shown in Figure 8.5.

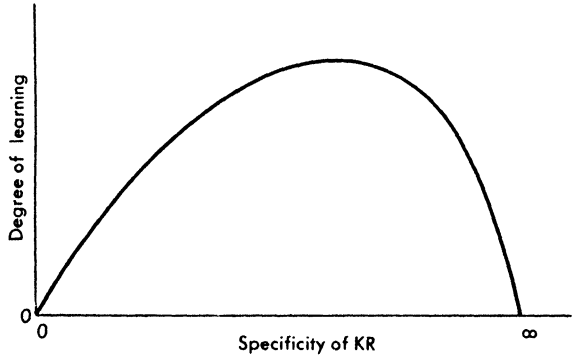


Figure 8.5. Hypothesized relationship between precision of knowledge of results and learning.

The curve in Figure 8.5 indicates that learning is facilitated by increased precision in feedback *up to a point*, but beyond this point learning is hindered with continued increases in precision. The explanation, of course, is that the trainee will reach a saturation point where the information given is just too much for him to handle, and he will have to spend time trying to simplify it in order to understand it. He becomes "overloaded," so to speak.

4. Delay time or lag time in KR seems to be generally related to learning *performance*, but the relationship is not a clear one. The research seems to demonstrate that lengthy delays between performance and knowledge of performance are detrimental to learning, perhaps due to lost information. However, at best the data are ambiguous.

ACCURACY VERSUS SPEED A problem related to the more general KR question concerns the kind of instruction or set given the trainee in a learning situation. All too often it is forgotten that the way in which the trainee is asked to view the task to be learned can itself have an exceedingly powerful effect on the learning process. As previously mentioned, one aspect of KR is its informational dimension—that is, the amount of knowledge it provides the trainee about his performance. KR only makes such information *available*, however, it does not guarantee that the trainee will pay any attention to it.

In most learning situations there are several kinds of KR information available, with the two most usual categories being some version of *quantity* and *quality* of performance. The instructions given to the trainee can have the effect of making him attend nearly exclusively to one of these. For example, if trainees are told that their *performance* task is to learn to produce as many units as possible, they may tend to ignore any KR information having to do with quality.

Actually, in most training situations the instructions given to the trainee are rather ambiguous. He is left on his own to determine which aspects of performance are the most critical and thus to which kinds of KR he will attend. How many times have we heard someone say to students "Do the very best you can!" or some other ambiguous instructional statement which really does not help the student a bit. It may be a well-intentioned expression, but it does not clarify what is meant by "best" performance.

Howell and Kreidler (1963, 1964) report several studies which relate directly to the problem of different types of instructional sets which can be given to a trainee. In their first study they gave each of four different groups a different instructional set, as follows:

Group 1: told to maximize speed
Group 2: told to maximize accuracy
Group 3: told to maximize both speed and accuracy ("contradictory" set)
Group 4: told to maximize amount of information processed[3] ("composite" set)

As might be expected, the speed group did best when judged by a speed criterion and the accuracy group did best when evaluated in terms of an accuracy criterion. When groups were evaluated in terms of a composite performance measure which took both speed and accuracy into account (amount of information processed), there were no differences between groups 2, 3, and 4, but the speed was noticeably poorer than the other three. The conclusion was that speed instructions are not as

[3] Formal definition of information in terms of "bits" was explained to each person.

desirable as other instructional sets unless speed alone is specifically desired. Unfortunately, in this study the feedback given to the groups corresponded to the instructions they had been given. Thus it was not possible to determine if instructions or feedback produced the effect. In their second study, which in general supported the first, this difficulty was not present.

SCHEDULES OF LEARNING

Probably one of the most well-established and well-documented learning principles is that distributed or spaced practice is superior to continuous or massed practice. This seems to be true for both simple laboratory tasks and for highly complex tasks. Actually, schedules of learning can be manipulated in three different ways: (1) duration of practice sessions, (2) duration of rest sessions, and (3) positioning of rest sessions.

Evidence seems to support the notion of short practice periods and moderate rest periods (McGeoch and Irion, 1952). Of course exact definitions of the terms "short" and "moderate" need to be established empirically for each individual task, as does the optimal positioning of the rest periods. However, it is usually much more effective to have short practice periods interrupted by short, frequent rest periods than to have only one or two long rest periods and one or two long practice periods.

TRANSFER OF TRAINING

Much of the research on the learning of different kinds of skills has been directed at the problem of transfer of training. There is a very practical reason for this. Very few training situations represent the actual job with perfect fidelity. Therefore, it becomes of real importance to understand something about the processes involved when one takes a skill learned in one setting and attempts to "transfer" this to a somewhat different skill in perhaps a somewhat different setting.

Transfer of training may be either positive or negative. *Positive transfer* is said to occur when something previously learned benefits performance or learning in a new situation. *Negative transfer* is said to occur when something previously learned hinders performance or learning in a new situation.

The explanation for transfer of training is based upon the concept of *identical elements*—that is, the greater the number of task elements held in common, the greater the transfer between the tasks. More specifically, transfer has been shown to be related to (1) similarity of the stimulus and (2) similarity of the response. For example, after learning to drive one make of automobile there is usually very little difficulty in learning to drive any other make. While one may find many minor variations in automobiles from make to make and even from year to year for a given make, the similarities far outnumber the dissimilarities. Therefore transfer should be (and usually is) quite highly positive. This is an example of both stimulus similarity (the dials, windows, and seats are located in approximately the same places) and response similarity (the brakes are located in the same place, the steering wheels work about the same, etc.).

However, sometimes even minor changes from car to car can cause some difficulty. Many of us have experienced the embarrassment in transferring a braking skill or habit acquired with mechanical brakes to a situation involving power brakes. For

even greater negative transfer, consider the American driver who visits England and who reports considerable difficulty in adjusting to a steering wheel on the other side of the car, plus driving on the left rather than the right side of the highway.

Naylor and Briggs (1961) have summarized some of the more important findings concerning transfer of training, particularly in terms of what other variables seem to influence the amount of transfer which can be expected to take place.

THE SHAPE OF THE TRANSFER FUNCTION OVER TIME In a series of interesting experiments, Bunch and his coworkers (Bunch, 1939; Bunch and Lang, 1939; Bunch and Rogers, 1936) investigated the amount of transfer, both positive and negative, as a general function of time. The curve for positive transfer appears to approximate quite closely the general curve for retention, at least up to periods of 120 days of no practice. Thus, the amount of positive transfer decreases as a function of time.

Research concerned with the amount of negative transfer, however, shows a very interesting trend. Again, the amount of transfer (negative) decreases markedly as the length of a no-practice interval is increased. However, beyond a certain point transfer does not disappear entirely but instead becomes positive, increasing in magnitude toward the longer intervals of no practice, only to finally decline gradually toward zero again. Because of this, Bunch (1939) has concluded that whether one habit is antagonistic to another is a function of the interval between the acquisition of the two habits.

TRANSFER AS A FUNCTION OF TASK SIMILARITY Baker, Wylie, and Gagné (1950) investigated the hypothesis that learning a new response to an old stimulus will lead to negative transfer. Using a basic tracking task, with the rate of cranking being the independent variable, they found positive transfer between rates in almost all cases. The amount of positive transfer was a function of response rate similarity, regardless of whether the initial task was faster or slower than the final task. Unfortunately, the no-practice interval was only two minutes in duration. Hauty (1953), in manipulating task similarity, found that the more identical the stimuli the more likely is negative transfer, while the more identical the response the more likely is positive transfer. Again, the no-practice interval was quite short. Defining task similarity in terms of the number of common stimulus-response relationships, Duncan (1953) found that all transfer was positive and directly related to the amount of task similarity within the context of a lever-positioning task.

TRANSFER AS A FUNCTION OF THE AMOUNT OF INITIAL LEARNING Britt (1934) obtained results which indicated that the amount of transfer of training was a function of the age of previous associations. However, his results could also be interpreted to mean the amount of transfer of training was directly related to the amount of initial learning on the initial task. In the study cited in the above section, Duncan (1953) also investigated transfer as a function of degree of original learning. Again, there was a direct relationship between amount of learning and amount of transfer.

TRANSFER AS A FUNCTION OF TASK DIFFICULTY Gibbs (1951), investigating the effects of task difficulty upon transfer as part of a larger experiment, found that transfer was about equal when the tasks were of equal difficulty, but he also found a definite transfer "bias" as the tasks became more dissimilar in terms of difficulty. The results indicated that there was greater transfer from a difficult task to an easier task than vice versa. Day (1956), in a review of the research in motor skills pertaining to the transfer effect as a function of task difficulty, broke down the

experiments into three classes depending upon what was modified to affect the difficulty level. He concluded that those experiments which regulated difficulty by varying the stimulus (target size, target speed, number of targets, etc.) have in general obtained negative results regarding differential transfer effects as a function of task difficulty. However, those studies which have varied difficulty in terms of response variations (strength of response, speed of response, number of responses required, etc.) seem to indicate that greater transfer is obtained by initial training on a more difficult task than on an easier task. Those studies which manipulate task difficulty in terms of response compatibility also tend to indicate that better transfer is obtained from the hard to easy sequence than from the easy to hard sequence.

TRANSFER AS A FUNCTION OF KNOWLEDGE OF RESULTS Briggs, Fitts, and Bahrick (1957) report results from a study designed to investigate the influence or knowledge of results (as affected by visual noise) upon learning and transfer in a complex tracking task. They obtained marked differences between groups during the acquisition trials, with the better performance being associated with the lower noise levels. However, there were no marked differences between group scores on the transfer trials. The authors concluded that the results support the position that knowledge of results has an effect on performance but does not affect learning.

PART VERSUS WHOLE TRAINING

The accepted goal of training is usually the achievement of a desired level of competence in the performance of a particular task or job. In addition, it is usually required that this training be accomplished as rapidly and as efficiently as possible. In seeking the best possible training procedures to accomplish these objectives, the question of the size of the unit to be learned becomes one of extreme interest to many individuals. Should one try to teach the entire task, or is it more efficient in the long run to teach only individual segments of the task initially, and then at some later point in training begin the process of combining the already learned subtasks? This question became labeled the problem of part versus whole learning, and in spite of its lengthy history (the first part—whole research study—was done in Europe in 1900) an unequivocal solution still does not exist.

TYPES OF TRAINING SCHEDULES There is a wide variety of different strategies which can be adopted in scheduling the learning of a task. However, they are all versions of one of three primary strategies—*pure part*, *progressive part*, and *whole* training. Suppose one has a task which can be conveniently divided into three distinct parts of subtasks, A, B, and C. The three different training procedures would proceed as shown:

	Phase I	*Phase II*	*Phase III*	*Phase IV*
Whole Training	A + B + C	A + B + C	A + B + C	A + B + C
Pure-Part Training	A	B	C	A + B + C
Progressive-Part Training	A	A + B	A + B + C	A + B + C

NAYLOR HYPOTHESIS One of the major frustrations concerning the different training schedules was that different researchers would get contradictory results when trying to establish which was superior. For example, Seagoe (1936) examined all studies on part versus whole training and found no interpretable pattern (see Figure 8.6).

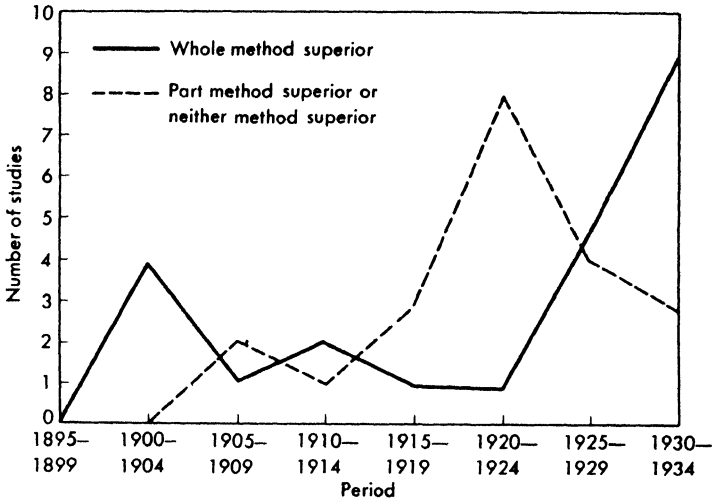


Figure 8.6. Comparative frequency of studies favoring whole methods for period 1895–1934. (Adapted from M. V. Seagoe. Qualitative wholes: A revaluation of the whole-part problem. *Journal of Educational Psychology*, American Psychological Association, 1936, 27, 537–545.)

Disturbed by this continued inconsistency, Naylor (1962) carefully examined all part versus whole training studies since 1930 and found that the outcome of the research seemed related to the characteristics of the task which the trainees were attempting to master. Table 8.1 shows his findings.

Table 8.1 *Percent of Post-1930 Studies Finding Whole or Part Method Superior as a Function of Task Complexity and Organization*

		Task Organization	
	Task Complexity	*High*	*Low*
Whole	High	100	0
	Medium	50	50
	Low	25	50
Part	High	0	25
	Medium	19	50
	Low	63	50
Inconclusive	High	0	75
	Medium	31	0
	Low	12	0

Before examining Table 8.1 it is necessary to define the terms *task organization* and *task complexity*. In his taxonomy Naylor proposed that nearly all tasks could be considered as being made up of several subtasks. Thus, driving a car involves at a minimum manipulation of the steering wheel and manipulation of the accelerator —both subtasks involving control over different types of vehicle movement. Task organization refers to the extent to which these subtasks are interrelated, that is,

are dependent upon or influence each other. In the case of driving a car, one usually adjusts his forward velocity, either by braking or letting up on the accelerator, when he approaches a sharp turn. Thus the action made in steering has an effect upon the action taken with respect to the accelerator. The two subtasks are not so completely independent of one another that you can perform one with complete disregard for the other.

Task complexity refers to the difficulty of each of the separate task components viewed individually. The demands placed upon a driver in terms of steering a car when involved in rush-hour traffic are not usually very great (assuming he stays in his lane). Most of the time the driver is primarily concerned with his forward motion. That is, he spends the majority of his attention and energy in stopping, starting, slowing down, speeding up, etc. Contrast this with driving on an open turnpike at a constant speed, where hardly any demands upon the forward motion subtask are involved. Certainly the reader can also recall situations where he has been forced to drive on a very curved but level piece of highway. Here almost all the task demand is being placed upon the steering component and very little upon forward control.

Naylor then proposed that the total difficulty of any task could be explained in terms of these two basic task characteristics

$$\text{Task difficulty} = \text{Task complexity} \times \text{Task organization}$$

Table 8.1 shows that there indeed appears to be a systematic relationship between which training methodology proved to be superior and the task characteristics of complexity and organization. Based upon the pattern shown in Table 8.1, Naylor suggested the following training principles:

Principle 1: Given a task of relatively high organization, as task complexity is increased whole task training should become relatively more efficient than the part task methods.

Principle 2: Given a task of relatively low organization, an increase in task complexity should result in part methods becoming relatively superior to whole task training.

Studies Supporting Hypothesis: There are several studies which now support, either directly or by inference, the Naylor hypothesis. In a pair of studies, Naylor and Briggs (Briggs & Naylor, 1962; Naylor & Briggs, 1963) investigated first the task complexity variable and then task complexity and task organization together in terms of their effort on the part-whole controversy. The initial study was somewhat ambiguous, but when interpreted in terms of the more comprehensive second study the data in both seemed to uphold the principles listed above. Bilodeau (1954, 1955, and 1957) has also presented evidence which supports, in particular, the notion of component interaction as being important to the part versus whole training problem.

FRACTIONATION VERSUS SIMPLIFICATION. An interesting training question raised by Briggs and Waters (1958) is whether the best training procedure utilizes some form of part practice (a process they call task "fractionation") or instead use a degraded version of the whole task (a process they call task "simplification"). Both procedures, of course, call for training the person on something "less" than the final task he is trying to learn to perform. Fractionation, the first strategy, simply divides the actual task into parts to be learned individually; simplification involves whole training, but on an easier or less demanding version of the task. Their research indicated

that (1) the number of task dimensions in the training task should duplicate the number of dimensions present in the actual task and (2) simplification rather than fractionation appeared to be a preferable training procedure.

The findings of Briggs and Waters would seem to speak well for all "simulated" training situations. It should be pointed out, however, that it may be possible to combine both part training and simplified training in certain situations so that training efficiency is made optimal.

MOTIVATION

The role of motivation is very important in learning; little if any progress occurs without it. Whereas students often spend three inefficient years in school attempting to learn a foreign language, the recent experiences of the military, the Peace Corps, and the foreign service have demonstrated that the average person can be trained to conduct a conversation in a foreign language in a matter of months. Only part of this vast difference in effectiveness is attributable to better teaching techniques; most of it is due to the increased motivation of the individuals. The learner in those situations understands the reason for studying a foreign language much more readily than the high school student; the latter was probably convinced that it was just a useless requirement imposed on him by a "bunch of old fossils."

Effective motivation is the essence of learning. Unless the person who is teaching a group takes pains to ensure correct motivation, little learning is likely to occur. A man who has been promised a promotion provided he takes a specific course will learn its content in a comparatively short time. A girl who hopes to obtain a secretarial job will learn to type and take dictation in as little as six weeks. Without the specific motivation of a job, both these individuals may dawdle along for years. Many college students insist that they learned the entire content of a course in a few nights. In the "cram" session they discover that there is much useful material that they should have learned during the course, but the motive for learning was too remote then. When passing the final examination is their only motive for learning, they learn.

SOME COMMENTS ON TRADITIONAL LEARNING PRINCIPLES

It was mentioned earlier that a dogged adherence to the traditional principles might *not* be the best way to approach the training problem. This has been eloquently argued by Gagné (1962) who cites numerous examples where research in military training situations has demonstrated that these principles were totally inadequate. Indeed, in some cases they turned out to be hindrances rather than aids to the training process.

Gagné argues for a somewhat different set of psychological principles for training (1962, p. 88):

> 1. Any human task may be analyzed into a set of component tasks which are quite distinct from each other in terms of the experimental operations needed to produce them.
> 2. These task components are mediators of the final task performance; that is, their presence insures positive transfer to a final performance, and their absence reduces such transfer to near zero.
> 3. The basic principles of training design consist of: (a) identifying the component tasks of a final performance; (b) insuring that each of these component tasks is fully achieved; and (c) arranging the total learning situation in a sequence which will insure optimal mediational effects from one component to another.

It is interesting to note that Gagné's first principle deals with the importance of task analysis into separate subtasks or components and is thus very similar to Naylor's point of view (discussed earlier under part and whole training).

SOME TRAINING GUIDELINES

Rather than talking about learning principles, it seems more desirable simply to set forth a series of suggested guidelines which, when followed or considered, may help to make most training situations more effective.

To be practical, when one is involved in the training of others, it is best to recognize that a number of fundamental considerations are necessary. The following check list of ten items can often serve to enhance the training process. Decisions are typically required in connection with each, and the difficulty is that too often facts and data are lacking which would determine the best manner of organizing the training course. This would simply indicate that research on the training process is as necessary as training itself.

A good start, however, involves consideration of the following items:

1. Motivation is not only a desirable but often also a necessary accompaniment of learning.
2. The number of units or lessons to teach most effectively must be a considered judgment; too often it is arbitrarily and artificially set.
3. The amount to be learned in any one unit must be planned. For optimal learning, the unit should be neither too large and complex nor too simple.
4. Any training is practically never comprehensive or exhaustive. The amount to be learned has to be related to the desired job performance. For example, a person need not know how a motor operates in order to drive a car.
5. The task to be performed should be demonstrated.
6. The demonstration must be immediately followed by the doing on the part of the learner.
7. A discussion and question period should follow the doing, to clear up any misconceptions on the part of the learner between the explanation of why and the demonstration of how.
8. Ample and adequate practice opportunity should be encouraged. Some learners tend to overestimate their performance and erroneously and prematurely believe that the task has really been learned.
9. Observable progress during practice goes a long way toward encouraging a sufficient degree of practice. Plotting the learning curve wherever practical is to be encouraged.
10. A summary and review of the entire learning process should be made by the learner, with the teacher available for last-minute pointers and for establishing that the task has been learned according to performance that meets the criterion or standard.

EXPERIENCE VERSUS TRAINING

Industrial workers are likely to regard with disfavor the inroads made by education. Many experienced workers seem to be members of a "secret society," with mysterious knowledge all their own which has come to them only after long years at a workbench. We often hear an experienced worker or a successful boss telling a young college graduate or a person who has studied the specific task in school

that "this education business" is nonsense. The argument runs somewhat as follows: "Look where I am, and I didn't have any education." How much of this is discrimination, or fear of young people, or resistance to modern trends, and how much of it is sour grapes is hard to tell. On the other hand, people with formal training sometimes tend to belittle those who have gained their knowledge by experience. This merely pours salt in the wound. A young person with training but with little or no experience should use judgment and curb his missionary zeal by not insisting that he must train all those who have not had his training. This is just as ridiculous as the experienced employee's objection to anyone who has been trained but does not have the experience.

Training cannot entirely take the place of experience, but when training is effective it can shorten the time required to reach maximum production. In fact, the training of new people is often the only means of introducing new and efficient work methods as a substitute for older and less efficient ways of doing a job. Experience can be a good means of training, but this result is sometimes accidental. Experience that depends solely on trial and error or on the prodding of a poor teacher means trouble and a much longer uphill struggle to achieve competence. However, experience and training are not incompatible as methods of learning; in the proper proportion, both together lead to the highest efficiency and maximum achievement in the shortest time.

Training, like education, can be regarded as a continuing process of life. This means that refresher courses are a "must" in any comprehensive training program. Of the many reasons for periodically retraining experienced employees, the most pressing one is the appearance, slowly but surely, of strange and inefficient work methods which interfere with the efficient methods learned originally. In addition to ironing out the kinks and removing the "bugs," refresher courses permit the continual introduction of new and improved work methods; they also serve to bolster morale since the individual employee is assured that the company is still aware of his existence.

KINDS OF TRAINING

Over the years a number of different types of training have emerged in industrial situations. For example, Tiffin and McCormick (1958) classify training into the following seven categories:

1. *Orientation training:* to acquaint new employees with the company
2. *Vestibule training:* simulated on-the-job training
3. *On-the-job training:* actual practice on the job
4. *Apprentice training:* formal apprenticeship for a craft
5. *Technical training:* training in technical aspects of job
6. *Supervisory training:* training in management
7. *Other training:* any unusual situation not included in the above

Actually, any classification of different types of training situations is rather difficult since there are a number of different ways that they might be classified. For example,

a more meaningful system might be to classify training experiences in terms of (1) purpose, (2) location, (3) trainee, and (4) technique. Such a classification is given below:

I. *Purpose of training experience*
   A. To impart knowledge (e.g., orientation training)
   B. To teach skills (e.g., technical training)
   C. To modify attitudes
II. *Location of training experiences*
   A. Actual work situation (on-the-job training)
   B. Simulated work situation (e.g., vestibule training)
   C. Classroom
   D. Unspecified (e.g., home study)
III. *Trainee characteristics*
   A. Experience
      1. No prior experience (e.g., apprentice training)
      2. Moderate prior experience ("refresher" training)
      3. Extensive prior experience
   B. Job Level
      1. Nonsupervisory employees
      2. Supervisory employees
   C. Number
      1. Individuals
      2. Groups
IV. *Technique used*
   A. Lecture (formal)
   B. Lecture (discussion)
   C. Work sample procedure
   D. Role playing
   E. Television (closed circuit)
   F. Programed instruction
   G. Group discussion

These categories are only a guideline which can be used for purposes of convenience. Any training experience is apt to have several purposes and may use a variety of different techniques during its program.

SYSTEMS APPROACH TO TRAINING

There is a current emphasis in modern industrial psychology to view many of the traditional problems in a somewhat broader context than before. Problems such as training are considered not only in terms of training objectives per se, but also in terms of the objectives and goals of the total organization or "system" in which the individual will be performing his task. This broader view is sometimes referred to as a *systems approach*. The name implies that one must be concerned with objectives of the total system rather than objectives of any component within the system.

Eckstrand (1964, pp. 1–2) has strongly advocated a systems approach to training technology:

> In this approach, the development of a training program is likened to the development of a weapon system. Here the systems engineer begins with an operational requirement: a precise statement of the objective to be achieved by the system. The system engineer then works backward from these objectives to produce an arrangement of subsystems which, when operated according to some operational plan, will fulfill the objectives. The process ends with a series of tests to assure that the design achieved does in fact fulfill the requirement The design of a training system can proceed in the same manner. The behavior . . . men must exhibit on the job becomes the objective which must be achieved by the training system The job of the training designer, then, is to select and sequence a series of learning experiences which will produce the desired behavior. A testing [evaluation] phase is required to assure that the training program designed succeeds in producing men capable of performing as specified.

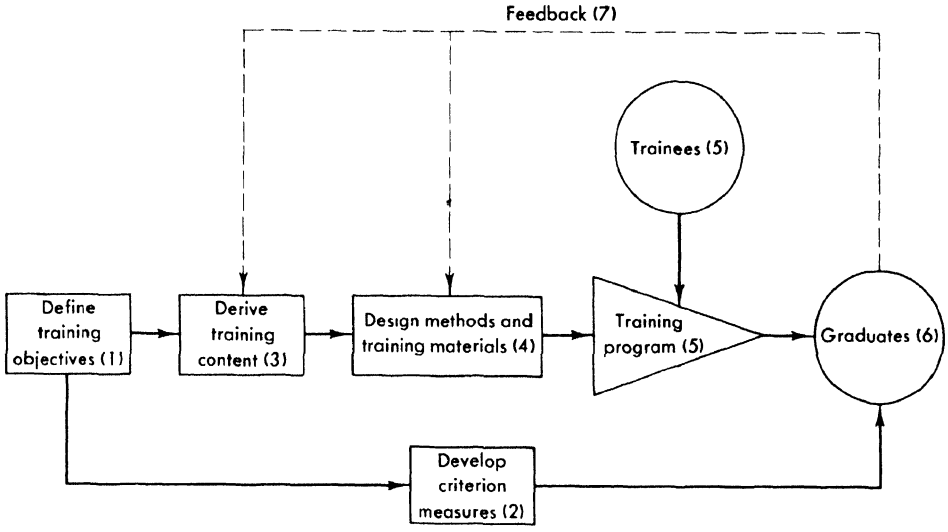


Figure 8.7. A systems approach to training. (From G A. Eckstrand. Current status of the technology of training. AMRL *Document Technical Report* 64–86, September, 1964, 3.)

The application of the systems approach to training can probably best be understood by examining Figure 8.7, taken from Eckstrand. It involves seven major steps:

Step 1. Define training objectives.
Step 2. Develop criterion measures for evaluating training.
Step 3. Derive training content.
Step 4. Design methods and training materials.
Step 5. Integrate training program and trainees.
Step 6. Compare graduates to criteria standards set in step 2.
Step 7. Modify steps 3 and 4 based upon results of step 6.

The advantage of the systems approach is that it helps the psychologist or training director in making his decisions. He is able to understand and keep in mind the total training process as he tries to accomplish his objectives.

SPECIAL TRAINING AREAS

Two of the most important, interesting, and challenging areas in training today are *management (executive) training* and *team training*. Because of their growing importance, we shall examine each of these in somewhat more detail.

EXECUTIVE TRAINING (LEADERSHIP TRAINING)

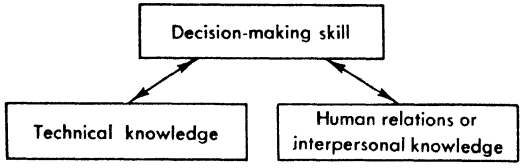
The term executive training may have several meanings, depending upon the goals and objectives of the training situation. Most executive training situations may be classified as belonging to one or both of two subtypes of training experiences: (1) human relations training and (2) decision training. Since they are qualitatively different, it is necessary to examine each separately.

Another term which is often used in place of executive training is *leadership training*. Actually, one might logically defend the position that leadership is only one aspect of executive or managerial ability. However, the distinction has never really been made clear by those most interested in the problem, and the terms *leadership training* and *executive training* or *executive development* have come to mean much the same sort of thing.

Chapter 14 is devoted exclusively to examining the problem of leadership; thus we will only touch briefly on it here from the training point of view. That is, we shall assume we can agree upon the appropriate things that we want to "train into" our people to make them good leaders. Thus our only problem here is what training methods are the best to use.

We have already said that leadership training has two separate components associated with it—a technical skill component and a human relations component. Let us examine these further.

The essence of any managerial position is decision making. Leaders, managers, executives (or whatever name they are given) exist primarily to make decisions. *This cannot be overemphasized!* If you strip a leader of decision-making prerogatives, you will no longer have a leader. This automatically means that to develop good leaders one must of necessity be concerned with developing his decision-making skills. These skills, then, may be viewed as shown:



TECHNICAL KNOWLEDGE TRAINING How does one train an executive in the technical aspects of his job (decision making)? Obviously there are a variety of methods which might be used.

1. *On-the-job training*—"breaking a new man in" by having him "understudy" a successful job incumbent.
2. *Simulated training*—having a man learn by giving him examples of typical situations to deal with and to study. There are a number of versions of this type of training, such as (a) case study method, (b) in-basket testing, and (c) management gaming.

3. *Classroom instruction*—using standard lecture and discussion techniques for training on the technical aspects of the job. For example, most industrial management programs are examples of how modern management principles are brought to the classroom.

HUMAN RELATIONS TRAINING Modern management theory accepts without question the notion that interpersonal relationships between a supervisor and his subordinates are an important component in the managerial decision process. Thus, optimizing the decision process must involve training in the understanding of the interpersonal relationships critical to the process. Such training is generally referred to as *human relations training.*

As in the case of technical knowledge training, there are a number of different training procedures which can be used to teach human relations skills.

1. *Simulated Situations* Almost any situation which is set up to simulate the technical aspects of the decision-making job can also be used (perhaps with minor modification) as a training device in human relations Two excellent examples are the methods of role playing and the in-basket technique.
2. *Group Discussion.* In the case of simulated training, the situation in which the trainee operates is a rather well-structured one. A different procedure is to form what are called "unstructured" groups for purposes of discussing problems of interpersonal relations among individuals. Examples of these are the "T" groups (training groups) and the *sensitivity training* methods.

Perhaps it would be helpful at this point to examine in somewhat greater detail each of the distinct training methods we have mentioned in order to give a better indication of their similarities and differences, particularly since we have implied that certain training methods may be used to teach both the technical and the human relations aspects of management.

IN-BASKET TRAINING This procedure has been mentioned earlier. It has many appealing features as a training device for both the human relations and the technical aspects of a managerial position. One of its major advantages is that quantitative indices of performance are possible with this tool which are not available with most other methods.

ROLE PLAYING The technique of role playing differs from the in-basket in that usually a number of individuals are given assigned roles to play and then requested to work out a solution to a given problem. Thus role playing permits practicing human relations skills in the relative safety of a laboratory environment. Role playing is *not* acting. Persons are asked to play the roles they have been assigned by "putting themselves into" these hypothetical situations. This permits the interactions between players to be representative of their own personalities. The roles and problem are structured only to facilitate the kinds of interpersonal relationships that are to be studied.[4]

MANAGEMENT GAMES Business or management games are a fairly recent phenomenon in executive training, although the use of games as a method of training decision makers has a lengthy history in the military; the first known war games were invented by Weikhmann in 1664. Games have always seemed ideally suited for teaching the strategy of move and countermove so frequently found in both land and sea en-

[4] For an excellent review on role playing see N. R. F. Maier. *Principles of human relations.* Wiley, New York, 1952.

gagements. Most games, of course, represent conflict situations, and in most games the consequences of a decision will depend partially upon the actions of the adversary.

Business games have become popular and practical with the tremendous growth in computer technology. Almost all modern business games are "computerized." The typical business game will be built around a major industry, a particular business firm, or perhaps only a particular aspect of a firm. To quote from Carter (1965, pp. 419–420):

> In typical games the designer makes assumptions regarding the influence on the business enterprise of such factors as price of the products, costs and capabilities of production, advertising and selling costs and results, research and development costs and results, investment costs and effects of plant and equipment, and dividend rates. Many of these functions are invariantly defined while others have randomness built into their effect. For instance, research and development usually produce product improvement as a function of the amount of money invested, but with the effectiveness of the outcome modified by random factors.
>
> The players are usually formed into groups of four or five without any organizational structure being defined. Often teams compete against each other, each team representing a business attempting to maximize its goals in a competitive market At the beginning of play each team is given identical consolidated reports of profit and loss, statement of financial condition, plant capacity, etc. Players are allowed to spend funds for plant expansion, research and development, advertising, market research, and similar items. The outcome of such expenditures is usually not known to the players, but has been programmed into the enterprise model used in playing the game. Play is in non-real time with a few hours or a day's play representing a calendar quarter or six months' real time. At the end of each time period a computer calculates the results of a period's operations and appropriate reports are furnished each team. On the basis of these reports and the players' records of their actions, they proceed to plan and execute their operations for the next time period and so on until five to ten years' operations have been completed At the end of a play a critique of the play and the decision made is conducted, with discussion of factors which might have improved the players' performance.

Business games are primarily used for the teaching of modern management marketing principles as opposed to human relations skills.

SENSITIVITY TRAINING Earlier we said that role playing involved having persons interact on a structured problem in predefined roles. If one should remove both the problem and the specified roles, about all that remains is an unstructured group of people with a common interest in human relations! This is the essence of what has come to be called *sensitivity training*. Historically, the sensitivity approach to training has grown from the training methods developed by the National Training Laboratory in Group Development and from certain schools of group psychotherapy and nondirective counseling. To quote from Tannenbaum, Weschler, and Massarik (1961, p. 132):

> While the characteristics of the sensitivity training process vary somewhat the educational method used is primarily trainee-centered. Trainees are permitted a maximum amount of leeway of action, and their interactions subsequently provide the subject for discussion and analysis. During the training experience, it is the process (the "how") rather than the content (the "what") that receives most attention. Trainees are encouraged to deal with their feelings about themselves and about others and to explore the impact they have given each other. Thus, they examine feelings, expressions, gestures, and subtle behaviors which in everyday life often are taken for granted.

> Sensitivity training is still experimental. As carried out in university workshops, group development laboratories and plant settings, no formula or "cookbook" for conducting this type of training has been nor is it likely to be developed. In general, however, certain essential elements have emerged.

The "essential elements" mentioned above are the following:

1. Sensitivity training is "process oriented" rather than "content oriented."
2. The training is unstructured in terms of method and content.
3. Frustration appears essential to the success of training.
4. The training is optimal in small groups
5. The atmosphere must be "permissive."

SUMMARY OF METHODS Table 8.2 summarizes the different executive training methods and illustrates their relative effectiveness.

EXAMPLES OF EXECUTIVE TRAINING Before leaving the topic of management training it would probably be helpful to cite some examples of research. Research efforts have generally been rather sparse, perhaps because of the difficulties surrounding such endeavors which were mentioned earlier in the chapter. First we will look at a study designed to teach decision skills using a game which abstracted the basic decision parameters inherent in the real task, and second we shall look at an example of the use of role playing as a training device in human relations.

TABLE 8.2 *Different Methods of Executive Training and Their Relative Effectiveness for Teaching Different Executive Skills*

Method	*Technical Decision Skills*	*Human Relations Decision Skills*
In-Basket Training	Yes	Yes
Sensitivity Training	No	Yes
Gaming	Yes	No
Role Playing	Yes	Yes

Kinkade and Kidd Study This study was one which repeated an attempt to use a special purpose "game" as a method of training for a complex decision-making task. The task to be learned was that of an air traffic controller, a very demanding and stressful decision-making task. Air traffic controllers are in a continuous management role, directing pilots into their approaches, etc. Kinkade and Kidd (1962) designed a game similar to Chinese checkers which seemed to embody many of the kinds of decisions, at least in an abstract sense, which are required of air traffic controllers. One group of trainees was given six hours of practice on the game prior to going into training on the air traffic control task. The other group went directly into air traffic control training. Figures 8.8 and 8.9 show the results.

The effect of the training on the game is very clear. Both the amount of delay and the number of errors made were reduced by means of this technique. The study is an excellent example of the effectiveness a game can have in learning the technical aspects of a complex management task.

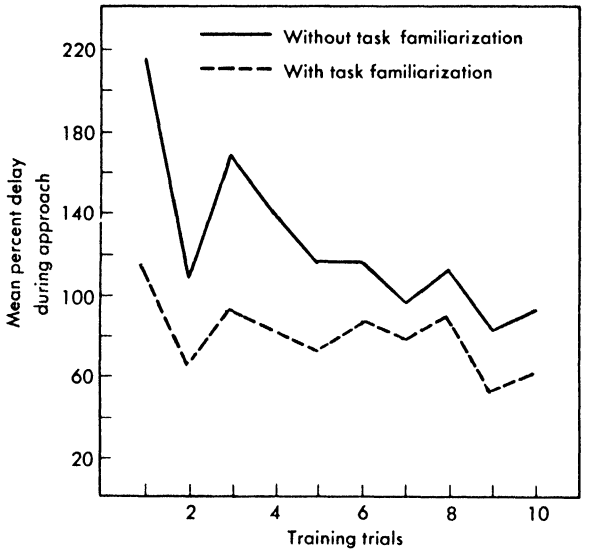


Figure 8.8. Comparative performance on training trials on the air traffic control task. (From R. G. Kinkade and J. S. Kidd. The use of an operational game as a method of task familiarization. *Journal of Applied Psychology*, 1962, *46*, 1–5.)

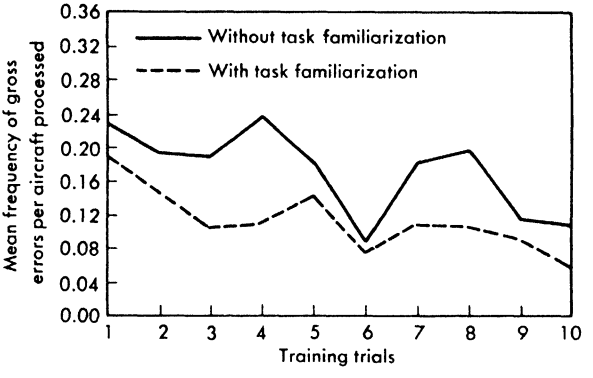


Figure 8.9. Comparative performance with respect to gross errors on the air traffic control task. (From R. G. Kinkade and J. S. Kidd. The use of an operational game as a method of task familiarization. *Journal of Applied Psychology*, 1962, *46*, 1–5.)

Harris and Fleishman Study This research was designed to evaluate the effect of a human relations supervisory training program in terms of the attitudes and behaviors of the trainees after they had returned for the actual work situations. It was one of the first such studies to follow up the effects of supervisory training. Most prior research had usually followed the procedure of evaluating the training immediately upon completion of the program.

Supervisory behavior was evaluated by Harris and Fleishman (1955) in terms of two criteria of performance as measured by a questionnaire filled out by subordinates (that is, ratings of the supervisors by those who worked under them served as the dependent variable). The two scores obtained from the questionnaire were *Initiating Structure* and *Consideration* (for a more detailed explanation of these basic dimensions of leadership the reader is referred to Chapter 14). At least three workers were drawn randomly from each of the work groups of 98 different foremen at two time periods 11 months apart. During the 11-month interval between the two evaluations, 39 foremen had attended the company's central school which administered the Super-

visory Training Program. The remaining 59 had received no human relations training during the same period. Table 8.3 shows the mean ratings for both groups on both criteria at the first and second evaluations.

TABLE 8.3 *Comparison of Supervisory Behavior Description Scores for Foremen With and Without Intervening Training*

Group	Leadership Dimension	Before Training		After Training	
		Mean	*SD*	*Mean*	*SD*
With intervening training ($N = 39$)	Consideration	70.6	13.8	72.1	13.0
	Initiating Structure	43.2	6.4	40.8	7.3
Without intervening training ($N = 59$)	Consideration	73 8	15.0	74.4	12.6
	Initiating Structure	41.5	7.1	39.6	6.1

SOURCE: E. F. Harris and E. A. Fleishman. Human relations training and the stability of leadership patterns. *Journal of Applied Psychology*, 1955, *39*, 20–25.

No significant differences were found in the mean scores made by either group before and after the training period. Also, and more important, no significant changes in mean scores occurred within either group, whether or not they had received supervisory training. Harris and Fleishman also found much the same sort of noneffect when they examined the influence of a refresher course upon supervisory behavior. The moral of the story would seem to be that the effects of human relations training may be only temporary at best, and exceeding care should therefore be taken before investing greatly in such a training program.

TEAM TRAINING

One of the results of modern-day technology seems to be an increased frequency in tasks which involve a team effort rather than an individual one. References are frequently made to "research teams," "management teams," "missile launch teams," "air defense teams," etc. A *team* is a group of individuals who are working together toward a common goal. All teams are really defined by their purpose or goal. Indeed, if two team members have different goals or objectives, the performance of the team is apt to become less efficient. Consider the basketball player who is more concerned with his own point production rather than whether the team wins or not.

Associated with the increased frequency of team tasks has been a related increase in the importance of team training within the larger training complex.

Is the team training problem any different from the usual training problem, or can one take the guidelines from the simpler learning situation and apply them to this more complex setting? The answer seems to be "perhaps." However, when one talks about such things as part or whole learning, feedback, etc., they become defined somewhat differently in a team training situation. Some of the more critical team training problems which researchers have tried to answer are:

1. Should team members be trained as individuals or as a unit?
2. What kinds of feedback are most appropriate to team members—individual or team?
3. Can team performance be predicted from the proficiency of individual team members?
4. What influence does a change in team membership have upon training?
5. What types of team structures appear best for learning different kinds of team tasks?
6. What kinds of proficiency measures should be used in evaluating teams?

GROUP VERSUS INDIVIDUAL TRAINING OF TEAM MEMBERS One of the more interesting questions relevant to team training is whether team members should be trained as a unit or whether individual training would be more efficient. Intuitively, it might be concluded that any task requiring several individuals for its successful completion would probably be acquired more efficiently through team training. Unfortunately, the scanty research currently available on the topic seems to be rather ambiguous. Horrocks, Krug, and Heermann (1960) found individual training superior, while several other studies found individual unit training most efficient (Glaser, Klaus, and Egerman, 1962; Egerman, Klaus, and Glaser, 1962). Little else has been done to examine this question in further detail, except for a recent effort by Naylor and Briggs (1965) which explored the effect of task complexity and organization upon team efficiency. Most important, there has been no attempt to provide any integrated approach to the question of individual versus team training.

There is, however, one aspect to the organized-unorganized dimension of teams explored by Naylor and Briggs in their 1965 team-training study which is worth further consideration, and that is with reference to its counterpart in the individual task domain. Earlier in the chapter we discussed the problem of part versus whole training when training a single person on a multidimensional task. We saw evidence that the relative effectiveness of the two approaches to training seemed to be a function of the kind of task for which one was trying to develop skill. In particular, evidence was given in favor of the Naylor hypothesis which said that part or whole training depended upon *task complexity* and *task organization*.

To the extent that task organization in an individual task represents a dimension identical to the team organization concept, data obtained from part versus whole research should prove to be a rich source of information for developing potential hypotheses about multi-man (team) tasks. Consider, for example, the Naylor and Briggs part-whole studies (1962–1963), which dealt with the advantages of whole tasks versus part task training for *individuals* as a function of task organization and complexity. One might view these studies as being analogous to a multi-man (team) training situation: The training on individual task dimensions is similar to individual training of team members, and the whole task training of Naylor and Briggs was similar to "team" training in a multi-man task. Naylor and Briggs manipulated task organization in their study by varying the amount of information which tied together two task dimensions. This definition of organization would certainly seem compatible to that typically used in a team situation of defining organization in terms of communication or information channels available between team members.

The 1963 Naylor and Briggs study indicated that for tasks of high organization, whole training appears to be most efficient regardless of task complexity; for tasks of relatively low (moderate) organization, whole training is best for tasks of low complexity but part training is best if the task is quite complex. Carrying this interaction over into the multi-man situation, it could therefore be hypothesized that for those tasks which are highly organized and which require a great deal of communication and cooperation between members, it would be best to employ team training. With tasks which only place low or moderate communication demands on team members, team training would be best if the subtasks are fairly simple, but individual training would be best if the subtasks are quite complex.

Some evidence for the above hypothesis can be found in the study by Kinkade and Kidd (1962) cited earlier, in which they investigated the effect of team communication upon decision-making performance using a highly abstract version of an air traffic control system. Their basic task situation was one of moderate organization and

low complexity, which would suggest that team training should be better than individual training. Indeed, the results indicated that performance for teams working under organized (whole) conditions was superior to those under the less organized (part) conditions. Additional support for the hypothesis may be seen in the data obtained from a study by Naylor and Briggs (1965) in which they showed that team performance in a radar-controlled aerial intercept task was influenced by the way the task was organized: Superior performance was obtained with teams in which the members worked independently of one another compared to performance in an organization which *encouraged* interaction (verbal communications, trading targets and interceptors, etc.) between radar controllers. Since the task involved was either of moderate (interaction encouraged but not required) or low organization and of either medium or high complexity, the results would appear to fit within the general framework explained above.

Largely untested, however, is the validity of the notion that individual training for tasks of low organization is more and more efficient as the tasks of each individual in the team is made more complex, and group training is more and more efficient for tasks of high organization as the tasks of each individual are made more complex.

FEEDBACK IN TEAM TRAINING The feedback or knowledge-of-results training variable becomes very complex and powerful in a team training situation. As Rosenberg and Hall (1958) have pointed out, the way in which feedback is handled in the team situation may often be the key to the actual communication structure of the team.

Of course, the primary difficulty in a team situation is whether you provide a team member with information about (1) his own performance only, (2) team performance only, or (3) both his own and team performance.

Knowledge of results about one's own performance may tend to produce behavior which is *self* rather than *team* oriented. The team member may become too enamoured with his own skill and not worry about the success of the team effort. On the other hand, there are certain problems involved in only providing team members with information about team performance, particularly when team performance is some composite of the performance of all team members. The reason for the difficulty is that under team feedback conditions an individual team member gets very little direct information about his own performance. Since we established earlier that KR is an essential ingredient to learning, it may be very hard for team members to learn under team feedback conditions.

Rosenberg and Hall have conducted several studies on this problem of feedback (1958). They have found individual performance to be tremendously affected by the kind of KR team members are given. Under conditions of team feedback (they used the term "confounded" feedback) the team members seemed to develop compensatory behavior. That is, if one team member was an "over responder," then the other team member would become an "under responder." The interesting finding was that confounded feedback did not result in less accurate performance as compared to an individual feedback condition.[5]

PREDICTION OF TEAM PERFORMANCE FROM INDIVIDUAL PERFORMANCE A rather interesting study of team proficiency is reported by Wiest, Porter, and Ghiselli (1961). They found that team performance was less than the sum of individual performances.

[5] For a more extensive discussion of the entire team training problem refer to M. Glanzer. Experimental study of team training and team functioning. In R. Glaser (ed.). *Training research and education.* Wiley, New York, 1965.

They also found that team performance was better predicted by knowing how good the most proficient team member was than by knowing how good the poorest member was. Also of importance was the discovery that the more similar two team members were in their individual proficiency, the more likely they were to form a proficient and effective team.

INFLUENCE OF CHANGES IN TEAM MEMBERSHIP UPON TRAINING Teams are often apt to be dynamic entities in terms of their composition. Membership in a team may not be a stable characteristic, particularly with large teams. Because of this instability of group membership, it is important that we know something about the effect of changes in team personnel upon the training process.

Horrocks, Heermann, and Krug (1961) found no effect on team performance when changes in membership of the team were made during training. However, their team task was a rather simple decoding requirement. Naylor and Briggs (1965) did find an effect on performance in a somewhat more complex task. In their study they introduced during training a new team member as a substitute for a regular team member who was either more or less skillful than the member being replaced. The more highly proficient substitute enabled the teams to improve significantly during a subsequent training session, while teams receiving a replacement with less proficiency than the man he replaced were found to deteriorate slightly in the subsequent session. However, both effects were quite temporary and disappeared with continued training. It would seem, then, that one could question the desirability of extra training for a replacement when on-the-job training can so quickly bring total team performance up to par following the stress of a replacement of one of the team members. Similar findings have been reported by Glaser, Klaus, and Egerman (1962) and Egerman, Klaus, and Glaser (1962).

TEAM STRUCTURE AND TRAINING EFFICIENCY Quite a bit of research has looked at the structure of teams with respect to various criteria such as flexibility and proficiency. Most of this will be examined in Chapter 16. Very little is known about the optimal structure of teams for purposes of training. However, there is good reason to believe that the characteristics of the task itself are the most important considerations in determining how the team should be structured during the training (Briggs and Naylor, 1965).

TEAM PROFICIENCY MEASURES One of the major problems of team training research is the difficulty in obtaining good measures of team proficiency. This often leads to a substitute procedure of evaluating a team in terms of the goodness of the performance of individual team members—a very tenuous procedure at best. Glanzer (1965) has strongly urged that more time be spent on developing good measures of team proficiency so that we can better evaluate our training methods in this important training area.

PROGRAMED INSTRUCTION

Today in education the term "teaching machine" is a common expression. It is but one of a number of labels given to a particular method of instruction which is becoming very popular in industry. Other terms with equivalent meaning are "auto-

instructional device" and "programed instruction." Basically, these terms all refer to a process of *individual self-instruction* on the part of the trainee, a process which requires the active individual participation of the student.[6]

In essence a teaching machine is a device which presents organized sequences of instructional material to the individual student in such a way that he is required to make definite responses to the material. The machine or "program" is also designed to provide the trainee with immediate feedback about the correctness of his response. S. L. Pressey, a professor at The Ohio State University, is credited with "inventing" the teaching machine around 1924. His original machine is now on display at the Smithsonian Institution. Pressey's technique for presenting material to trainees received very little attention until the famous learning theorist, B. F. Skinner, wrote a paper which provided theoretical support for Pressey's procedure (Skinner, 1954). This paper had the effect of making programed instruction "scientifically respectable," and everyone started to become interested.

There are a variety of actual devices for presenting programed material to students. These range from rather expensive electronic devices which display material and provide buttons which, when pressed, indicate the correctness of the response to the question asked, down to simple paper-and-pencil devices which provide spaces for a person's response to each question and give the correct answer when a tab section is raised. Figure 8.10 gives an example of how questions are phased in a typical programed learning text.

	Answers
The numbered units in a programed textbook are called frames. A blank line within a frame indicates that a single word or number is to be supplied.	
Example. A textbook which asks the reader to supply answers within each of its numbered units is called a _____ textbook.	programed
If the blank requires two or more words to be supplied, the blank will be a double line.	
Example To use such a textbook most effectively, a student always conceals the printed answers until he has _____	written his own answers
As these examples have illustrated, the answer to a blank appears on the same line of print with the blank to which it belongs. When two blanks appear on the same line, the two answers are separated by a dash	
Example· Another suggested rule for using a programed textbook is that a _____ period should be _____.	study—short
Occasionally, the answers which belong to two successive blanks may be interchangeable; that is, they would be correct in either order. In such cases, the dash is replaced by a double-headed arrow, and such an arrow is included even when the two blanks do not fall in the same line of print.	
Example· When a coin is tossed, it will turn up _____ or _____.	heads ⟷ tails

Figure 8.10. An example of a series of frames in a programed text. (Examples are from the introduction to C. McCollough and L. Van Atta. *Statistical concepts.* McGraw-Hill, New York, 1963.)

[6] For more complete discussions on programed instruction the following sources are recommended: J. L. Hughes (ed.). *Programmed learning: A critical evaluation.* Educational Methods, Inc., Chicago, 1964; and A. A. Lumsdaine and R. Glaser (eds.). *Teaching machines and programed learning: A source book.* National Educational Association, Washington, D. C., 1960.

The extent to which various forms of programed material are in use in industry today is hard to estimate. The incidence is probably quite high, however, and is likely to continue to grow. Those who favor programed instruction cite a number of advantages of this training procedure:

1. *Low training costs.* The total cost of developing a program for a particular training situation plus the cost of administration is usually lower than the cost incurred through other forms of instruction.
2. *Flexibility.* Since programed instruction is *individual* instruction, it is often more flexible than other methods.
3. *Greater organization.* The process of writing a program forces a close examination of teaching objectives, thus often improving the teaching process by bringing about a better clarification of goals and a removal of extraneous material. (This is often a time and dollar savings also.)
4. *Rigid control over material.* With programs one has a rather complete control over the learning process and also over the feedback process.
5. *Complete record of student performance.* Because the trainee is continually making responses to each frame and these responses can be made a matter of record, it is possible to keep exceptional watch on the progress of the learner.

Programed instruction does, however, have certain disadvantages. Among the primary objections are:

1. *No chance for extra clarification.* It is impossible to provide the trainee with answers to questions which might be raised related to the material being presented. Thus, all that will be learned is what is presented in the program. The program must be assumed to be a sufficient learning experience.
2. *No individualized training.* All trainees, regardless of ability, receive the same program. The only flexibility is in terms of the speed with which different trainees complete the program.

Interestingly enough, Pressey himself has become somewhat dismayed at the rapid growth in the use of teaching machines. To quote from his article (1963, p. 5):

> Teaching machines and programmed materials are now being used all over the country in schools and colleges and in industrial and military training. Manufacture and sale of such products are a major enterprise of many publishers and equipment makers. Ambitious young people are embarking on careers in such work. The whole subject has become an accepted topic of everyday talk. However, there is disturbing evidence that current auto-instruction is *not* up to the claims made for it, that the current "boom" might be followed by a "bust" unfortunate for those involved—and for psychology. This paper is first of all a plea that to guard against such a danger the whole situation be soon given close critical inspection, and not merely to assure (as is now being attempted) that programs are good; but critically to consider whether the whole current concept of programing may be at fault, and an almost totally different approach than now orthodox to all ideas about auto-instruction be called for.
>
> The archvillain, leading so many people astray, is declared to be learning theory. No less a charge is made than that the whole trend of American research and theory as regards learning has been based on a false premise—that the important features of human learning are to be found in animals. Instead, the all-important fact is that human has transcended animal learning. Language, number, such skills as silent reading, make possible facilitations of learning, and kinds of learning, impossible even for the apes. Auto-instruction should enhance such potentials. Instead, current animal-derived procedures in auto-instruction destroy meaningful structure to present fragments serially in programs, and replace processes of cognitive clarification with largely rote reinforcings of bit learnings.

TEACHING MACHINE RESEARCH

In spite of the popularity of the teaching machine, there has not been a great deal of research carried out to determine in any precise fashion its relative merits as an instructional device. There is evidence (Coulson and Silberman, 1959; Fry, 1960) that a "write-in" program (see Figure 8.10) takes longer to complete than a program involving multiple choice responding, but also produces a higher level of learning. However, such research deals with the way in which programs can be made most effective and does not indicate whether programed instruction is superior to other teaching methods.

More recently, Goldberg and Dawson (1964) reported a study done with 47 clerical employees in which they compared a conventional instruction method (lecture-discussion) with two different methods of programed instruction (programed text and teaching machine). The employees were divided into three comparable groups and taught statistics by one of the three methods. No significant differences were found in achievement among the groups at the end of training. But when employees were retested six months later, differences were obtained. As Table 8.4 shows, the group taught by the conventional instructional methods exhibited the greatest amount of retention.

TABLE 8.4 *Mean Scores on Achievement Tests for Those Trainees Retested After a Six-month Period*

		Achievement Test 1			*Achievement Test 2*		
Group	*N*	*Test*	*Retest*	*Percent Loss*	*Test*	*Retest*	*Percent Loss*
Classroom	12	39 3	34.4	12.5%	21.4	18.8	12.2%
Machine	15	41.6	32.3	22 4	17.1	11.7	31.6
Textbook	13	42.2	32.8	22.3	16.6	13.3	19.9

SOURCE: M. H. Goldberg and R. I. Dawson. Comparison of programmed and conventional instruction methods. *Journal of Applied Psychology*, 1964, *48*, 110–114.

EVALUATION OF TRAINING EFFECTIVENESS

It was pointed out in Chapter 6 that the criterion problem would be a common problem throughout all the other "content" areas to be discussed later. Here we find it popping up in the context of training. How should one evaluate the "goodness" or "badness" of a training method or a training program? What criterion or criteria should be used? Certainly we would hope that training procedures implemented in industry would be subjected to some form of critical assessment to determine if they are accomplishing their objectives.

Actually, the particular measure of effectiveness is probably not so critical as the way in which the effectiveness data are gathered. For example, suppose one gave all trainees a pretest, then put them through a ten-week training course, and then gave them a post-training examination. Suppose further that the average test score was

significantly higher after training than it was before. Does this indicate that the training was successful? Not really, since we have no way of knowing whether the improvement was due to training or due to any other factor. *The only way training effectiveness can be truly evaluated is by comparing the change in performance of a group receiving training to the change in performance of a comparable group which did not receive the training.* If the two groups (training and control) are both equated at the beginning, then one need only compare their relative performance at the end of the training period. However, since equating groups is always a difficult task, it is usually safer to stick with the process of comparing the degree of change of the two groups, a process which adjusts for original skill level (see Figure 8.11).

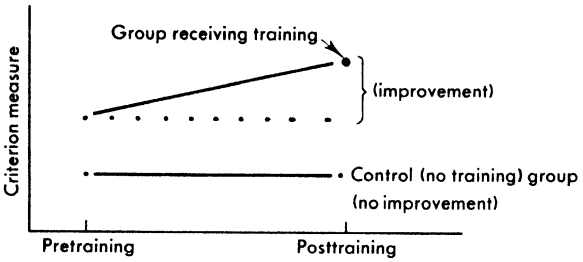


Figure 8.11. Illustration of method to determine the effectiveness of training using a control group which did not receive training.

It is therefore important to remember that it is not enough merely to recommend desirable training procedures and describe some of those in operation. It is necessary for the industrial psychologist to conduct research in this field. The essential difference between a scientific and a nonscientific training program is that the former requires objective evaluation. Research methods must be used to determine the relative merits of training methods, aids, and results.

Mahler and Monroe (1952) attempted to gain some information in connection with four rather important and basic questions related to industrial training:

1. How are training needs determined?
2. What are the relative merits of different training methods?
3. What are the relative merits of different training aids?
4. What has training accomplished?

In an attempt to answer these questions they used three sources of information —a review of the literature, a questionnaire survey which brought responses from 150 companies out of a total of 253 selected companies solicited, and field visits to 30 selected companies. The results of the study, while interesting in terms of the reported case histories, are necessarily disappointing to the student of training research. They indicate that training research tends to be minimum in quantity and elementary in nature. More optimistically, they show a trend toward more and better training research, indicating that some of the companies doing rather impressive jobs never manage to report their results in the literature where it would be accessible to those interested in the problems of training research. Not more than one company in 40 actually studied the relative merits of various training methods or aids. Training effectiveness was usually evaluated by subjective judgment rather than by research methods. Practically no attempts were made to show that the alleged improvements could be attributed to the training rather than to other causes.

Before we can be certain that our industrial techniques lead to positive results rather than wishful thinking, we must be able to obtain a measure of the training accomplishment. In other words, criteria are needed.

The criteria suggested by Lindahl (1949) for checking the results of training are:

1. Better quality of production
2. Increase in the number of operators able to meet job standards
3. Reduction in time required to do a specific job
4. Decrease in breakable supplies or tools
5. Decrease in absenteeism
6. Reduction in separation rate
7. Reduction in operational costs
8. Better performance on personnel tools such as tests, rating scales, and attitude surveys

McGehee has been conducting rather valuable work in the area of training. Two of his studies are refreshing because of their practical concern for real problems. In one study (1948) he established that fast and slow learners could be differentiated as early as the second week of training. By the end of the second week predictions could be made differentiating the two groups with a 20 percent better than chance accuracy. By the end of the sixth week, the advantage over chance is 63 percent.

In another study (1952) a 61.6 percent reduction in waste was effected as a result of a realistic awareness of training and its concomitant problems. McGehee recommends that management present frankly, fully, and sincerely the reasons for any change to all who are affected by it, then enlist the participation of the individuals involved by assigning specific tasks at their level of competence. In addition, such individuals must be informed of results and given recognition for assisting in the program.

McGehee's idea is that one explains the need for change, secures participation, and keeps participants informed. Group decision is omitted from McGehee's requirements, and this view is in contradiction to those who believe that group decision on goals is a most important part of the training.

An analysis of the conflicting hypotheses and some of the data lead to the theory that if the trainees can accept management's need and if the problem is sincere and not subterfuge, then group decision may not be necessary at all. Group decision can even be a subterfuge or "gimmick," in which case it will surely boomerang like any of the other tricks that will work for only a short time.

Incidentally, McGehee and Livingstone (1954), in a follow-up of the study in which waste was cut by 61.6 percent, found that with no reinforcement of training the reduced waste rate continued for 80 weeks. While this is not evidence of permanence, it is at least evidence of a continued effect over a reasonably long period.

SPECIALIZED TRAINING

The trend in education in recent years has been toward specialized training. The rise in popularity of business and engineering colleges, coupled with a general decrease in the proportionate enrollments in schools in liberal arts, illustrates this point. One recent development has been the formation of junior colleges aimed at specializing for industry.

A recently organized school in New York City, which functions at the junior college level, is known as the Fashion Institute of Technology and Design. It is

supported in part by the Board of Education of New York and in part by an educational foundation of the needle trades. The school offers college subjects, but in addition gives specific training for work in any of the many divisions of the needle trades, such as gloves, underwear, suits, dresses, shirts, blouses, and millinery. Its graduates are trained in illustrating, designing, and production control, and also in the techniques of sewing and draping. A feature of this school is the cooperative education in effect in the last semester. During this semester each senior goes to school for 2 five-week periods and works in industry for 2 five-week periods. When students are in school, alternates are on the job. The school makes a real attempt to place a student in the specific industrial job for which he has been training. This coordination of job experience and classroom work gives the graduates working experience and a knowledge of a job in a factory, sample room, or other department of an industrial establishment, as well as the knowledge gained in the classroom. Graduates know what the pressure of production means. They see clearly the relationship between education and job training. Since most people must eventually enter an office, factory, or profession, such real experience in addition to a more general education is of decided value. However, the "set" and "shift" of alternating classroom and work activities has its disruptive influences in the educational process. The week prior to work and the week of return to the classroom is noticeably more chaotic than the usual classroom scene.

Cooperative education has its perils. Students often do not respond smoothly to the shifting scene—now school, now work, now school. Tensions are built up concerning the job the student hopes to get. If economic conditions are bad, he may not even get a job.

Some employers are more interested in "exploitation" than training. They are served high-caliber personnel at minimum wages. The administrators of the program are subjected to pressure from employers to send top-ranking students and from students to obtain top-ranking jobs. All students and jobs are not top-ranking.

Cooperative education does have its shortcomings and limitations, and in extreme instances it may interfere with the academic training process and not even provide industrial training. All this is mentioned not to discourage the development of cooperative education but rather to improve it by pointing out the potential trouble spots so that efforts may be made to avoid them.

Some industries and specific companies conduct their own training schools, not only to make sure of having acceptable applicants but also because they recognize the importance of this training. An outstanding example is the Ford school in Dearborn, which trains young men in automobile mechanics and related fields.

Industry generally does not recognize its responsibility for training workers. Possibly the chief reason for this is the lack of facilities and of "know-how." In the past, because of an undue fear of the taint of commercialism, schools and colleges have too often refused to recognize the service they can perform. Cooperation between these two great forces in the training of workers has been conspicuously absent. The trend in this country is toward more and better education, with enrollment figures steadily increasing. Now is the time for both industry and educational institutions to cooperate in making training in industry more realistic and more valuable.

The International Harvester Company of Chicago and the University of Chicago afford a rare example of what can be achieved when "company funds" and "college brains" get together. The training courses offered in the Central Sales School are only part of the program. These courses are broken down into four areas: orientation,

personnel development, management operation, and technical aspects of the job. The program is handled on a mass scale; 17,000 men graduated from the school in a single year. In addition, plants are being developed to train the supervisory personnel along the lines of centralized and individualized needs. Harvester's training program recognizes not only the need for employee orientation but also the value of off-hour training directed toward the employee's intellectual pursuits, recreational inclination, and personal and domestic situation. This company regards its entire training program as a social contribution to the need for better informed adults in America. It is interested in developing individuals, not simply in training men.

The relationship between Harvester and the University of Chicago is exceedingly interesting. Since it is hoped that their cooperation will serve as a model for future collaboration between industry and colleges, the opening paragraphs of their contract are quoted:

> The International Harvester Company is interested in developing a program of training and education which will increase the feeling of unity in the organization; develop individuals in the company for responsible positions; facilitate the absorption of new employees and returned servicemen and women into the organization, and, in time, contribute to a better understanding of the application of machinery in agriculture. To achieve these ends, the company plans to develop the appropriate staff and physical facilities within its organization. In the development of these plans, the advice and counsel of specialists in adult education, business education, vocational education, etc., on the staff of a leading university would be welcome.
>
> The University of Chicago through its Department of Education in the Division of Social Sciences and through its School of Business is interested in studying methods and procedures for communicating information of varying intellectual contents to individuals of varied backgrounds. The experience attained within a university and its laboratory schools is restricted both in regard to intellectual content and the heterogeneity of the individuals with whom it comes in contact For these reasons the University welcomes the opportunity to cooperate in the formulation of objectives, development of instructional methods, and the measurement of the achievement of the objectives made possible by the educational plan of a large industrial organization
>
> It is therefore proposed that the International Harvester Company and the University of Chicago enter into a working relationship to achieve their mutual ends

The role of this university in training instructors for the Harvester Company shows that college professors can improve instruction for industrial training and that trained instructors should devote their attention to what the students are learning rather than concentrate on lecture organization. The company instructors are taught to define objectives not in terms of content but in terms of the students' changes in behavior. They are also taught to eliminate nonessential material from classroom courses.

TYPICAL TRAINING DEPARTMENTS IN ACTION

The average worker insists that he was never trained for his job, regardless of whether he has worked for small or large organizations. By and large this is true. Business organizations must recognize their responsibility and create training departments on a larger scale than has been done in the past.

The training responsibilities of an industrial or business organization are clear-cut, and its training program can be considered as having two functions. The first, a

general orientation in company policies, routines, and rules, should attempt to assimilate the new employee as rapidly as possible so that he may overcome his feeling of being a stranger. The second function concerns specific job training. The training programs of companies, large and small, generally are weak in both these aspects.

One author is reminded of the time he was hired as a factory operator in a large manufacturing concern. After a series of interviews and a physical examination he was assigned a number, given a time card, and sent to a specific department. After showing his credentials to the department supervisor, he was seated at a work bench and told: "This is Joe. He will tell you what to do." Joe, who seemed rather annoyed, said merely, "The best way to learn is to help me do this." At the end of the day the supervisor came over and asked for the new man's production. Obviously there was none because he had been helping the man who was supposedly training him. Rather than say this and get into trouble with Joe, the author said that he had not produced anything because he thought he was not supposed to work the first day. The supervisor looked at him in a manner that he clearly understood —it suggested that he was pretty dumb—and said, "Be sure that you produce tomorrow." The next day he went to work, asked as few questions as he could, and did not annoy Joe. Of course, his production would have been greater if he had had at least an hour's demonstration and training. Some time later he discovered that Joe disregarded him not for personal reasons, but simply because he himself had to produce his usual amount of work and was given no allowance for the time spent in training a new man.

In some smaller organizations the employee is hired, put to work, and told to ask any questions he wants to; but since it is never made clear who is supposed to answer these questions, the average worker gains experience by struggling along meekly on a hit-or-miss basis. Industry can and should do a better job. It is not necessary for an organization to make a fetish of training, but hiring an experienced educator or psychologist on a consulting basis one hour a week for a month or two will often enable it to set up a training program that is far better than no training at all.

The Electric Auto-Lite Company emphasizes visual aids in its training program. In fact, Walt Disney's first industrial production was made for this company. The film, which ran 18 minutes and cost about $75,000, was designed to teach the correct installation of spark plugs.

General Motors has a gigantic training program covering many areas. The training in the field of management and supervision includes many different specialized programs. The company has a large apprentice training program and does considerable on-the-job training. It also conducts an institute featuring cooperative education. The courses offered vary from special classes of a few days' duration to a four-year course which, together with certain prescribed postgraduate work in the fifth year, leads to an engineering degree.

THE TRAINING STAFF

Training in industry is given most frequently by an experienced operator. The supervisor and the special instructor follow in that order, but there is often an inverse relationship between this order and effectiveness in training. There is absolutely

no reason for assuming that an experienced operator is a good teacher; such an operator is often not capable of even describing the method he uses. People who are excellent ice skaters often cannot teach anyone else to skate; the same applies to expert tennis players, carpenters, electricians, etc. It is necessary to train the experienced operator in training techniques before he can be considered an effective trainer. As for the supervisor, although it is true that he may know how to handle people, he may not know how to do the specific task in question. The best method of training in industry is to have a special instructor who can do the job adequately and knows how to teach as well; the latter is essential. Very frequently this means training the trainer.

Bavelas, formerly of the Industrial Relations Section of the Massachusetts Institute of Technology, has demonstrated this point. He has conducted several studies which indicate that the training of trainers results in better performance by the people being trained. He suggests that trainers be given training through discussion methods so that they will better understand the implications of attitude and motivation in a training program. In one study he devoted eight hours to such a course; the results indicate that differences in employee ability exist when performance is measured under three conditions: without training of the trainers, with four hours of training, and with the full eight-hour course.

REFERENCES

Annett, J. The role of knowledge of results in learning a survey. U.S. NAVTRADEVCEN Technical Document Report 342-3, May 1961.

Bahrick, H. P., P. M. Fitts, and G. E. Briggs. Learning curves. Facts or artifacts? *Psychological Bulletin*, 1957, 54, 256–268.

Baker, K E., R C. Wylie, and R. M. Gangné. Transfer of training to a motor skill as a function of variation in rate of response *Journal of Experimental Psychology*, 1950, 40, 721–732.

Bilodeau, E. A. Patterns of internal consistency in multipart skilled performances *American Journal of Psychology*, 1957, 70, 550–559.

Bilodeau, E. A. Variations in knowledge of component performance and its effects upon part-part and part-whole relations. *Journal of Experimental Psychology*, 1955, 50, 215–224.

Bilodeau, E. A., and Ina McD. Bilodeau. The contribution of component activities to the total psychomotor task. *Journal of Experimental Psychology*, 1954, 47, 37–46.

Briggs, G. E., P. M. Fitts, and H. P. Bahrick. Learning and performance in a complex tracking task as a function of visual noise. *Journal of Experimental Psychology*, 1957, 53, 379–387.

Briggs, G. E., and J. C. Naylor. The relative efficiency of several training methods as a function of transfer task complexity. *Journal of Experimental Psychology*, 1962, 64, 505–512.

Briggs, G. E , and L. K. Waters. Training and transfer as a function of component interaction. *Journal of Experimental Psychology*, 1958, 56, 492–500.

Britt, S. H. The relationship between transfer of training and age of previous association. *American Journal of Psychology*, 1934, 46, 113–116.

Bryan, W. L., and N. Harter. Studies in the telegraphic language. *Psychological Review*, 1899, 6, 346–376.

Bunch, M. E. Transfer of training in the mastery of an antagonistic habit after varying intervals of time. *Journal of Comparative Psychology*, 1939, 28, 189–200.

Bunch, M. E., and E. S. Lang. The amount of transfer of training from partial learning after varying intervals of time. *Journal of Comparative Psychology*, 1939, 27, 449–459.

Bunch, M. E., and M. Rogers. The relationship between transfer and the length of the interval separating the mastery of the two problems. *Journal of Comparative Psychology*, 1936, 21, 37–52.

Carter, L. F. Exercising the executive decision-making function in large systems. In R. Glaser (ed.). *Training research and education*. Wiley, New York, 1965.

Coulson, J. E., and H. F. Silberman. *Results of an initial experiment in automated teaching*. System Development Corporation, Santa Monica, Calif., July 1959.

Day, R. H. Relative task difficulty and transfer of training in skilled performance. *Psychological Bulletin*, 1956, 53, 160–168.

Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *Journal of Experimental Psychology*, 1953, 45, 1–11.

Eckstrand, G. A. Current status of the technology of training. *AMRL Document Technical Report 64-86*, September 1964.

Egerman, K., D. J. Klaus, and F. Glaser. Increasing team proficiency through training. 3. Decremental effects of reinforcement in teams with redundant members. Technical Report, 1962, American Institute for Research, Palo Alto, Calif, Contract No. Nonr-2551(00), Office of Naval Research.

Fleishman, E. A., and W. E. Hempel, Jr. The relationship between abilities and improvement with practice in a visual discrimination task. *Journal of Experimental Psychology*, 1955, 49, 301–312.

Fry, E. B. A study of teaching machine response modes In A. A. Lumsdaine and R. Glaser (eds.). *Teaching machine and programmed learning. a source book*. National Educational Association, Washington, D.C., 1960, 469–474.

Fuchs, A. H. The progression-regression hypotheses in perceptual-motor skill learning. *Journal of Experimental Psychology*, 1962, 63, 177–182.

Gagné, R. M. Military training and principles of learning. *American Psychologist*, 1962, 17, 83–91.

Gibbs, C. B. Transfer of training and skill assumptions in tracking tasks. *Quarterly Journal of Experimental Psychology*, 1951, 3, 99–110.

Glanzer, M. Experimental study of team training and team functioning In R. Glaser (ed.). *Training Research and Education*. Wiley, New York, 1965.

Glaser, R, D. J. Klaus, and K. Egerman. Increasing team proficiency through training. 2. The acquisition and extinction of a team response. Technical Report, 1962, American Institute for Research, Palo Alto, Calif., Contract No. Nonr-2551(00), Office of Naval Research.

Goldberg, M. H., and R. I. Dawson. Comparison of programmed and conventional instruction methods. *Journal of Applied Psychology*, 1964, 48, 110–114.

Harris, E. F., and E. A. Fleishman. Human relations training and the stability of leadership patterns. *Journal of Applied Psychology*, 1955, 39, 20–25.

Hauty, G. T. Response similarity-dissimilarity and differential motor transfer effect. *Journal of Psychology*, 1953, 36, 363–378.

Horrocks, J. E., E. Heermann, and R. Krug. Team training III. An approach to optimum methods and procedures. United States Navy Training Device Center Technical Report, 1961, No. 198-3.

Horrocks, J. E., R. E. Krug, and E. Heermann. Team training II. Individual learning and team performance. United States Navy Training Device Center Technical Report, 1960, No. 198-2.

Howell, W. C., and D. L. Kreidler. Information processing under contradictory instructional sets. *Journal of Experimental Psychology*, 1963, 65, 39–46.

Howell, W. C., and D. L. Kreidler. Instructional sets and subjective criterion levels in a complex information-processing task. *Journal of Experimental Psychology*, 1964, 68, No. 6, 612–614.

Kinkade, R. G., and J. S. Kidd. The use of an operational game as a method of task familiarization. *Journal of Applied Psychology*, 1962, 46, 1–5.

Lindahl, L. G. How to build a training program. *Personnel Journal*, 1949, 27, 417–419.

McCollough, C., and L. Van Atta. *Statistical concepts.* McGraw-Hill, New York, 1963.

McGehee, W. Cutting training waste. *Personnel Psychology*, 1948, 1, 331–340.

McGehee, W., and D. W. Livingstone. Training reduces material waste. *Personnel Psychology*, 1952, 5, 115–123.

McGehee, W., and D. W. Livingstone. Persistence of the effects of training employees to reduce waste. *Personnel Psychology*, 1954, 7, 33–39.

McGeoch, J. A., and A. L. Irion. *The psychology of human learning*, 2nd ed. Longmans, New York, 1952.

Mahler, W. R., and W. S. Monroe. How industry determines the need for and effectiveness of training. *PRS Report* 929, Department of the Army, Washington, D.C., 1952.

Maier, N. R. F. Principles of human relations. Wiley, New York, 1952.

Naylor, J. C. Parameters affecting the relative efficiency of part and whole practice methods: A review of the literature. *United States Naval Training Devices Center Technical Report*, No. 950-1, February 1962.

Naylor, J. C., and G. E. Briggs. The effect of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology*, 1963, 65, 217–224.

Naylor, J. C., and G. E. Briggs. Long-term retention of learned skills. A review of the literature. ASD Technical Report 61-390, Wright-Patterson Air Force Base. Aeronautical Systems Division, Acrospace Medical Laboratory, 1961.

Naylor, J. C., and G. E. Briggs. Team-training effectiveness under various conditions. *Journal of Applied Psychology*, 1965, 49, No. 4, 223–229.

Pressey, S. L. Teaching machine (and learning theory) crisis. *Journal of Applied Psychology*, 1963, 47, 1–6.

Rosenberg, S , and R. L. Hall. The effects of different social feedback conditions upon performance in dyadic teams. *,ournal of Abnormal Social Psychology*, 1958, 57, 271–277.

Seagoe, M. V. Qualitative wholes: A revaluation of the whole-part problem. *Journal of Educational Psychology*, 1936, 27, 537–545.

Skinner, B. F. The science of learning and the art of teaching. *Harvard Educational Review*, 1954, 24, 86–97.

Tannenbaum, L. R., I R. Weschler, and F. Massarik Leadership and organization: A behavioral science approach. McGraw-Hill, New York, 1961.

Tiffin, J., and E. J. McCormick. *Industrial psychology*, 4th ed. Prentice-Hall, Englewood Cliffs, N.J., 1958.

Wiest, W. M., L W. Porter, and E. E. Ghiselli. Relationships between individual proficiency and team performance and efficiency. *Journal of Applied Psychology*, 1961, 45, 435–440.

9 ATTITUDE MEASUREMENT

All people have attitudes which result in tendencies to respond positively or negatively to another person, a group of people, an object, a situation involving objects and people, or an idea. Very often the possession of an attitude predisposes the individual to react in a specified direction. This being so, a knowledge of the attitude sometimes allows for the prediction of behavior, both in industry and in other aspects of life.

An attitude is acquired or learned by the individual, and at times the learning process is so subtle that he may not even recognize the attitude he possesses or truly understand how it was acquired. Attitudes may be rational or irrational, and the possession of intelligence is by itself no guarantee of a larger number of rational attitudes. The individual who has a certain attitude often has to rationalize or self-justify his reasons for it by drawing upon only the illustrations that back it up.

The acquired determining tendency in one's thinking or behavior may be the result of fact or opinion, but for the person who holds the particular attitude this has no real importance. Attitude based on opinion may be more intense than attitude based on facts.

ATTITUDES AND SOCIAL PHENOMENA

Interpersonal relations, to a large extent, are determined by the attitudes people hold. Similar attitudes tend to produce more cooperation and dissimilar attitudes produce more friction among individuals.

Before discussing how attitudes are formed, changed, and measured, we should agree on what attitudes are so that we do not have different attitudes about attitudes. By *attitudes* we mean the beliefs, feelings, and action tendencies of an individual or group of individuals toward objects, ideas, and people. Quite often persons and objects or ideas become associated in the minds of individuals and as a result attitudes become multidimensional and complex.

For most attitudes this complexity is the rule despite the fact that attitudes indicate a general and almost oversimplified direction toward favor or disfavor. It should be remembered that the constituent parts of an attitude contributing to this generalization are never as logical to the observer as they are to the holder of the attitude. Thus, your friends may point to inconsistencies in your beliefs, feelings, and action tendencies whereas you find that the components are not inconsistencies but rather are reinforcements leading to the "logicalness" of your position for or against the idea, object, or person.

Attitudes are reflected by the opinions or views that we hold. Attitudes are often formed by the judgments we make as a result of information we receive from various sources. When we find this information credible and congruent with our predispositions and previous judgments, we develop a favorable or positive attitude toward the new object, idea, person, or group. However, when the information, no matter how received, is incongruent it can be judged as incredible and the result is a negative attitude.

Prejudices are attitudes. Half humorously, prejudices are the attitudes of others that we do not share. Prejudices are, therefore, often called "wrong" or "bad." More seriously, prejudices have an important impact on all of us and on our interpersonal relations. Every one is in favor of reducing prejudice, but not everyone is willing to admit that the attitude held is a prejudicial one. Someone who is anti-Negro, anti-Semitic, or anti anything presents "facts" to prove that he is not prejudiced and that his beliefs are supported in fact.

Regardless of the level of intelligence, most individuals hold attitudes on most topics. High intelligence does not necessarily release people from prejudices, nor does it cause people to seek a full and objective inquiry before forming opinions. In other words, people with varying intelligence levels often share the same attitudes and to the same intense degrees.

Attitudes are not always a function of degree or amount of knowledge about the object of the attitude. Justifications of attitudes we hold are often a function of the rationalizations we use to justify the knowledge, or lack of it, we have on almost any topic.

For any person, attitudes are variously related to motives, values, personality, and emotions. The individual has internal needs and external social pressures. In relating these needs and pressures he manifests wants, aspirations, and expectancies. In further relating these to other people he expresses his attitudes. When attitudes are weak he may not even choose to defend them; but when attitudes about an idea, object, person, or group are intense, he will indulge in as many defense mechanisms as necessary to appear logical, lofty, and righteous.

The intense holders of attitudes probably are the rabid reformers; they may even be the most sublime of the idealists. In other words, attitude intensity often produces strange bedfellows. Politicians seem to recognize this with cunning insight. As a result they can, when they are successful, get peculiar combinations of backers supporting with intensity *their* candidate who cannot possibly be the same man to such different people.

Attitudes are also related to the groups with which a person wants to identify and belong. People at all ages seek to belong to certain groups of a formal or informal variety. This is reflected in clubs (good) and gangs (bad). We have "good colleges" and "poor colleges"—and which is which depends in part on *who* and *what* you believe. Think of the large number of organized groups: Republicans, fraternities, Daughters of the American Revolution, CORE, Ku Klux Klan, Black Muslims, Anti-Defamation League, Anti-Vivisectionists, etc. Such organizations, as well as all others, tend to have some rabid and inspired members, and for each of these there is a nonmember just as strongly opposed to the organization. Most groups also have many members who belong but are passive in their group role.

How does it happen that groups can have staunch supporters, passive onlookers, and rabid attackers? Only part of the answer comes from joining for social approval or gratification. The other part of the answer must stem from the attitudes that

people have prior to joining the group. They expect the group to help them achieve their expectancies or at least to find congruencies in their attitudes and the attitudes of others. Whereas the motives for joining a group such as a nudist colony may differ among individuals, the satisfaction of finding a number of people with similar positive attitudes toward nudism becomes a common denominator. And then there are those who must prevent others from being nudists, and so they form groups to prevent nudist clubs from existing, even if it means such sacrifices as spying on nudists to be able to produce the evidence.

ATTITUDE FORMATION

In essence, a large part of our social lives is a result of our attitudes. These attitudes in turn are a result of the relationships we have with those who contribute to them, such as members of our families, our friends, our teachers, our neighbors, and our religious advisors.

The role of parents and siblings in contributing to our personality formation is well known. The role of the family in contributing to the formation of our attitudes operates similarly. Attitudes toward the opposite sex, religion, tolerance or prejudice, education, occupations, political parties, and almost all other areas where attitudes are capable of expression are to quite an extent the result of our accepting or rejecting the attitudes held by members of our family. The reader might well pause at this point and review his attitudes on any of various topics and by self-analysis attempt to determine the contributing role of his mother, father, sister, brother, aunt, uncle.

The neighborhood we live in has a certain structure in terms of its housing, cultural facilities, religious groupings, and possibly ethnic differences. Further, it has people who are neighbors. The neighbors—adults or children—tolerate, condone, or deny certain attitudes and behavior, and as a result we are either New Yorkers, Midwesterners, Southerners, etc. Further, we accept these mores and conform, or we deny them and possibly rebel. The conformity or rebellion in some respects are the evidences of the attitudes we hold.

Regardless of one's religion, there seems to be the same opportunity to conform and practice orthodoxy or rebel and either become irreligious or convert to another religion that has attitudes more acceptable to the self. The interplay of family, neighbors, and religious advisors combine to help form our attitudes not only toward the many aspects of religious living, but also toward politics, sex, food, and our interpersonal relations with others.

Consider your college campus and those whom you choose as friends. Can you see how attitudes play back and forth in your choice of friends and in turn their choice of you? Why are some friendships long lasting and others of much shorter duration? The beliefs, feelings, action tendencies, and behavior manifestations are more likely to be similar toward objects, ideas, and people among friends. In other words, friendships are more likely to continue as attitudes are similarly formed and held in common.

Further, our economic and occupational position and aspirations also contribute to our attitudes. They determine in part our attitudes toward unions and management, and our belief that certain laws are "good" or "bad." In short, our entire socioeconomic background influences our present and future attitudes.

As another example of the kinds of things that contribute to attitude formation we refer to the mass communicators. All varieties of mass communicators—tele-

vision, radio, newspapers, and magazines—"feed" their audiences large quantities of "information." The presentation of news or information is constructed so as to cater to the attitudes of the audience. In turn, the audience selects the specific form of mass communication that best reflects its attitudes on various subjects. The items of interest include sex and teen-agers, crime, divorce, politics, religion, dope addiction, civil rights, and pornographic literature—among many others. The material we select helps us either to substantiate our opinions or to establish new ones.

In essence, opinions are attitudes, and almost all of us have opinions on almost everything. Sometimes we are strongly for or against; sometimes we reserve judgment because we are confused or not sure. We rarely find ourselves with no opinion at all, but at times we are only mildly for or against something.

The mass communication media and individuals interact with each other. The former are capable of forming or influencing the attitudes of their audience, and the latter are capable of determining the success or failure of the mass communicator by subscribing, reading, viewing, or rejecting. For example, we hear a great deal about the "low level" or content and programing of the typical radio station, television station, or popular magazines. Yet those who are successful, that is, are catering to mass audiences, must be doing something "right." They cater to the prevailing attitudes and opinions of their audience and use surveys to prove that they are right. They hire "creative" people to see that they have the pulse of their public.

The mass communication media use effective but nonscientific methods of attitude presentation. They represent particular points of view about foreign countries, the United Nations, color television, the President, taxes, federal aid to education, dieting, smoking, fashion, and art. They act to form attitudes or change existing ones. This may be done subtly or not so subtly. In any event, they are influential in attitude formation and the reflection of which common attitudes are considered "public opinion."

ATTITUDES AND BEHAVIOR

Attitudes do not always predict behavior. Attitudes act to reflect more of the emotional components of life rather than behavior. For example, if grades in college are an indication of performance or behavior, then it should be clear that students' attitudes do not always predict grades. Part of the reason is that attitudes are capable of change, and so one's present attitude may or may not be the same at a different time. Another reason is that attitudes may be generalized or specific, and in the latter instance the specific attitude may not be typical of the general. For example, most people have a favorable attitude toward Boy Scouts in general, but they may not have the same attitude about the local troop after they have entertained them at a barbecue.

Other things being equal, it is necessary to establish a relationship between attitude and specific behaviors. Attitude toward sex may or may not be predictive of behavior. Attitude toward tolerance of minority groups may be favorable, but behavioral intolerance may be demonstrated toward individuals of the group. The favorable attitude toward a certain brand of a television set may not result in that purchase because of price, availability, and the attitude of the salesman.

The original attitude may have been measured quite accurately, but it may change as a result of circumstances at the time of the behavioral decision.

If attitudes are subject to change, when can attitudes predict behavior? They are likely to do so when all the variables related to the behavior are known and when new variables are not introduced.

ATTITUDE CHANGE

Since attitudes form as a result of varieties of formal or informal learning experiences, they obviously are capable of change as a result of new and different learning experiences. Changes in attitudes can be classified as one of two types. The more readily obtained change generally can occur in the *degree of the already established direction.* When a person is for (or against) an object, idea, or person, it is possible to change the degree of the attitude held. Thus the degree can become more or less but still remain in the same direction (that is, either pro or con).

The second type of change is usually more difficult to achieve but is entirely within the realm of predictable possibility. It is the change in the *reversal of the direction of the attitude.* This change is measurable in behavioral terms such as change in retail store purchasing, change in voting for a different political candidate, change in the spouse, and resigning from an organization or joining one.

Krech, Crutchfield, and Ballachey (1962) refer to changes in degree of existing attitudes as congruent and changes from positive to negative (or the reverse) as incongruent. Attitude modifiability according to them is a function of seven attitude characteristics: (1) extremeness, (2) multiplexity, (3) consistency, (4) interconnectedness, (5) consonance, (6) strength and number of wants served by the attitude, and (7) centrality of value to which the attitude is related.

With reference to changes in attitude, the holders of more extreme attitudes are less likely to change. The greater the multiplexity of an attitude, the less likely is a change in direction of attitude to occur, but the more likely is a change of degree in present direction to occur. Attitudes with consistency among its components tend to be stable and least likely to change, but an attitude with inconsistent components can change more readily.

The more an attitude is connected to other attitudes, the less likely is change to occur. For example, if one is a conservative in all walks of life, then changing conservatism in one part of it will not be readily accomplished. Related is the consonance of attitude clusters. When an attitude exists in a state of consonance with other attitudes, then change is not likely to occur. Since attitudes can serve many wants and needs of an individual, the possibility of change will depend upon the number and strength of the wants served. And, last, the closer the attitude is to a basic value held by an individual, the less likely is change to occur.

Brown, Galanter, Hess, and Mandler (1962) studied attitude change by model construction and reached the conclusion that attitude change is the consequence of disequilibrium when positive and negative bonds are associated. This disequilibrium initiates change, and change operates in the direction of equilibrium restoration. They cite the work of Osgood, Suci, and Tannenbaum (1957) as developers of the Congruity model; Festinger (1957) as the sponsor of the Dissonance model; and Abelson and Rosenberg (1958) as originators of the Balance model.

In many respects the Congruity, Dissonance, and Balance models have a great deal in common, although it is true that they are somewhat different in their fine

points. The dimensions of congruity-incongruity, consonance-dissonance, and balance-imbalance enable one to understand the conditions conducive to attitude change better.

The congruity theory states that some associations between positives, some associations between negatives, and some dissociations between objects (both positive and negative) do not constitute equilibria, and this produces attitude change. Festinger's dissonance theory considers the importance of drives in attitude change. For him, equilibrium is consonance and disequilibrium is dissonance. Reduction of dissonance is achieved by activity leading to consonance. According to the Balance theory, there is equilibrium when elements of identical sign are linked by positive relations and when elements of opposite sign are linked by negative relations.

The essence of these three theories is that incongruity, dissonance, and imbalance are conditions of disequilibrium, and under such conditions attitude changes can occur and be demonstrated. Educators, politicians, religious advisors, parents, manufacturers, salesmen—almost everybody you can think of—work to change attitudes so that their "good" object or idea can have more followers and users. The work of the "model constructors" sets a theoretical framework to enable one to understand how attitudes can be changed.

Much experimental work on attitude change has been conducted in the laboratory, and two such studies will now be cited to illustrate the flavor. Kelman and Eagly (1965) report the results of two experiments. In the first, three communications to a group of Negro college students were the same. However, one-third of the students heard the tape-recorded communication from a communicator represented as a pompous, paternalistic authoritarian. The second group heard a communicator represented as a modest, humble, and objective scholar (a college professor). The third group heard from a person represented as a Negro reverend who spoke as a member of the Negro community.

The second experiment had for its subjects high school students. They heard taped messages emphasizing the juvenile delinquency problem. The negative communicator projected the image of an ignorant enemy, and the positive communicator projected the image of one who would be personally attractive to a teen-age group. The main conclusion drawn is that "the tendency to perceive communication content in line with one's attitude toward the communicator is most likely to come into play when the communicator arouses strong feelings." Kelman and Eagly hypothesize that "misconception is a function of the degree to which the incongruous situation raises questions at *self-definition* in the subject."

Considering the concept of disequilibrium as a propensity for attitude change, we suggest three signs that are conducive to attitude change. The first requires spotting trends. It can be assumed that a trend occurs when a need for change exists. Among the more socially stable (older folks), trends are least likely to be effective, except in instances where such people are personally involved as in social security or medicare. However, among the less socially stable (teen-agers), trends and fads can be established almost overnight. As a result, new-style singers, dances, bathing suits, manner of dress, and haircuts do become popular in a very short time. Why does this happen? Because of attitude incongruence and imbalance leading temporarily to balance or equilibrium which in turn leads to still newer trends if the original need was only temporarily satiated.

Another method of obtaining attitude change is to be able to spot an unfulfilled need. In this respect, research is more important than intuition. Convenience, price,

satisfaction, and durability are factors that can suggest unfulfilled needs that are not present in objects. The new product promises to fulfill a need not presently satisfied, and so faster relief from headache pain, easier methods to lose weight, the political candidate who will reduce taxes and increase government services, and that really different shiny safe new car all cater to the imbalance and disequilibrium in the consumer's attitude allowing for change in attitude and eliciting behavior in a specific direction.

The third sign of potential attitude change is to be aware of new members joining an organization, or more impressively, observing a new organization in its formative stages. The zeal and enthusiasm manifested is related to the need for establishing an equilibrium between the attitudes of the joiner and his new-found brethren who are expected to match his attitudes, especially since the older brethren in the older organization created the imbalance.

STUDYING EMPLOYEE ATTITUDES

One of the most fruitful pursuits for the industrial psychologist is to study determinants of the attitudes of employees and employers. Although some work has been done in this field, most attitude research has been aimed at gaining specific knowledge of employee attitudes in a specific situation. This is understandable since these studies are usually paid for by employers who are motivated by a desire either to promote efficiency or to iron out some of their difficulties with employees.

Employers, of course, have attitudes on as many things as employees, and the two are usually somewhat different. Employers' attitudes may lead to incorrect assumptions about employees, or sometimes, to employee behavior that is the antithesis of that which was assumed or predicted. Of course, the same is true of employees. There is a need to study employer as well as employee attitudes in order to get more complete insight into the problem of employer-employee relations.

In studying the attitudes of employers and employees, one must not assume the absence of predetermining factors prior to the job situation or, for that matter, prior to the work history. Such assumptions are fallacious and are likely to lead to sterile results. That a person is likely to be a "little liberal" or a "little conservative" before his first job has been determined by the many interesting influences of school, church, home, community, etc. People are likely to favor or reject things in accordance with their background—or, sometimes, in spite of it. Thus, facts are too often interpreted in the light of one's predetermined attitude toward the other fellow's behavior.

An employer may believe that his employees are interested only in salary and that they are not concerned with his problems. To prove this, he cites "facts" such as restricting production, willingness to break rules, etc. An employee may believe his employer is interested only in profits and that he treats him with less consideration than he does his machines. He also cites "facts" to prove this—low salary, deductions for minor infractions of rules, poor working conditions, and lack of interest in his problem of making a living. Not only is this a situation in which present and immediately past conditions contribute to the respective beliefs, but it is likely to be one that draws upon much of the past life of the individual, especially the part of it that is colored by emotion. The employer with an unfavorable attitude toward his employees is no more likely to understand his employees than an employee with an unfavorable attitude toward employers is likely to understand

his own employer. Poor employer-employee relations and industrial warfare are the inevitable impasse. A better mutual understanding of attitudes will not remove emotion from the situation, because attitudes are the very essence of emotions; but it will enable one to predict behavior with more accuracy and possibly to avoid conflict by bringing about changes in attitudes along fruitful lines.

Two additional characteristics of attitudes should be repeated before we go into the subject of their measurement. The first is that attitudes are not necessarily a result of intelligence or comprehension. They are part of our hedonistic life. Beginning with simple sensory feeling of pleasantness and unpleasantness we develop likes and dislikes. We further develop emotions, moods, and sentiments. When an individual has an attitude toward a person, subject, or thing, some aspect and degree of feeling accompanies it. It may be a like or dislike, a mood, a sentiment, or even an emotion or passion. A favorable attitude toward a work place means that it is generally a pleasant office or factory and that we like to work there. We might prefer not to accept another job because we are sentimental about the place. We then find ourselves in a favorable mood and at times exhibit various emotional forms of behavior on or about the job.

An unfavorable attitude has similar hedonistic aspects, except that they are negative. We dislike the setup. We are generally unhappy and in a depressed mood, hate our colleagues and bosses, and fly into a rage upon the slightest provocation.

While it is rare that attitudes change overnight, it is nevertheless true that they do change. This is their second characteristic. The fact that attitudes are susceptible to change makes their measurement more practical. To measure attitudes with scientific accuracy is one thing. To understand the formation of attitudes and to attempt to change the factors contributing to it is another. Although ordinarily this lies within the province of social psychology, all too often in the past the social psychologist has avoided the problems that confront industry. It is also true that the industrial psychologist has often overlooked the methods and techniques employed by the social psychologist. Actually, there is much overlapping of subject matter in these two fields, and work directed to the free exchange of knowledge and efforts between these two groups of psychologists, even to the point of ultimate integration, would be of great value.

If the industrial psychologist understands the complexities of attitude formation and the mutability of attitudes, he can do a better job in measuring industrial attitudes. Industrial psychology can make use of the techniques that have been developed by psychologists and modify them for the specific purpose at hand. As more work is done in the field, new and better methods will become available.

Attitude measurement technology is a most useful device in the hands of industrial psychologists. Specific information about job satisfaction and industrial morale, when properly obtained, can be very useful. The effects of changes in working conditions and environment, incentives, training programs, and many other factors can be measured in more ways than through production records. If the attitudes of employees are known both before and after a change is made, its ultimate success can be predicted more accurately. For example, an employer inaugurates a change —a bonus, a new workbench, or whatever—because he believes his employees will like it. Afterwards he finds that conditions are worse instead of better and accordingly concludes that his employees are ungrateful. This conclusion may be wrong. If he had known their attitudes in the first place, instead of that particular change

he might have made one that would have improved their attitudes, with resulting benefits to him and his company.

When management wants to discover the source of dissatisfaction and correct them, attitude surveys are justified. Surveys alone practically never increase production.

METHODS OF MEASURING ATTITUDES

THE SAMPLE

The previous sections on attitude have implied measurement. Changes in attitude, degree, or direction demand measuring techniques of established, reliable, and valid variety. Accordingly, psychologists and other social scientists devote much time and attention to sampling techniques and instrument construction so that attitudes can be measured accurately.

The results of any attitude measurement have to be generalized against or compared with a population with known characteristics. Predictions are possible when the group included for measurement is typical of the population being studied. In a national election the sample should be representative of the total number of persons who actually vote. For a new product, the sample of potential users studied must resemble those who will ultimately use the product. Various techniques have been developed to draw a sample.

Kish (1953) describes simple random sampling, probability sampling, area sampling, stratified sampling, and cluster sampling, among others. As can readily be inferred, the choice and size of subjects included in the sample is a matter of budget and time, but some systematic method of selecting the sample from the universe of the population is mandatory if the results are to have any reliability and validity.

In drawing a simple random sample, the size of the population, such as the number of employees, is known and a number is assigned to each. Those interviewed will be the total decided and those whose numbers were drawn by chance as if out of a hat. The probability sample selects respondents based upon a knowledge of the elements comprising the population; in other words, it is a statistical refinement of the random sample. Area sampling selects its respondents from boundaries that are defined and identifiable. Most often the respondents in area sampling are identified with dwelling units within the area.

Stratified sampling divides the population into subpopulations called strata. From each strata a sample is selected. For example, subsamples might consist of product users versus nonusers, or teen-age males and teen-age females, or almost any variable in the characteristics of a population about which one hopes to obtain data leading to a solution of the problem investigated. Cluster sampling requires the selection of respondents from defined groups or areas. Selecting a sample from five cities, eight blocks, three classes of general psychology students, or any other selected and defined segment of a population is the essence of cluster sampling.

Regardless of the method of sampling and the statistical refinements introduced to reduce sampling errors, the ultimate value of the sample obtained depends upon the quality and integrity of the interviewers. One cannot hope to obtain an adequate

sample without paying great attention to the briefing, supervising, and auditing of field interviews. Unfortunately, this fact is often overlooked by quite a few so-called research organizations, and when they do not obtain the sample characteristics desired, they "correct" for the imbalance by "projecting" from within segments of their own sample. Statistically this may be correct, but methodologically it is a poor and unsatisfactory substitute.

Six methods of measuring attitudes, each with its advantages and disadvantages, will be described and illustrated. Which one should be used will often depend on the person or group upon whom the decision rests. Attitude measurement can be used by an individual employer, a trade association, a union, or an informal employee group. A satisfied staff, a strike call, an expansion of the group, or a series of changes in the plant may result from the facts uncovered. The methods to be described are: (1) impressionistic, (2) guided interview, (3) unguided interview, (4) questionnaire, (5) attitude scale, and (6) indirect.

IMPRESSIONISTIC METHOD

The impressionistic method is nonstatistical in that it does not lead to quantitative knowledge. It is based upon the observation of behavior and attitudes. From the point of view of science, it is the least desirable of the six methods but because it is a method whereby attitude measures attitude, it is the most widely used. The industrial psychologist can only condone it. It is excusable when regarded as a preliminary to the other methods or when a very rapid spot survey must be made. The validity of the impressionistic method varies from very poor to rather good, depending to a large extent upon the training of the observer—whether he can remain neutral in the situation he observes, whether his background and identifications preclude the possibility of correct conclusions, and whether the results are forced in a certain direction. Since it is hard to tell who is a saint or a sinner either before or after the report is made, this method has its danger points and must be taken with at least a few grains of salt.

It must be recognized that the very source of data can be highly subjective. The reporter's biases, point of view, and previous attitudes toward similar experiences can determine what is perceived. Further, one's acceptance or rejection of the person reporting the impression often determines whether the impression is regarded as fact or fiction.

Industrial towns, factory sites, work conditions, and employee morale are often measured by the impressionistic method.

An illustration—and one intended to bring out its chief characteristic, impressions—is given below. It concerns one author's impressions of Hershey, Pennsylvania, the site of the Hershey Chocolate Corporation. The author claims the report is factual; some readers may claim that it is fiction.

Numerous visits to Hershey have led the author to conclude that this is the garden spot of America, at least as far as industrial towns are concerned. It seems to him that the employees are given much more than merely a chance to work for a living, to earn money. This small town contains more opportunity to live and enjoy oneself than does any other town of comparable—and, in many cases, much greater—size. For example, there is a community building that houses the community theater, a little theater, a junior college, a hospital, a gymnasium, a swimming pool, bowling alleys, game rooms, a social room, library, dining room, cafeteria, club rooms, and dormitories. Elsewhere in the town there is a park ballroom, an

amusement park for picnicking and riding, trout pools, a zoological garden, a rose garden, playhouses and playgrounds for children, a sports arena, an outdoor stadium, and a department store. The school system is understood to be the largest consolidated system in the country, and includes grade school, junior high, and senior high—with its academic, commercial, and vocational divisions—an industrial school, and a junior college. There is no tuition for any of this schooling. A sidelight on the school system is the program whereby 1000 orphan boys are housed on nearby farms and taught trades as well as given an education. At age 18 they receive a year's supply of clothing and $100 in addition to what they may have saved from their weekly allowances. There is a palatial hotel on the hilltop and an inn which, although much less pretentious, is clean and well-kept. Guests receive miniature Hershey bars. Everything in the town, from the factory to the office building and the homes, is spotless. The people look happy and well-fed, and the homes are in good repair. The streets are named after the various items used in the chocolate bars; the only thing that is not named Hershey is the post office.

To compare this city with some of the mining towns in the same state is like getting a glimpse of paradise and its opposite. The author's impression of Hershey, its factory, and its workers is extremely favorable. However, it is only fair to say that many people, professional psychologists included, who have had the same opportunities to observe the town do not agree with this view. To them the entire setup indicates an excessive paternalism, and they feel that the employees have been lulled into a false security and have consequently lost their zeal, ambition, and drive.

The reader may wonder about the possibility of industrial strife in such a company. The fact is that in 1937 Hershey was the scene of bloodshed. Newspaper files reveal that on March 18, 1937, Hershey signed an agreement with the United Chocolate Workers of America (CIO) recognizing the union as the collective bargaining agent for its members. On April 2, 1937, about 500 Hershey workers went on a sit-down strike; they occupied the main plant and forced a complete shutdown of the company. On April 8, 1937, the front-page headlines of the city's newspaper read: "Farmers oust 500 sit-down strikers in battle at Hershey plant. Many injured. M. S. Hershey in tears." Finally, on April 24, 1937, an election supervised by the National Labor Relations Board was held, and 1542 men voted for no union while 781 voted for the United Chocolate Workers.

Depending upon preconceived attitudes, many different conclusions can be drawn from this series of episodes. One might be that it pays to treat employees fairly; another might be that it does not. Since the authors hold no brief for either generalization—because neither is warranted—it is not necessary to draw any conclusion except to observe that the impressionistic method needs bolstering. More rigid and exacting techniques for measuring attitudes are necessary, and the remaining five methods supply them. The first two are specific types of interviewing used in the measurement of attitudes. They differ in organization and in the type of question asked.

GUIDED INTERVIEW

The second method of measuring attitudes, the guided interview, is a purposeful conversation in which the interviewer tries to obtain honest and complete answers to a specific number of questions. (See the discussion on the structured interview

given in Chapter 5.) Like all interviews, it has the advantage of face-to-face contact.

This type of interview is used most frequently in industry when considering an applicant for a job. In addition, an employer or his representative may use it in handling group complaints of workers. However, it has not been used very often in determining employee attitudes.

In the guided interview, the interviewer must abide by certain rules. He must limit his talking to the minimum, asking questions and saying a few words here and there to impress the interviewee with the importance of what the interviewee is saying. The interviewer should never argue or give advice; he should have skill in refraining from both of these. He should not express his private convictions. He must be sympathetic and encouraging but make no suggestions. The interviewer must try to have all his questions answered, and he must ask all his subjects the same questions in the same way. The questions must be fair; they must allow for an answer and should not be leading. They must not be embarrassing. A fair and complete record of the interview should be kept. Usually the best time to make this record is immediately after the interview.

The reader must be cautioned about such interviewing. It is difficult, much more difficult than one suspects. The pitfalls are many. In the hands of the inexperienced, the interview is likely to turn into a pep talk, a biased series of questions, or an argument.

The guided interview is a relatively expensive method of determining employee attitudes. Unless an experienced person conducts it, the results are likely to be as inaccurate as those obtained with the impressionistic method. To do a survey of 100 employees usually requires from one hundred to two hundred hours. This time includes planning, preparation, interviews, analysis of data, and writing the report. While it is not necessary to argue how much the services of an industrial psychologist are worth, the prevailing scale paid to consultants varies, and such a study can cost between $2000 and $4000. This cost deters a great many employers and encourages them to use the impressionistic method.

The guided interview is valuable insofar as it yields information on the specific frustrations of employees. It shows what has gone wrong with the beautiful blueprint of organization and communication.

A study by Stagner, Rich, and Britten (1941) illustrates a guided interview with the "closed" type of answer. In such an interview the respondent is asked a series of specific questions and is expected to answer with one of a number of answers that are provided. In its most simple and highly structured form, the "closed" type of answer is either "yes" or "no." A group of 159 machine tool workers were interviewed in their homes. Thirty-four brief questions were read to them, and the answers were recorded on a five-point scale from emphatic "yes" to emphatic "no." Each interview was short, lasting only about eight minutes. A numerical scoring system was used on which a number from 1 through 5 was assigned each answer. Since 19 questions, according to the authors, were related to job satisfaction, it was possible to get a total score for this attitude. Perfect satisfaction would yield a score of 19 and complete dissatisfaction a score of 95. In this study the range was from 27 to 67, with an average of 43.5 indicating that the average worker in this group was "satisfied" (neutrality would be 57).

Extremely satisfied and extremely dissatisfied workers were then selected on the basis of their scores in an effort to determine the questions that most clearly differentiated these two groups. Some of the questions used in the interview are re-

TABLE 9.1 *Factors Influencing Job Satisfaction of Machine Tool Workers*

Questions	Percentage Satisfied (N = 159)		Critical Ratio Between Extremely Satisfied and Extremely Dissatisfied Groups
	Interviewer A	Interviewer B	
1. Do you feel the factory could afford to pay more?	10	15	6.92
2. Do you like the kind of work you do on your job?	87	80	5.86
3. Do you get as much enjoyment from your work as from your spare time?	48	55	5.77
4. Are you told when you are doing a good job?	40	52	4.47
5. Do you feel allowed to offer suggestions as to methods of improvement?	74	80	3.83
6. Do you believe that the bosses and supervisors are always fair to you?	91	92	3.76
7. Do you think you could do better if given a chance at another job?	44	45	3.48
8. When you make a mistake in your work, do you always get a square deal from those deciding the case?	97	83	3.48
9. Do you feel that your present hours are too long?	67	54	3.30
10. Do you feel sure of your job as long as you do good work?	90	87	3.13
11. Do you feel your pay is fair as compared with equally important jobs in the factory?	75	80	2.96
12. How do you like your foreman?	92	85	2.95
13. Would you rather be sure of steady work at your present job than have the responsibility of being a boss?	60	63	2.28
14. Should the mill where you work be fixed up in light, heat, ventilation, etc.?	79	67	2.22
15. Do you think it makes a difference to the company that you're on the job?	52	56	1.99
16. Do you think the management should tell the men more about when the mill is going to close, reopen, or when new orders are coming through?	51	66	1.50
17. Do you feel free to carry your troubles about your work to your boss?	91	90	1.33
18. Do you feel that orders from your bosses many times disagree with one another?	44	57	0.76
19. Does the boss interfere too much in your work?	94	91	0.59
20. Are you friendly with the men who work alongside you?	96	98	0.57

SOURCE: Stagner, R., J. Rich, and R. H. Britten. Job attitudes—defense workers. *Personnel Journal*, 1941, *20*, 90–97.

ported in Table 9.1, together with the percentage of satisfied workers and the critical ratios (CR, significance of the difference between the satisfied and the dissatisfied groups defined in a study; a CR of 3 or more indicates a statistically significant difference). The main object in presenting this table is to illustrate the type of question asked in a guided interview on job satisfaction. The results are to be considered of secondary importance. (The questions are arranged by their critical-ratio value and not in the order asked.)

The table shows that although a small proportion of the workers were satisfied with their pay, this question nevertheless differentiated to the greatest extent the satisfied workers from the dissatisfied ones. Each question in Table 9.1 can be examined in terms of the way the entire population would probably answer, and thus a list of satisfactions and dissatisfactions can be constructed. It can also be analyzed in terms of the way dissatisfied workers differ from satisfied workers. For example, "pay," "liking the kind of work," and "relation of enjoyment of work and spare-time activity" successfully differentiate the two groups. However, "friendly with the men," "boss interferes," and "bosses' orders disagree" are items that do not successfully distinguish the two groups.

Another example of the guided interview technique is the work of Kornhauser (1952). While he was broadly concerned with the attitudes of Detroit people toward Detroit, at least two points in the study have particular reference here.

A total of 324 employed people rated their job satisfaction as follows:

	Percent of Satisfaction
Very satisfied	62
Fairly satisfied	35
Rather dissatisfied	2
Very dissatisfied	1

In response to the query "What do you like about your job? What don't you like about it?" the principal likes and dislikes mentioned are listed in Table 9.2.

A quotation from the study is most interesting in view of its implications:

> "Not less interesting is a related discovery Whereas 68 percent of skilled and 51 percent of nonskilled factory workers mention "inherent interest, nature of work" and the like as reasons for liking their jobs, only 38 percent of skilled and 27 percent of nonskilled workers outside of factories cite similar reasons. This tends to refute the notion that auto-plant jobs are especially robotlike, deadly or devoid of interest. At the same time, however, it is to be noted that factory workers speak much less than nonfactory workers of freedom, personal responsibility and opportunities for advancement as sources of job satisfaction. Among all the occupational groups, moreover, there are remarkably few references to chances for advancement."

Heron, working in this same area, has proposed a 15-item job satisfaction inventory (1954). The nature of the questions allows either a multiple-choice type of response on a five-point scale or a considerable degree of verbalism. By way of illustration his inventory is presented (Heron, 1954). It is a good model.

1. How do you feel you have *got on* since first coming here?
2. As a *place to work*, how does transport compare with other places in this area (or where you lived before)?
3. How much does your job give you a chance to *do the things you are best at*?
4. How fresh do you *usually* feel at the end of the day?

5. Not counting all the other things that make your job good or bad, how do you like the *kind of work* that you do?
6. How do *your mates* think this job compares with most other jobs?
7. How convenient are the hours on this job?
8. How do you find the transport department as an *employer*?
9. How well do your average earnings supply a *decent standard of living*?
10. How *interesting* is this job?
11. What is your opinion about the *speed* at which a guard has to work during *peak hours*?
12. How well is the transport department *run*?
13. How do you *like* your job?
14. How do you feel about your *prospects of advancement* in the transport department?
15. How *satisfied* are you with your job?

TABLE 9.2 *Principal Likes and Dislikes on the Job*

		Percent of 324 Employed People[a]
Things Liked About Job	Kind of work done; nature of job	48
	People I work with	32
	Pay	23
	Freedom; personal responsibility	16
	Work environment; working conditions	14
	The company; the bosses; people I work for	13
	Hours	7
	Chance for advancement	5
Things Disliked About Job	Kind of work done; nature of job	13
	Work environment; working conditions	13
	Hours	8
	Pay	7
	People I work with	6
	Irregular; not a steady job	3
	The company; bosses; people I work for	2
	No chance for advancement	2

[a] Many persons mentioned two things they liked; thus the percentages total more than 100 percent.
SOURCE: A. Kornhauser. *Attitudes of Detroit people toward Detroit.* Wayne State University Press, Detroit, 1952. Reprinted with permission from the author and publisher.

UNGUIDED INTERVIEW (NONDIRECTIVE)

The third method of determining employee attitude is the unguided interview or the nondirective interview. Although there are differences between the two terms, for purposes of industrial psychology they can be considered similar. (The reader is encouraged to refer back to the discussion of the unstructured interview in Chapter 5.)

The unguided interview is characterized by the free nature of the discussion and by the fact that it is the person interviewed who really defines its limits. There are no specific questions that the interviewer must ask; his main concern is to probe and establish the emotional content of the interview. Carl Rogers is one of the leaders in promoting this type of interview as a form of psychotherapy; the reader will benefit greatly by reading his book *Counseling and Psychotherapy* (Rogers, 1942).

The Hawthorne group (see Chapter 10) has been responsible for many advances in the application of this technique to industry. One of their contributions is a series of rules of orientation and of conduct. The rules of orientation are as follows:

1. The interviewer should treat what is said in an interview as an item in a context.
   a. The interviewer should not pay exclusive attention to the manifest content of the conversation.
   b. The interviewer should not treat everything that is said as either fact or error.
   c. The interviewer should not treat everything that is said as being at the same psychological level.
2. The interviewer should listen not only to what a person wants to say but also for what he does not want to say or cannot say without help.
3. The interviewer should treat the mental contexts described in the preceding rule as indices and seek through them the personal reference that is being revealed.
4. The interviewer should keep the personal reference in its social context.
   a. The interviewer should remember that the interview is itself a social situation and that therefore the social relation existing between the interviewer and the interviewee is in part determining what is said. The interviewer should see to it that the speaker's sentiments do not act on his own.

The rules of conduct are:

1. The interviewer should listen to the speaker in a patient and friendly, but intelligently critical, manner.
2. The interviewer should not display any kind of authority.
3. The interviewer should not give advice or moral admonition.
4. The interviewer should not argue with the speaker.
5. The interviewer should talk or ask questions only under certain conditions:
   a. To help the person talk
   b. To relieve any fears on the part of the speaker which may be affecting his relation to the interviewer
   c. To praise the interviewee for reporting his thoughts and feelings accurately
   d. To veer the discussion to some topic which has been omitted or neglected
   e. To discuss implicit assumptions, if this is advisable

These rules are not cited because they are either self-explanatory or above debate, but because they give an idea of the general conduct of the interview. Excerpts from two interviews show both the varied nature of the discussion and the underlying principles involved.

EMPLOYEE: Things went along pretty well for a long time, although at times I was a little discouraged, as during the time I was laid off for three or four months and there were quite a number of changes in supervisors; and when I came back to work in Department ——— I was very much surprised.

INTERVIEWER: How's that?

EMPLOYEE: It seems that it was my destiny to be working for a man who had been my supervisor three times before on outside jobs. We had always got along together then, but there seemed to be a certain coolness developed between us—why, I don't know—but I did my work and said nothing. His attitude toward me did not get any better and many a time I had reasonable cause for complaint, but I kept still.

INTERVIEWER: Is that so?

EMPLOYEE: Yes, he used some very abusive language at times. . . . Last year I was hit a terrible blow. My seventeen-year-old girl was taken away from me. She was sick not quite a week. She died of spinal meningitis.

INTERVIEWER: That's too bad.

EMPLOYEE: Yes, she was a dandy young lady. She would have graduated from high school this February.

My daughter's death caused my wife to have a general nervous breakdown a week after my girl was buried. That meant I had to send her to the hospital right away. In the course of her treatment in the hospital, the doctors advised me that in addition to her nervous condition she was in a very delicate condition. I could hardly believe it, but later on I was convinced. Well, my wife was in the hospital for about nine weeks and then came home.

About seven and a half months after that I was the father of twins, a girl and a boy, and the birth of twins, along with my wife's nervous condition, left her in a very bad shape. She came home from the hospital three weeks after the twins were born. She was unable to walk; in fact, she was almost an invalid. A week or two later, while my other girl who is fifteen years old went to the store and there was nobody else around, my wife made an attempt to walk, and in doing so she was so weak that she fell and knocked one kneecap out of place and injured herself internally. I had to send her back to the hospital. She was there from three to five weeks, I think, and now she is practically an invalid.

I have been advised by the doctors that what she needs the most is rest and quiet, and I am saving every penny so that I may be able to send her to a sanitarium.

Mr. Interviewer, aren't you getting tired of listening to me?

INTERVIEWER: No, indeed I am not. Go right ahead. I am very much interested.

EMPLOYEE: Well, all the time that I was having this trouble my supervisor, a man whom I worked with twelve years, treated me like a dog.[1]

Excerpts from another interview are as follows:

INTERVIEWER: You feel that there is a little politics played, is that it?

EMPLOYEE: A little? Well, I think there is a great deal of it, if you are asking me. This friendship stuff, stepping out with the boss, goes a long way around here. A blind man could see that.

INTERVIEWER: You feel that stepping out with the boss gives a person a drag?

EMPLOYEE: A drag? Say, he is sitting on top of the world. It doesn't make any difference whether he knows anything or not. He is put on a job and is sure to remain there as long as his friend remains a department head. Usually the man has ample time to get experience and with the department head coaching him along, he has probably developed himself well enough so that by the time his friend is transferred he is rather familiar with the job.

INTERVIEWER: You mentioned that you were a supervisor one time. What capacity were you in?

EMPLOYEE: I was a section head in the X department at one time. I was later made a section head in charge of the Y department.

INTERVIEWER: Were you given any reason why you were taken off this supervising work?

EMPLOYEE: No, they never told me a thing. They took me off and made me like it. That's what makes me mad. They do these damn things and they never give a fellow any explanation. They put anything they feel like down on this personnel record, and it goes upstairs and the employee never knows what is on that record. I don't see how they can do that. If they put anything on record, I don't see why the employee is not allowed to see it. I think if they would show these things to the employees, an employee would have an opportunity of correcting these wrongs if he only knew what they were. When raise time comes along, you don't get a raise and they never give you any reason why. They just tell you that you are doing a good job, to keep it up, that they are very sorry but they didn't have enough to go

[1] Reprinted by permission of the authors and publisher from F. J. Roethlisberger and William J. Dickson. *Management and the worker—an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago*. Harvard University Press, Cambridge, Mass., 1939.

around. Of course, that's very possible, that everybody can't get a raise every time, but I think they should arrange it so that certain ones would get a raise one time and the others another time. They also tell you that you are not under limit of the job, but they don't give you any more money. I can't figure that thing out.

If a fellow gets up around $50 a week, he is at a standstill. It's been two and a half years now since I have had an increase. I am working just as conscientiously as I ever did. I am always living in hope that the next time I'll get a raise. When a fellow is married and has a family, there are always certain places for your money every week.[2]

The greatest advantage of this type of interview is that when it is conducted correctly the interviewer is fairly sure of getting at what is on the worker's mind and thus is able to understand the attitudes of employees. Its disadvantages are that it is difficult to summarize and requires laborious study. It is also time consuming and costly, and sometimes presents problems that most industrial concerns believe are out of their province.

QUESTIONNAIRE

The fourth method, the questionnaire, lends itself to the mass-production techniques of determining employee attitudes. In some respects there is more similarity than difference between this method and the guided interview. For example, the study by Stagner, Rich, and Britten (1941) might just as well have been a questionnaire study. The fact that eight minutes is reported as the length of the interview means that they went at a very rapid pace. However, it may be that these authors preferred to question the workers in their homes because they felt they would get more honest answers. It is also possible that they wanted workers from various plants or were afraid they would not be given permission to conduct their survey at the plant. In any case, they could have distributed these questionnaires at one time if they could have gotten the subjects together. On the other hand, the interview usually affords an opportunity for the interviewer to observe the subject's feelings and manner of answering questions.

Although these two methods overlap, the questionnaire is more economical because one person can administer it to a large group at one time. This method also has the advantage of eliminating any effect the interview may have on the respondent and of not requiring as much experience or training on the part of the interviewer as the other methods do. It has the disadvantages of securing no more information than that provided by the answers to the specific questions and of lacking the spontaneity of the unguided interview.

The use of the questionnaire method in determining employee attitude is exemplified in Kolstad's study (1938). Kolstad constructed a questionnaire designed to measure the attitudes of certain employees in a department store toward specific items and the overall job morale. He defines morale by listing the following ten attitudes or beliefs that were expressed by employees with high morale:

1. Feels very sure of holding his job as long as he does good work
2. Has been made to feel in every way that he is really a part of the organization
3. Feels that the management does a great deal more than could be expected to maintain good working relationships between him and the people with whom he works
4. Feels that the management of this store is more interested in the welfare of the people in jobs such as his than are the other department stores in the city
5. Has never been dissatisfied with his job or if he has, such dissatisfaction was hardly ever the store's fault

[2] *Ibid.*

6. Believes that this department store treats its employees better than the other department stores in the city
7. Feels that the management is always fair with the employees in jobs such as his
8. Feels that his immediate superiors are always fair in their treatment of him
9. Can always find out whether his work is improving or not
10. Knows of no other department store in the city in which he would rather work in the same job at the same salary

Ten multiple-choice questions were used to cover these topics. Each question had five answers; the subject checked only one. A representative question was:

How much does the management do to have good working relations between you and the people with whom you work?
( ) as *little* as possible
( ) much *less* than one would expect
( ) about as much as one would expect
( ) a little *more* than one could expect
( ) a *great deal more* than one could expect

The questionnaire was scored, the possible scores ranging from + 48 to — 48. The specific items that were found to be most closely related to morale (as defined by Kolstad) were:

1. Promotion of best-qualified persons
2. Help available to get results expected
3. Encouragement to offer new ideas and suggestions
4. Fair hearing—square deal for grievances
5. Pay increase when deserved
6. Invitation to offer suggestions when plans are being made
7. Freedom to seek advice when problems arise
8. Reasons given when changes are ordered in work
9. Information about plans and results
10. No contradictory or conflicting orders

The findings of this study, shown in Table 9.3, are based upon 740 nonselling and 660 selling employees. The table is presented not in order to demonstrate the nature of morale, but rather to illustrate how scoring a questionnaire leads to quantitative results. These data indicate that the morale of the employees who sell is higher than that of employees who do not sell. Kolstad finds no significant statistical differences between the scores of men and women, married and single men, or married and single women. He does report statistically significant differences in the scores based on length of service. Thus the group employed one to five years had a lower morale than either the short-term employees or those employed over five years.

Since the main point of this chapter is to explain the principles of measuring attitudes, a minor comment may be made about the table in the original article. It is incomplete. Kolstad goes to the trouble of reporting averages and standard deviations for twenty groups but does not indicate the number in the subgroups. He merely reports the total number of employees in the selling and nonselling groups. As a result, it is impossible to check the reliability of the differences reported or to compute others. To fail to indicate the number of persons in a group is a serious error in statistical technique.

Kolstad also reports on the items investigated. He queried the employees on 34 specific items related to store pride, relations with superiors, promotion, pay, and

TABLE 9.3 *Average Morale Scores*

	Selling Employees	*Nonselling Employees*
All	22.5	10.5
Men	19.6	11.5
Women	23.6	9.2
Married men	19.0	15.4
Single men	20.1	9.7
Married women	24.2	10.7
Single women	23.3	9.0
Length of service:		
Less than 1 year	22.5	12.3
1–5 years	20.5	6.6
Over 5 years	24.6	12.6

SOURCE: A. Kolstad. Employee attitudes in a department store. *Journal of Applied Psychology*, 1938, *22*, 470–479.

factors influencing employee results on the job. The four items that he found most closely related to morale among the selling employees were: (1) promotion of best-qualified persons, (2) encouragement to offer new ideas and suggestions, (3) understanding of difficulties of job by superiors, and (4) help available to get results expected. The four items for the nonselling employees were: (1) help available to get results expected, (2) encouragement to offer new ideas and suggestions, (3) fair hearing—a square deal for grievances, and (4) promotion of best-qualified persons.

A novel questionnaire technique has been proposed by Kerr (1948). This system not only guarantees anonymity, but also makes it unnecessary to do any writing or marking on the responses. The "Tear Ballot for Industry" has eleven appropriate questions. Each question furnishes five answers, and all the person does in responding is to tear the appropriate arrowhead at the end of the answer.

Weitz and Nuckols compared the direct and indirect question technique as used in a questionnaire (1953). As an example of the indirect approach the following question was used: "Approximately what percent of the agents in your company think that the training they received was good? 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 percent." As an example of the direct approach the following question was used: "The training I received for my present job was _____ poor; _____ adequate; _____ excellent."

The authors found that the direct and indirect items correlated with each other. They found that both, to some extent, could predict the criterion which they used, that is, survival or continued employment. They also found that the direct items, in general, did a slightly better job of predicting survival, and so they see no advantage in using the indirect question.

The methodological controversy over the use of direct and indirect items as a system of gathering data will of course continue. A review of all the literature in this connection indicates that there is as much evidence in favor of one as in favor of the other. Apparently the bias of the researcher enters into the situation. Both will continue to be used; the advantages and disadvantages of each should be carefully considered in relation to the particular situation.

ATTITUDE SCALES

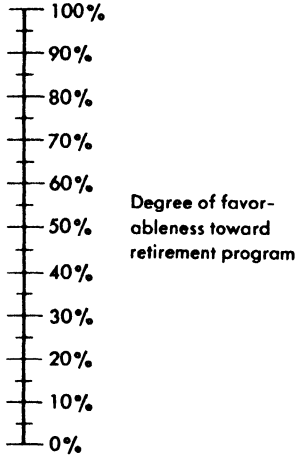
The fifth and best method of formally measuring employee attitudes is through the use of *attitude scales*. As the name implies, an attitude scale is a kind of "psychological yardstick" which can be used to measure attitudes in a quantitative manner.

There are a number of different techniques available for constructing good attitude scales. Also, attitude scales can themselves be grouped into different classes, depending upon their rationale. The following taxonomy is a useful one:

1. Rating-Scale Instruments
2. Scaled-Item Instruments
   a. Rank Order Scaling
   b. Paired Comparison Scaling
   c. Equal Appearing Intervals Scaling
   d. Successive Intervals Scaling
3. Criterion-Group Instruments
   a. Likert Scale
   b. Error-Choice Scale
4. Other Methods
   a. Guttman Scale
   b. Osgood Semantic Scale

A detailed examination of these different types of scales is not possible here.[3] However, a brief discussion of each will be useful.

RATING-SCALE INSTRUMENTS Appropriately named, these are simple scales which are given to a person to use to *rate himself* on a given attitude dimension. To illustrate, suppose you wished to measure workers' attitudes toward their retirement program. One way of doing this would be to present each worker with a scale as shown below.



Please indicate the degree to which you like your current retirement program by placing a check on the above scale.

[3] The interested reader is referred to A. L. Edwards. *Techniques of attitude scale construction*. Appleton-Century-Crofts, New York, 1957, and J. P. Guilford. *Psychometric methods*, 2nd ed. McGraw-Hill, New York, 1954.

The resulting rating is a single, global estimate of how favorable the worker is to the particular attitude object involved (in this case the attitude object was the retirement program).

Some of the major advantages of this method of obtaining a person's attitude are its simplicity and its ease of use. Its major disadvantages are the ease with which it can be faked (a person can easily falsify his attitude if he so desires) and its lack of specificity (it only gives a global assessment of the attitude without giving any more detailed information).

SCALED-ITEM INSTRUMENTS The rationale of this attitude-measuring procedure is quite simple also. In sequence, it goes something like this:

1. *A large number of statements are obtained about the attitude object.* If we wish to measure attitudes toward the retirement program, then statements might look like (a) "Our retirement program is excellent in most respects," (b) "Our retirement program is hard on certain people," or (c) "Our retirement program is hard to understand." These statements can be gathered from any number of sources, such as the employees themselves or from general discussion sessions among the staff.

2. *Each item is scaled for "favorableness."* The second step is to obtain a scale value for each statement which expresses how favorable that statement is toward the attitude object. Thus statement (a) above, which seems to be saying something nice about the retirement program, would probably end up with a high scale value, while statements (b) and (c), which are less flattering, would probably end up with lower scale value.

In other words, this step is simply one of attempting to determine exactly how favorable each statement is toward, in this case, the retirement program. In Chapter 7 several different scaling methods were discussed in connection with determining the relative merits (scale values) of individuals. These same techniques are applicable here. Thus one can use (1) the rank-order method, (2) the paired-comparison method, (3) the equal appearing interval method, or (4) the successive-interval method, depending upon one's own preference.

3. *Selection of best statements for final instrument.* After all items have been scaled, the best items are selected for use in the final attitude measuring instrument. Statements covering the entire range of scale values are included, and statements where there was high agreement on the scale value are preferred.

4. *Use of instrument to measure attitudes.* The final form of the instrument is now ready for use. The person responding to the attitude questionnaire is simply asked to check those statements with which he agrees. His score is the median scale value of the items with which he agrees.

An example of such a scale is provided by Uhrbrock (1934) who has developed a scale to measure the attitude of employees. The items in it and the values assigned to each item are shown in Table 9.4.

On this scale 3934 factory workers obtained an average score of 6.34. Ninety-six clerks averaged 6.84, and 400 foremen had an average score of 7.19. Care must be exercised in interpreting these results, lest one come to the conclusion that the attitude of the typical factory worker is exemplified by the statement "I think a man should go to the hospital for even a scratch, as it may stop blood poisoning." After all, this statement has a scale value of 6.3 and is closest to the average of 6.34. But such a conclusion is of course foolish, because 6.34 is the average of the mathematical weights of all the statements checked by the employees.

TABLE 9.4 *Statements Used in Uhrbrock's Scale for Measuring Attitude of Employees Toward Their Company*

Scale Value	*Statement*
10.4	I think this company treats its employees better than any other company does.
9.5	If I had to do it over again, I'd still work for this company.
9.3	They don't play favorites in this company.
8.9	A man can get ahead in this company if he tries.
8.7	I have as much confidence in the company physician as I do in my own doctor.
8.5	The company is sincere in wanting to know what its employees think about it.
7.9	A wage incentive plan offers a just reward for the faster worker.
7.4	On the whole, the company treats us about as well as we deserve.
6.3	I think a man should go to the hospital for even a scratch, as it may stop blood poisoning.
5.4	I believe accidents will happen, no matter what you do about them.
5.1	The workers put as much over on the company as the company puts over on them.
4.4	The company does too much welfare work.
4.1	Soldiering on the job is increasing.
3.6	I do not think applicants for employment are treated courteously.
3.2	I believe many good suggestions are killed by the bosses.
2.9	My boss gives all the breaks to his lodge and church friends.
2.5	I think the company goes outside to fill good jobs instead of promoting men who are here.
2.1	You've got to have "pull" with certain people around here to get ahead.
1.5	In the long run this company will "put it over" on you.
1.0	The pay in the company is terrible.
0.8	An honest man fails in this company.

SOURCE: R. S. Uhrbrock. Attitudes of 4430 employees. *Journal of Social Psychology*, 1934, *5*, 365–377.

In fact, Uhrbrock reports that statements expressing a favorable attitude toward the company were checked by more than twice as many workers as checked the unfavorable statements. In addition to finding that foremen had more favorable attitudes than factory workers, he found that this was true more of women than of men, and that employees who had worked for the company more than six years were slightly more favorably inclined than those with a shorter period of service. It will be noted that Uhrbrock finds a difference in attitude between males and females, whereas Kolstad does not. This is due to different samples, different levels of employment, and other similar factors.

CRITERION-GROUP INSTRUMENTS The third class of attitude-measuring devices consists of those which are constructed using the standard test construction (item-analysis) procedures discussed in Chapter 2. This method also involves a sequence of events.

1. *Collection of items.* (Same as previously discussed.)
2. *Item analysis against a criterion.* All items are examined to see if they discriminate significantly between a group of individuals which is favorable toward the attitude object (Group F) and a group which is unfavorable (Group U). If the item does so discriminate, it is kept on the final scale.

The two major versions of this type of attitude scale are the *Likert Scale* and the *Error-Choice Scale*. A Likert item would look like this:

(A) Our retirement program is excellent in most respects.

a.	Strongly disagree	(1)
b.	Disagree	(2)
c.	Indifferent	(3)
d.	Agree	(4)
e.	Strongly agree	(5)

Score for each alternative

The person reads the item and then selects one of the alternatives. Each alternative has a score or weight associated with it. A person's score on the final attitude scale is simply the sum of the weights of the alternatives he has checked. Weights are usually assigned so that high scores indicate favorable attitudes.

The following example illustrates this method. Richardson, Bellows, Henry and Company (see Chapter 1) did a considerable amount of employee attitude measurement for its clients. As an example, for one such study (Stagner, Rich, and Britten, 1941) 94 statements were prepared based upon preliminary interviews with a sample of employees' discussions with management and upon general considerations based on experience. These statements were intended to cover ten specific areas:

Area	*Statement*
The company	1. Headquarters gives proper attention to divisional operations.
	2. Management adds equipment and facilities as needed.
Working conditions	3. The times for starting and ending our workday are satisfactory.
	4. Lunch facilities are adequate.
Pay	5. The general plan for reviewing salaries and merit raises is good.
	6. My job pays what it should, compared with similar jobs in other divisions of the company.
Benefits	7. Hospital-medical-surgical insurance coverage is satisfactory.
	8. I have as much information as I need about the pension plan.
Future opportunities	9. If business goes well generally, I feel sure of holding my job.
	10. There is adequate provision here for developing employees for future supervisory jobs.
Personal satisfaction	11. I enjoy my work.
	12. My job gives me a chance to do those things I can do best.
Organizational efficiency	13. I always know who is in charge of my work.
	14. Higher-ups don't pass the buck if I go to them about decisions for problems.
Supervisor	15. My supervisor lets me work things out in my own way.
	16. My supervisor is well-liked by the people who work for him.
Teamwork	17. People get along well with each other in my section.
	18. In my section, people do not try to get ahead at another's expense.
Communications	19. I am kept informed on what goes on here.
	20. Other sections regularly tell us what we need to know about their work.

Employees responded by indicating: (1) definitely agree, (2) inclined to agree, (3) inclined to disagree, or (4) definitely disagree. All questionnaires were filled out, and anonymity was guaranteed. The results are presented in Figure 9.1. They compare the respondents who respond favorably in each of four categories: supervisory, nonsupervisory, in the company, and in other companies.

The second major version of criterion-group scales is *the Error-Choice Attitude Scale,* originally proposed by Hammond (1948). It is somewhat similar to the Likert Scale in that (1) the items are presented in multiple-choice format and (2) only items which statistically discriminate between high and low criterion groups are included on the final form. However, it differs from the Likert in one very important respect—*it is disguised to look like a test of general knowledge rather than a test of attitude!*

Each item in an error-choice attitude scale asks a factual question and includes usually two or four alternatives. The catch is that *none* of the alternatives are correct! Half are underestimates of the "true" or correct answer and the other half are overestimates. The rationale goes something like this:

1. Even though a person is given an item without the correct answer he is still forced to respond. That is, he is required to select one of the incorrect alternatives by the very nature of the test, even if he is reasonably sure that none of them are correct.
2. When forced into picking an incorrect answer, the decision to overestimate or underestimate will be related to an individual's attitude.
3. Thus by seeing if a person continually overestimates or underestimates on factual items one can measure attitudes.

Perhaps an illustrative item would be helpful to demonstrate. Suppose one wants to measure attitudes toward women drivers. Suppose also that we know that there is no difference in the number of accidents experienced by female drivers and male drivers when frequencies are adjusted for amount of miles driven. We could then write an item which might look like this:

> The number of accidents per mile driven for women is what ratio to the number of accidents per mile driven for men? That is, how much more often do women have accidents than men after one adjusts for mileage?
>
> (a) 2 female to 1 male accident (4)
> (b) 1½ female to 1 male accident (3)
> (c) 1 female to 1½ male accidents (2)
> (d) 1 female to 2 male accidents (1)

What do you think would happen if this item were given to a random sample of 100 men and 100 women? In all likelihood the men would get higher scores than the women, since they would be more apt to bias their responses toward (a) and (b), while the women would be more apt to select (c) or (d).

The problem of faking on the attitude scale is virtually eliminated with the error-choice method, simply because the testee does not know his responses are being used to reflect his attitude. To him, it looks like his knowledge about driving and accidents is being measured—not his attitude toward women drivers.

Weschler (1951) has demonstrated the use of this technique by constructing a test to measure attitude toward labor-management relations. While 24 of the items were real, 16 were of the error-choice type. An illustration of the type of question he

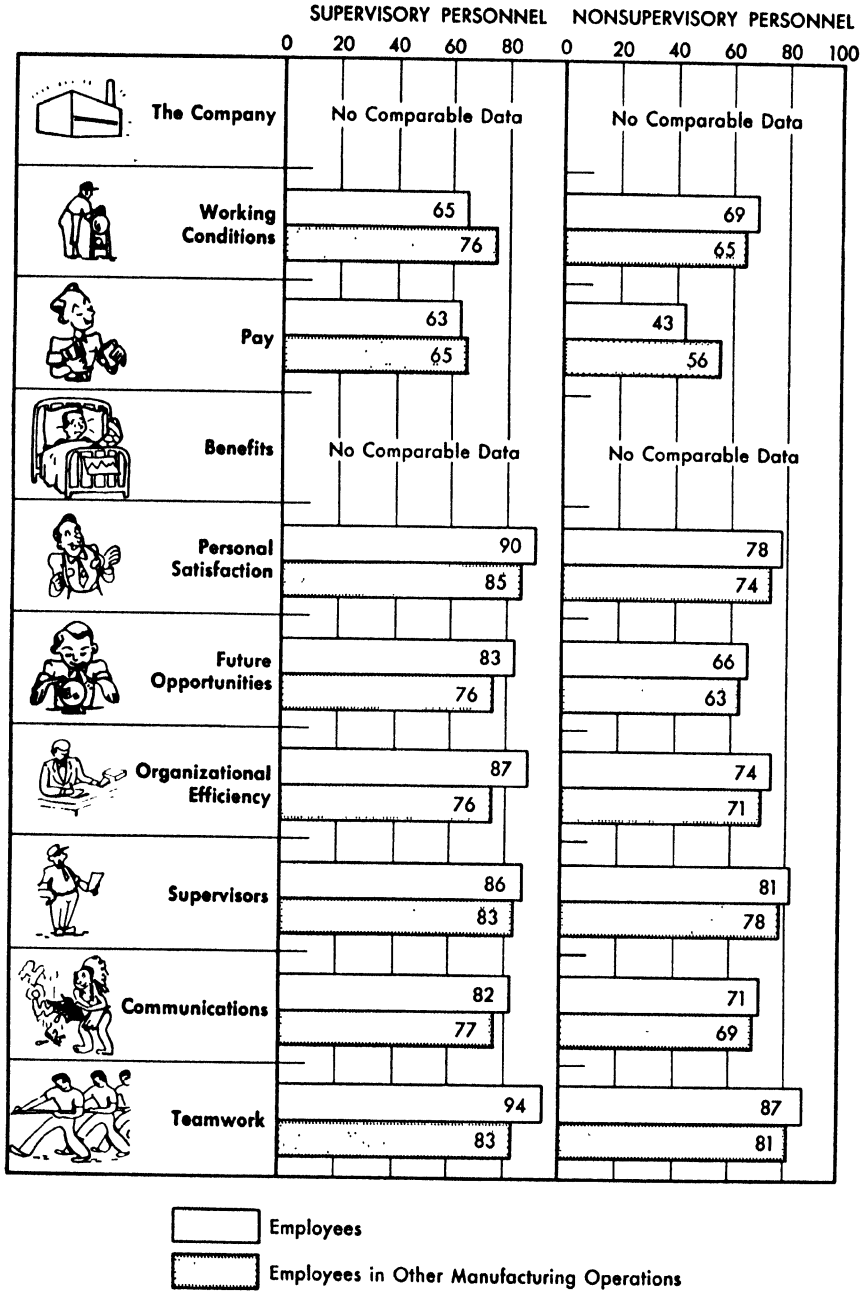


Figure 9.1. Comparison of questionnaire category responses with data from other surveys. (Courtesy Richardson, Bellows, Henry & Co., Inc. From R. Stagner, J. Rich, and R. H. Britten. Job attitudes—defense workers. *Personnel Journal*, 1941, 20, 90–97.)

used is: "At present, the following percentage of people in the United States are entirely dependent upon jobs and have very few savings: (a) about 55 percent; (b) about 85 percent." (Note—correct answer is 70 percent.)

OTHER SCALING METHODS Guttman (1950) has proposed a scaling method where the items have a special cumulative property. For example, the respondent who responds positively to the eighth item of a ten-point scale will have responded positively to all other seven items; similarly, the one who responds to the fourth item will have responded positively to only the first, second, and third. In a perfect Guttman scale, an individual's score bears a one-to-one relationship with his response pattern. In practice the construction of a perfect scale, using this method, has many difficulties because of the multiplexity of attitude dimensions. However, the scale does present an interesting methodological departure from the Thurstone and Likert methods.

The final scaling method of attitude measurement to be discussed has been proposed by Osgood, Suci, and Tannenbaum and is known as the Semantic Differential (1957). The method requires the respondent to rate on a scale of, let us say, four or six or eight points, the associate meaning of a series of descriptive polar terms. Examples of such polar terms are innumerable: rough-smooth, weak-strong, small-large, tense-relaxed, wet-dry, fresh-stale, cold-hot, fair-unfair, etc.

A valuable reference for those interested in building attitude scales in this area is the work of Uhrbrock (1961). He has gathered over 2000 short descriptive statements which have been given scale values. This material can serve as a source to those who need to construct such scales.

INDIRECT METHOD

The sixth method of attitude measurement has been described as the indirect method (Weschler and Bernberg, 1950). It is intended to provide a more free rein of expression. The objective is to explore the "deeper levels rather than to deal only with the manifest verbal content." This method deliberately attempts to conceal the intent of the measurement and allows the experimenter to observe and measure without producing an effect on the attitude itself. Varieties of techniques have been included within this category: word associations, sentence completions, or picture and story theme completion.

Weschler has raised four interesting questions as a result of the indirect technique (1951) which point up the problem as to whether this is trickery or scientific method: (1) Do I have the right to investigate other people's attitudes? (2) Do I have the right to "deceive" (authors' quotes) people in order to get at their attitudes? (3) Do I have the right to report on new indirect attitude-measuring devices at a time when these can be used by unscrupulous politicians or by other selfish interests? (4) What is my responsibility for seeing that the findings which I report are properly interpreted?

These are significant questions and must be answered. Admittedly, the answers reveal the authors' attitude—and without a survey. The answer is an unequivocal "yes" to question 1—we have the right to investigate attitudes. The answer is "yes" to question 2, except that in our opinion the indirect method is not deception. The false answers willingly given by the respondents when direct techniques are used are the deceptions. As for question 3, we will always have unscrupulous politicians

and selfish interests, so why wait for Utopia in order to do worthwhile research? In answer to question 4, we believe that our research responsibility is great, but in a free society and with a press interested in circulation one may expect others to misinterpret. This should not deter the scientist but rather should encourage him to communicate more directly with the public.

The indirect method of measuring attitude is the newest development, is the most subtle, and conceivably has dangerous implications in the hands of the unscrupulous. Nevertheless, it is a technique and will be used. Little good can come from wishing that the H-bomb were not in existence. More good can come from knowing of its existence and trying to work within its framework of possible destruction. In a much smaller and possibly insignificant way, let us not hide our heads in the sands with reference to this technique of attitude measurement.

Friesen (1949) has developed an incomplete sentence blank which is an attempt to standardize the indirect-method technique of measuring employee attitudes. A rather novel approach has been suggested by Baumgarten (1952). She has collected a large number of proverbs concerning human, labor, and social relations. The testee selects those proverbs which he believes to be correct and incorrect. While this technique theoretically could evidence attitude, for the time being it has to be regarded as interesting and speculative.

The Evans and Lascau (1950) research known as "My Job Contest" (MJC) conducted at General Motors is an example of research using the indirect method of attitude measurement. On the surface, this was a letter-writing contest on the topic "My Job and Why I Like It." Five thousand prizes were awarded, including such items as a Cadillac and other General Motors cars on down to rear-view mirrors. A tremendous amount of employee participation occurred—almost 50 percent of the 297,401 eligible employees entered. The letters varied in length from one hand-written sentence to twenty typewritten pages. About 700 letters were written in languages other than English.

Management recognized that this served a greater purpose than merely a letter-writing contest. It presented the opportunity to analyze thematically the relatively unstructured reflections of employees. It was, therefore, an indirect method to measure attitude. Analysis of the content of the letters resulted in establishing 58 themes and formed the basis for reporting to each division the findings about its employees' attitudes toward their job and related conditions. Although the analysis made use of rather exacting statistical techniques and recognized necessary procedural controls, it was communicated to management in an interesting and nontechnical manner. This is very important as a general principle. Too many industrial psychologists become so technical that they lose even their colleagues, let alone the management people who must understand and translate the findings into action.

ATTITUDE SURVEYS AND THEIR APPLICATION

Too often an attitude survey performed in industry, by whatever method chosen, is regarded as an interesting piece of literature that some people in management ought to read the first chance they get. This is money down the drain. Attitude survey results must be put to use and necessary changes introduced.

Executives must be prepared to face the real possibility that the attitudes they *believe* employees have may not coincide with *actual* employee attitudes. They must recognize that this difference or the revelation of the unexpected is not a threat to personal integrity.

Bradshaw and Krugman (1948) have proposed coupling attitude surveys with conferences. In one company they had a team of two members from each of six levels of management participate in a series of conferences. The conferences began prior to the interviews and continued for a discussion of the report findings. Within six months after the completion of the survey, action had been taken on the following projects: a supervisory training program, a revision of promotion policies, a foreman selection program, a reorganization of personnel and work procedures in a major department, a revision of the suggestion system, and a program of providing employees with various types of factual information.

Baumgartel (1953) reports an interesting experiment which clearly shows that a discussion of attitude survey results leads to substantial positive change. The experimental groups with the feedback discussions had decidedly more favorable attitudes. The control group and the experimental group with no discussion remained the same.

THE ATTITUDES OF EMPLOYEES VERSUS THE ATTITUDES OF EMPLOYERS

A 1939 report in *Personnel Journal* indicated clearly that employers and employees have different ideas on the subject of morale. Table 9.6 shows the rank assigned to various morale factors by these two groups.

TABLE 9.5 *Rank Assigned to Morale Factors*

Item	*Employee Ranking*	*Employer Ranking*
Credit for all work done	1	7
Interesting work	2	3
Fair pay	3	1
Understanding and appreciation	4	5
Counsel on personal problems	5	8
Promotion on merit	6	4
Good physical working conditions	7	6
Job security	8	2

SOURCE: Employee Relations Bureau, National Retail Dry Goods Association. Your employer wants you to know him. *Personnel Journal*, 1939, *17*, 357.

Inspection of this table shows that there is room for misunderstanding between the two groups. Thus employees rate "credit for all work done" and "counsel on personal problems" as more important than employers do, whereas the latter rate fair pay and security as the two most important items. It is also interesting to note that both groups rate good physical working conditions as relatively unimportant, and yet a great amount of money is spent on such things as air conditioning, light

control, oil-absorbing floors, and music—and very little on attitude measurement. More will be said about this later; here it is necessary only to say that employers continually underestimate the importance of attitude on the assumption either that they know employee attitudes or that the attitudes of their employees are similar to their own.

Not many studies of employer attitudes have been made; work is necessary in this field. Even though the content of this chapter might seem to justify the title "Employee Attitudes" rather than the one used, the title was chosen in order to drive home the point that not only can attitudes be measured by the six methods described, but these methods can be used to measure *employer* attitudes as well. The evidence indicates that both groups have attitudes toward each other that differ from their attitudes toward themselves. A clear exposition of this fact by a skillful technician would lead to greater mutual understanding. At the present time industry, as presented by employer and employee, lacks such insight.

One of the rare studies of executive attitudes is Cherington and Bergen's interesting report (1941) of the attitudes of 51 senior executives in a large company. Prior to undertaking a study of the attitudes of its employees and supervisors, this company surveyed the attitudes of its 51 senior executives by means of a questionnaire. The questionnaire was "designed to bring out specific attitudes toward the various points of relationships between superiors and associates, the plan of organization, and the effectiveness with which its various units had been coordinated, impartiality of executives, compensation, fairness in selecting junior executives for promotion, the effectiveness of executive training, union relations, and the like." In addition, the questionnaire was designed to measure the general morale of this group and their "total" attitudes toward their work and the company. The item-checking method of filling out the questionnaire was used, and precautions were taken to ensure anonymity.

The results show that the most unfavorable specific attitude held by this group concerned their compensation. Over 50 percent felt that the salaries for their positions were lower than those for other comparable positions in the company. Another source of dissatisfaction was the promotion procedure, as evidenced by the fact that less than one-third believed that the best man always received the promotion. These executives felt a strong need for a clear-cut organization plan and for more effective coordination of the various executives. Some respondents showed lack of any clear understanding of their own duties, responsibilities, and authority and those of their fellow executives. Only half of them believed that certain specialized departments—such as personnel, industrial engineering, and the laboratory—helped them in discharging their duties. Only half the group felt that they were really part of the management of the company. The group expressed a desire for more intensive executive training and a need for information which they were not getting—on plans, the company's long-range programs, general business conditions, and development of labor relations in the company.

The executives ranked ten selected personnel practices in the following order of relative importance:

1. Fair pay
2. Clear-cut definition of duties, responsibilities, and authority
3. Promotion on merit
4. Credit for work accomplished

5. Job security
6. Understanding leadership by superiors
7. Adequate job instruction and related information
8. Prompt and fair adjustment of grievances
9. Fair layoff procedure
10. Adequate retirement benefits

Incidentally, the general morale of the group was good. A score of from 0 to 100 was possible. The average was 77, with a range of 55 to 97.

Two important points are revealed in this study. The first shows that there is room for misunderstanding because the employee's attitude and that of the employer differ. For example, an employee is not ordinarily inclined to think of senior executives as being concerned about equality of salary. The average employee believes an executive earns so much money that such problems cannot exist for him. How often does an employee think of a senior executive as feeling a strong need for clear-cut organization and for a definition of duties, responsibility, and authority? How many employees realize that a senior executive may not consider himself part of the management of the company? The second point shows that the employer's impression of the employee's attitude is not well grounded. This study indicates that an executive projects upon the employee his own evaluation of the importance of money. Fair pay ranks first in the list of personnel practices. Similarly, in Fosdick's study the employers rank fair pay first. Yet Fosdick's study and other studies of employee attitude show that employees do not rank this first. The attitude toward money will be considered more fully later. Here it is presented merely as food for thought.

REFERENCES

Abelson, R. P., and M. J. Rosenberg Symbolic psychologic A model of attitudinal cognition *Behavior Science*, 1958, 3, 1–13.

Baumgartel, H. The survey feedback experiment· A study of a program for the use of attitude survey data in a large organization. Survey Research Center, University of Michigan, Ann Arbor, 1953 (Summary of Technical Report).

Baumgarten, F. A proverb test for attitude measurement. *Personnel Psychology*, 1952, 5, 249–261.

Bradshaw, F. F., and H. E. Krugman. Making the most of attitude surveys. *Personnel*, August 1948.

Brown, R., E. Galanter, E. H. Hess, and G. Mandler. *New directions in psychology*. Holt, New York, 1962.

Cherington, P. T., and H. B. Bergen. What do the bosses think? *Advanced Management*, 1941, 6, 66–69.

Edwards, A. L. *Techniques of attitude scale construction*. Appleton-Century-Crofts, New York, 1957.

Evans, C. E., and L. N. Laseau. My job contest. *Personnel Psychology Monograph #1*, Washington, D.C., 1950.

Festinger, L. *A theory of cognitive dissonance*. Harper & Row, New York, 1957.

Fosdick, S. J. Your employee wants you to know him. *Personnel Journal*, 1939, 17, 357.

Friesen, E. P. The incomplete sentences technique as a measure of employee attitudes. *Personel Psychology*, 1949, 5, 329–340.

Guilford, J. P *Psychometric methods*, 2nd ed. McGraw-Hill, New York, 1954, 195.
Guttman, L. The basis for scalogram analysis. In Stouffer, *et al. Measurement and predictions.* Princeton University Press, Princeton, N J., 1950.
Hammond, K. R. Measuring attitudes by error-choice: An indirect method. *Journal of Abnormal and Social Psychology*, 1948, *43*, 38–48.
Heron, A. Satisfaction and satisfactoriness complementary aspects of occupational adjustment. *Occupational Psychology*, 1954, *28*, 140–153.
Kelman, H. C., and A. H. Eagly. Attitude toward the communicator perception of communication content and attitude change. *Journal of Personality and Social Psychology*, 1965, *1*, 63–78.
Kerr, W. A. On the validity and reliability of the job satisfaction tear ballot. *Journal of Applied Psychology*, 1948, 32, 275–281.
Kish, L. *Selection of the sample in research methods in the behavioral sciences.* Edited by L. Festinger and D. Katz. Holt, New York, 1953
Kolstad, A. Employee attitudes in a department store *Journal of Applied Psychology*, 1938, 22, 470–479.
Kornhauser, A. *Attitudes of Detroit people toward Detroit* Wayne State University Press, Detroit, 1952.
Krech, D., R. S. Crutchfield, and E. Z. Ballachey. *Individual in society.* McGraw-Hill, New York, 1962.
Osgood, C. E., J. Suci, and P. H. Tannenbaum. *The measurement of meaning.* University of Illinois Press, Urbana, 1957.
Richardson, Bellows, Henry & Co Survey of attitudes of salaried employees. New York (unpublished, 1954).
Roethlisberger, F. J., and W. J. Dickson. *Management and the worker—an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago.* Harvard University Press, Cambridge, Mass., 1939.
Rogers, C. *Counseling and psychotherapy* Houghton Mifflin, Boston, 1942.
Stagner, R , J. Rich, and R. H. Britten. Job attitudes—defense workers. *Personnel Journal*, 1941, *20*, 90–97.
Uhrbrock, R. S. Attitude of 4430 employees. *Journal of Social Psychology*, 1934, *5*, 365–377.
Uhrbrock, R S. 2000 scaled items. *Personnel Psychology*, 1961, *14*, 375–420.
Weitz, J., and R. C. Nuckols The validity of direct and indirect questions in measuring job satisfaction. *Personnel Psychology*, 1953, *6*, 487–494.
Weschler, I. R. Problems in the use of indirect methods of attitude measurements. *Public Opinion Quarterly*, spring 1951, 133–138.
Weschler, I. R , and R. E. Bernberg. Indirect methods of attitude measurement. *International Journal of Opinion and Attitude Research*, 1950, *4*, 209–229.

10 THE HAWTHORNE STUDIES

The Hawthorne Studies are referred to because of their scope, significance, inclusiveness, and design. They form an interesting model. Although these experiments were conducted years ago, the passage of time has allowed them to take on the flavor of a classic. In the authors' opinion they represent the most significant research program undertaken to show the enormous complexity of the problem of production in relation to efficiency. Just as Hugo Münsterberg in his *Psychology and Industrial Efficiency* set the stage in 1913 for the work of psychologists in industry, so the Hawthorne Studies, carried on over a period of 12 years at the Hawthorne Works of the Western Electric Company, can be considered as "starting the show." They were begun in 1927.

In order to understand how these studies could pass from one type of experiment to another without the forced conclusions and forced pace of so many studies in this field, we need only quote Elton Mayo, professor of industrial research at Harvard and guiding star of the Hawthorne Studies.

> Researches of this type are usually impossible because of a foolish convention that institutions engaging in industrial research are expected to "pay their way" or "earn their keep." This means, in effect, that any such institution, living from hand to mouth, is committed to the futility of endless repetition of some former discovery. The interesting *aperçu*, the long chance, may not be followed; both alike must be denied in order that the group may "land another job." This confusion of research with commercial huckstering can never prosper: the only effect is to disgust the intelligent youngster who is thus forced to abandon the quest for human enlightenment [1]

The series of experiments that comprise the Hawthorne Studies are significant because those responsible for the work were able to investigate many of the dependent variables found in human experimentation, especially those among workers. Ordinarily the assumptions made as to the independent variables preclude the possibility of valid findings. Probably the most significant results of these studies are the fact that workers are affected by factors outside the job to an even greater extent than by those on the job itself, and that they organize into informal social groups. These organizations take precedence over management-employee organizations and determine production to as great an extent as do changes of a job-environment nature. The disregard of "outside the job" factors and employee self-grouping has led many studies conducted by management to erroneous conclusions.

[1] Reprinted by permission of the authors and publishers from F. J. Roethlisberger and William J. Dickson. *Management and the worker—an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago.* Harvard University Press, Cambridge, Mass., 1939.

Although the Hawthorne Studies are not to be considered the acme of perfection or even a model, they nevertheless represent a vast improvement over all the other work that had been done in the field. It is for this reason that they are reviewed in their entirety in this chapter. The serious student of industrial psychology will gain from them a panorama of the field. He will be able to understand the need for experimentation, and he will see the difficulties of it. He will note that verbalisms are poor substitutes for facts. Further, he will acquire a sense of continuity. The Hawthorne Studies show the complete interrelatedness of the various problems and demonstrate that changes in work environment, rest pauses, hours of work, hours in the working week, fatigue, monotony, incentives, employee attitudes, employee organization both formal and informal, and employee-employer relations are all intimately related. To treat them as if they were separate is to introduce such artificiality as to make the setup unreal.

The Hawthorne Studies can be conveniently divided into five major parts, one of which has two subdivisions. Each part is an outgrowth of the preceding one. In many respects each follows the other logically, but no one could have predicted the complex nature of the findings or the ramifications of procedure from the simple, unpretentious beginning. The five studies are referred to as:

1. Experiments on Illumination
2. Relay Assembly Test Room
   a. Second Relay Assembly Test Room
   b. Mica Splitting Test Room
3. Mass Interviewing Program
4. Bank Wiring Observation Room
5. Personnel Counseling

STUDY 1. EXPERIMENTS ON ILLUMINATION

The first experiment on illumination was conducted in three selected departments. The work performed in one department was inspection of small parts; in the second department relays were assembled; and the job in the third was winding coils. For the control situation, all employees worked under existing lighting installations so that production could be measured.

In the first department the various levels of average illumination intensity were 3, 6, 14, and 23 foot-candles. The production of the workers varied *without* direct relation to the amount of illumination.

In the second department the illumination intensities were 5, 12, 25, and 44 foot-candles. Production increased, but not solely as a result of changes in illumination. The third department showed similar results. The conclusions drawn "brought out very forcibly the necessity of controlling or eliminating the various additional factors which affected production output in either the same or opposing directions to that which we can ascribe to illumination" (Snow, 1927).

Here, apparently, was a problem that was not as simple as it looked. A second experiment with more refined techniques was set up; it took place in only one department and two groups of workers participated. These groups were equated for numbers, experience, and average production. The control group worked under

relatively constant illumination. The test groups worked under three different illumination intensities. The influence of any spirit of competition (a factor not part of the experiment) was guarded against by having the two groups work in different buildings.

This second experiment on illumination resulted in perplexing but enlightening results. *Both* groups increased production appreciably and to an almost identical degree. Since this experiment did not show what increase in production could be attributed to illumination, the third experiment was undertaken.

In this experiment further refinements in procedure were introduced. Only artificial light was used, daylight being excluded. The control group worked under a constant intensity of 10 foot-candles. The test group began with 10 foot-candles but the intensity was reduced by 1 foot-candle per period until they were working under only 3 foot-candles. This group of employees maintained efficiency to this point despite the discomfort and handicap of insufficient illumination. In a fourth experiment two volunteer girls worked in a light-controlled room until the intensity equaled that of ordinary moonlight. At this stage they maintained production and reported no eyestrain and less fatigue than when working under bright lights.

The fifth and last of the illumination experiments was conducted with the girls who wound coils. During this experiment there was no real change in production. At first the intensity of the lights was increased daily, the girls reporting that they liked brighter lights. An electrician then changed the light bulbs but kept the same intensity. The girls commented favorably on the increased illumination. In the last part of the experiment the illumination was decreased; the girls said that less light was not so pleasant. However, they felt the same way when the lights remained constant, even though the electrician was supposedly reducing the illumination.

In most such experimentation the sponsors would have thrown out the evidence and the "crackpots" responsible; it would have been considered a wild nightmare, to be repressed and suppressed. Fortunately in this instance, although the specific problem was not solved, there was an awareness that more knowledge concerning the problems involving human factors was essential. The Relay Assembly Test Room, the second of the Hawthorne Studies, was the result.

STUDY 2. RELAY ASSEMBLY TEST ROOM

The initial purpose of this experiment, which lasted about five years, was to exercise more direct control over the many variables that could have influenced the performance of the operators in the first study.

In order to exercise greater control it was decided to use a small group of employees in a separate room away from the regular working force. The task chosen was the assembly of small relays, because it is simple and highly repetitive (approximately 500 a day), requires no machinery, and permits an accurate measure of production.

Two experienced operators who were friendly with each other were invited to participate; they in turn selected three other assemblers and the layout operator, who assigns the work and procures the parts. The situation was similar to that in the regular Relay Assembly Department except that in the regular department there was one layout operator for six or seven girls. The only other person in the room

was a man who had worked on the illumination experiment. He was to keep records of what occurred and create and maintain a friendly atmosphere. A special room was fitted out, but chairs, fixtures, and work layout were similar to those in the regular department. An accurate recording device was added to the regular equipment

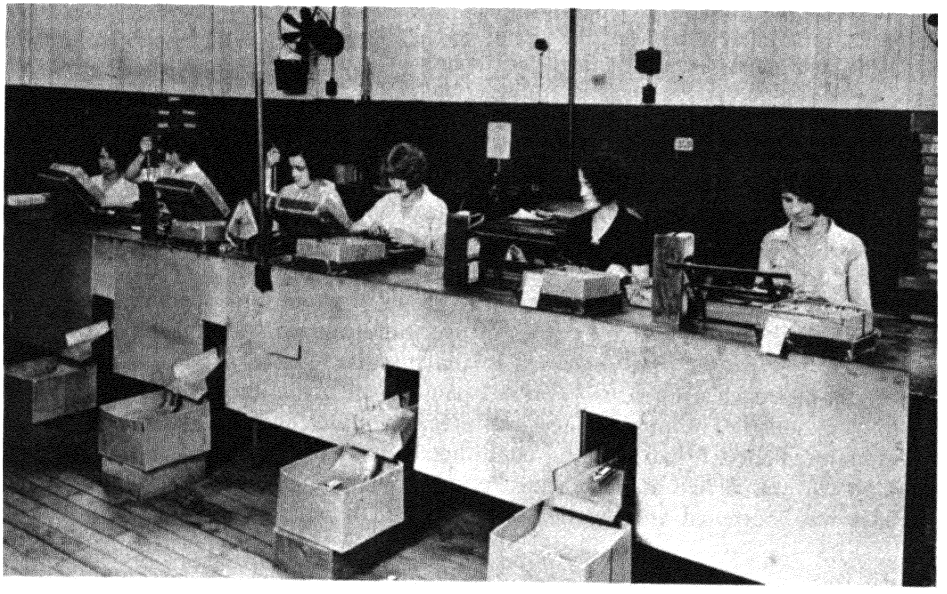


Figure 10.1. Relay assembly test room operators. (See p. 311, footnote 2.) Photograph taken during Hawthorne experiment.

so that it was possible to measure not only the number of relays assembled but also the time taken per relay. Temperature readings and humidity were recorded hourly. The girls took physical examinations every six weeks.

A complete daily history was kept to give an accurate account of what went on in the test room as changes were introduced. The following is an excerpt from this history:

Monday, June 13, 1927
- Operator 1A. Said she was tired today.
- Operator 2A. Tired also and said her head ached.
- Operator 3. Was asked if she thought she did more, less, or about the same amount of work. Ans. "More, I'm almost up to Operator 4 and I have a bigger relay."
- Operator 4. "I feel fine today, not tired or anything."
- Operator 5. "I'm tired today and sleepy."

Tuesday, June 21, 1927
- The foreman informed the group of their low activity for the past week. The weather was more favorable for work, cloudy and raining.
- Operator 1A. "I feel fine today. Just right for work."
- Operator 2A. "Today is fine for work."
- Operator 3. "I went to bed at 9 o'clock last night and feel O.K. today."
- Operator 4. "I feel great today."
- Operator 5. "A day like this is much better for work than yesterday."

Study 2 started with six questions:

1. Do employees actually get tired out?
2. Are rest pauses desirable?
3. Is a shorter working day desirable?
4. What are the attitudes of employees toward their work and toward the company?
5. What is the effect of changing the type of working equipment?
6. Why does production fall off in the afternoon?

To answer these questions 13 test periods were introduced.

The first test period made it possible to get an accurate measure of each girl's production under typical work conditions. The second test period attempted to determine the effects of a change in work place upon production. The third period introduced a change in the method of payment; the six girls were paid directly according to the output of their own group, whereas previously they had been paid on the basis of the production of the entire group of about 100 operators.

In the fourth period two rest pauses of five minutes each were introduced. In the fifth, these rest pauses were increased to ten minutes. Period 6 had six rest pauses of five minutes each. In period 7 two rest pauses were given and a "free midmorning lunch" was added. The morning rest was fifteen minutes and the afternoon one was ten minutes. Period 8 had working conditions similar to those in period 7, but the working day was a half-hour shorter. Period 9 was the same as period 8, but the workday was shortened another half-hour. Period 10 returned to the conditions in

TABLE 10.1 *Relay Assembly Test Room Data*

Period	*Length in Weeks*	*Experimental Conditions of Work*	*Experimental Working Hours per Week*[a]	*Percent Decrease from Standard Working Hours*
1	2	Standard	48	
2	5	Standard	48	
3	8	Standard	48	
4	5	Two 5-min. rests	47:05	1.9
5	4	Two 10-min. rests	46:10	3.8
6	4	Six 5-min. rests	45:15	5.7
7	11	15-min. rest and lunch in morning		
		10-min. afternoon rest	45:40	4.8
8	7	Same as 7, but 4:30 stop	43:10	10.0
9	4	Same as 7, but 4:00 stop	40:40	15.3
10	12	Same as 7	45:40	4.8
11	9	Same as 7, but Sat. morning off	41:40	13.2
12	12	Standard	48	
13	31	Same as 7	45:40	4.8

[a] By "experimental working hours" is meant the total time lapse between the official starting and stopping time for the day (standard working hours), from which the time decreases due to the experimental conditions of work listed above have been deducted.

period 7; that is, the day was one hour longer than in period 9. In period 11 the group went on a five-day week. Period 12 saw the group return to the conditions of period 3—no rest periods, no free lunches, and a full 6-day work week. Period 13 was essentially a repetition of periods 7 and 10 except that the operators furnished the morning lunch and the company supplied the beverages.

The variation in conditions produced extremely interesting results. The production of these girls climbed. Each test period had higher production records than the preceding one. Figure 10.2 and Table 10.1[2] present the general results.

However, to present the results of the second study in such an objective form as production records is to lose sight of the meaning and implications of the study. The important qualitative aspects will therefore be summarized.

Periods 1 through 3 successfully accomplished the transfer of the girls from the factory to the test-room situation. A group formation began to take place among these girls. They were allowed to talk more freely, and the supervisor's role, as they had known it, changed.

The second phase carried through four test periods and was mainly concerned with rest pauses. Although, on the basis of the data acquired, the experimenters believed that a rest period should occur somewhere between 9:30 and 10 A.M., they had no idea as to where to place the afternoon rest pause. In accordance with the general policy followed in the test room, the girls were consulted. The girls suggested that rest periods occur at 10 A.M. and 2 P.M. Some of their comments were as follows:

Operator 1A. Gee, that's the berries! It rests you to have five minutes like that. (To 3) Don't you like it?
Operator 3. Yes, I do.
Operator 4. I like the rest period, but I think one every hour would be better.
Operator 5. It rests you a little bit.

Since output increased even though the rest period reduced the workweek from 48 to 47:05 hours per week, the two rest periods were doubled, the workweek thereby being reduced to 46.10 hours. An immediate and definite rise occurred in the average output. At this point the girls expressed happiness over the increase in production and earnings but disagreed among themselves as to how to explain it. They also showed some apprehension as to whether they really would receive the increased earnings—one girl said, "We earn 80 percent but we'll only get 60 percent." The girls were offered the opportunity of getting their increases once a month but did not like the idea, for they believed that the company would find some way to avoid paying them. The experimenters state that these fears were irrational and unjustified; they sound as if they were proud that the girls were willing to express such fears. Right then they could have realized that the bonus system of payment was involved and that there was distrust. In this experiment the bonus idea was dropped. The implications are important and should be noted.

The girls were then put on six five-minute rest pauses, although they were unanimously in favor of two fifteen-minute periods. They did not favor this workday even though the workweek was reduced to 45:15 hours. Typical comments were:

[2] Figs. 10.1, 10.2, and 10.3 and Table 10.1 are reprinted by permission of the authors and the publishers from F. J. Roethlisberger and William J. Dickson. *Management and the worker —an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago.* Harvard University Press, Cambridge, Mass., 1939.

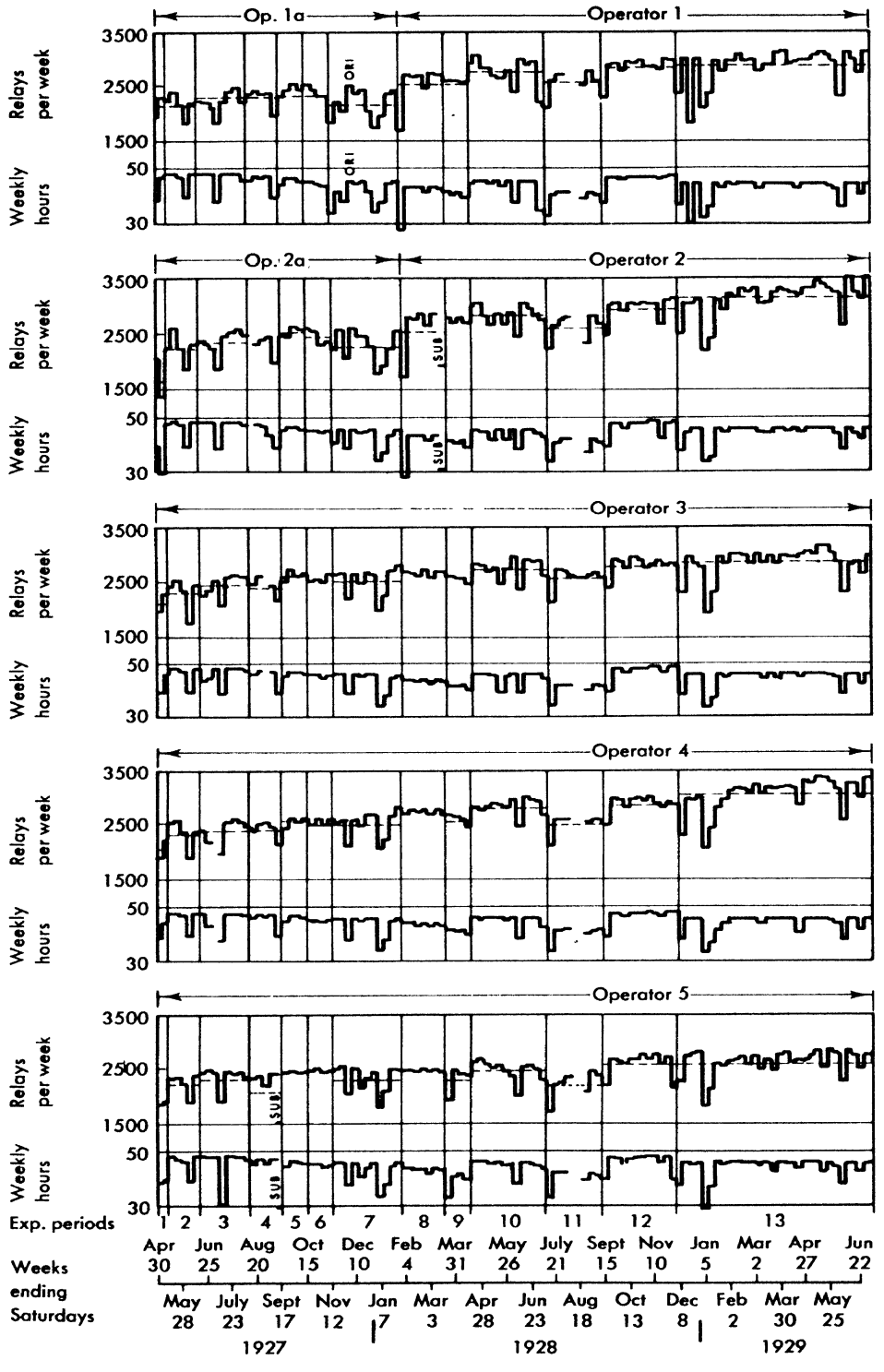


Figure 10.2. Total weekly output, experimental periods 1–13, relay assembly test room. (See p. 311, footnote 2.)

"I'm getting 'nuts' on this job." "I don't know what I'm doing." "I feel 'goofy' today." "I don't like these rest periods." "I just get started to work and then have to stop." In period 7 the workweek was increased to 45:40 hours and a midmorning lunch was served. The girls favored this period.

Analyzing individual production records during these first seven periods shows that operators 3 and 4 tend consistently to increase production. Operator 5 tends to maintain production consistently with only one spurt. Operators 1A and 2A are similar in production. Their trend is downward in periods 2 and 3, rises in 4 and 5, and drops in period 7. These two girls had been doing lots of free talking and were considered antagonistic and uncooperative. They were finally replaced in the experiment. This is another important clue that was overlooked by the experimenters.

The third phase of the Relay Assembly Test Room experiment was concerned with a shorter workweek (periods 8, 9, and 11). The three remaining periods (10, 12, 13) were checks. The operators were given the choice of starting work a half-hour later or ending a half-hour sooner. They chose the latter unanimously. This brought the workweek down to 43:10 hours. Output did not diminish and the girls liked the shorter schedule. Some comments were: "It's fine, and we still make our rate too." "I'll never get tired of it. Last night I had a whole streetcar all to myself." "I like it." In period 9 the workday was reduced another half-hour, the total workweek being 40:40 hours. Although the girls were in favor of a shorter working day, this was "too drastic." The total weekly output was lowered. The rate of output did not diminish, but the girls became quarrelsome. Operator 2, for example, scolded the other girls for not working hard enough. In period 10 they returned to the full 48-hour week with two rest periods of fifteen minutes in the morning with free lunch, and ten minutes in the afternoon. They did not like it; they claimed greater fatigue and expressed disinterest in their increased earning capacity. It is interesting to note that this "fatigue" was not mentioned in period 7, which was identical.

At this time the girls were given a questionnaire on health habits and psychological attitudes. They reported that they felt better in the test room but gave varied reasons. In answer to the question "What do you think has made it possible for you to increase your earnings since you have been on the test?" they replied: "Greater freedom," "absence of bosses," "more personal attention," "opportunity to set one's own pace." Further questioning brought out the fact that freedom from rigid and excessive supervision was the important factor in determining the girls' attitude toward work. In other words, rest pauses, a free lunch, a shorter workweek, and higher pay did not, in the minds of these girls, count for as much as freedom from such supervision.

Period 11 introduced the five-day week with the same rest pauses as in period 7. The girls were paid the basic hourly wage for the hours not worked on Saturday. No clear-cut drop in production was apparent; but since this period was disrupted by vacations, it is possible that it was introduced mainly as a "marking-time" period for all concerned. This is unfortunate in view of the popularity of the five-day week now, but in 1928 this was not so prevalent. Again, the experimenters missed an important clue.

Period 12 returned to the full 48-hour week without rests. The girls were informed that this period would last about three months and were asked to work in their normal fashion. Although the average hourly output dropped, total production reached the highest peak of any period. However, the girls took their own rest pauses during this period. They would eat candy and get thirsty, so they went to the drink-

ing fountain; in addition, there was much laughing and boisterous joking. The girls still liked the test room better because they could scream and have fun and because they had no bosses. Operator 3 at this stage told the test-room observer to "shut up."

In period 13 the rest periods were resumed. The girls furnished their own lunch, but the company furnished the tea. During this period it was observed that instead of antagonistic competition the girls worked toward a common goal. If one girl slowed up because she felt ill, the others would speed up. The total weekly output rose.

Since a careful record of "personal time" taken by employees was kept throughout the experiment, there is clear-cut evidence to show that organized rest periods decreased this personal time but also decreased the workweek. For example, in periods 3 and 12 (no rest periods) personal time averaged fourteen minutes. In periods 7, 10, and 13 with twenty-five minutes of rest periods it averaged seven minutes. In other words, the working day was shortened by eighteen minutes instead of twenty-five minutes. Physical examinations showed that the health of the employees was not impaired during this study. Absenteeism for these six operators while working in the factory averaged 15.2 irregularities but dropped to 3.5 irregularities in the test room. This is most probably to be attributed to a change in employee attitude.

The girls liked the workweek with no Saturdays best of all. (Next they preferred the 4:30 P.M. stop with rest and lunch.) The experimenters report that it is doubtful if the girls would have been unanimous in their approval had they not been paid for Saturday morning off. The present writers do not understand the doubt on the part of the experimenters!

During the course of this experiment the girls' attitude toward those in charge of it changed continually. In the early stages there was apprehension and fear. This finally disappeared as a result of the fact that the girls were consulted and informed about the various stages. By period 13 they liked the test room and its pleasanter, freer, and happier working conditions. In other words, they preferred the lack of restraint and excessive supervision. The test-room observer was regarded as a friendly representative of management and not as a supervisor who "bawls us out."

Another change that took place was the solidifying of the group. Four of the girls started going out with each other and all of them helped the others with their work. They no longer worked as individuals.

The experimenters were perplexed by the general trend toward increased production independent of rest pauses and shorter hours and by the improved mental attitude of the girls. Five tentative explanations were suggested to explain these facts:

1. Improved material conditions and methods of work
2. Shorter hours, which provided a relief from cumulative fatigue
3. Shorter work periods, which provided a relief from monotony
4. Wage incentive plan
5. Changes in method of supervision

The first explanation was dismissed. T. N. Whitehead (1933), who analyzed the entire body of data, finds no conclusive evidence to support it. There were no significant correlations between physical conditions and production. The second and third explanations were dismissed after careful examination of the daily work curves of

the operators. There was no evidence that these curves resembled fatigue curves, and only one operator's work curves showed a resemblance to the monotony curve.

In order to establish the truth or falsity of the wage incentive as an explanation, two minor studies were made. These were known as the Second Relay Assembly Test Room and the Mica Splitting Test Room. The original Relay Assembly Test group had a change of environment as well as of payment, whereas the Second Relay Assembly group had only a change of payment. Five experienced operators were put on a payment basis similar to that of the original group but were kept in the factory. This group increased production almost immediately by 12 percent and they continued at this level; but when they went back to the original method of payment there was a drop. These facts were interpreted by the experimenters to mean that (1) the formation of a small group to determine piecework earnings was an important factor in the Relay Assembly group and (2) subsequent rises in test-room production could not be explained in terms of this factor alone. A very important point is that the morale of the entire department was shattered. The other operators wanted similar consideration (and possibly the chance to earn more money). Because of the friction in the department, this experiment had to be discontinued.

In the Mica Splitting Test Room experiment the basis of wage payment remained at the same individual piece rate, but changes in working conditions similar to those in the test room were made. The primary purpose was to study the effects of the change in payment in the test room. Five operators took part in the experiment; two were selected at random and they chose the other three. Their production rose an average of 15 percent in 14 months. In the Second Relay group production rose 12 percent and the Relay Test Assembly Room rise in output averaged 30 percent. It can be assumed that 15 percent is the maximum that can be attributed to a change in financial incentive; the remainder must be attributed to other factors. The experimenters presented two conclusions based upon the evidence:

> 1. There was absolutely no evidence in favor of the hypothesis that the continuous increase in output in the Relay Assembly Test Room during the first two years could be attributed to the wage incentive factor alone.
> 2. The efficacy of a wage incentive was so dependent on its relation to other factors that it was impossible to consider it as a thing in itself having an independent effect on the individual. Only in connection with the interpersonal relations at work and the personal situation outside work, to mention two important variables, could its effect on output be determined.

By period 13 the experimenters realized that they had not studied the relation between output and fatigue, monotony, etc., but had performed a sociological and psychological experiment. By trying to control variables they had introduced a new one, a social situation that involved changed attitudes and interpersonal relations.

The chief result of years of work had been to demonstrate the importance of employee attitudes. For management, however, there were other practical results. They introduced rest pauses on a wide scale. They began to question many assumptions they had previously made, and they realized the errors of oversimplification. They saw that one could not predict the effect of a single factor if it were part of a total situation. They also realized the importance of gaining more knowledge about employee attitudes; this led to Study 3.

STUDY 3. MASS INTERVIEWING PROGRAM

Study 2 allowed the surmise that a relation existed between employee morale and supervision. It was believed that an improvement in supervision would improve morale. Since there was a dearth of facts on how to improve supervision and the material that was available was highly opinionated, it was decided to interview the employees in order to secure information. The Relay Assembly Test Room experiment showed that the employees were exceedingly disparaging about the supervision, whereas management had believed that the supervision, especially in this department, was good. This raised the possibility either that management knew little about what constitutes good supervisory methods or that they knew nothing of the employee attitudes on the subject. Thus, in this experiment the emphasis shifted from a study of changes in environmental work conditions to a study of human relations or of attitudes concerning human relations.

The Relay Assembly Test Room also showed that as the girl's attitudes improved toward each other, their work, the supervisor, and the company, their production increased. In other words, their morale affected their production. The company had been conducting supervisor training courses, and it decided that instruction in improving employee morale should be part of this course. Five meetings were planned, but two sessions made it clear that no factual data were available for this purpose. Consequently, as has been said, it was decided to interview the employees.

The interviewing program was launched in the inspection branch, in which about 1600 workers were employed. Three men and two women supervisors were chosen to conduct the interviews; they were told that the purpose was to gain information about employee attitudes, *not* to spy on supervisors or anyone else.

The first interview was conducted as follows:

1. Each interviewer was assigned a certain territory to cover. From the foreman of each department in his territory he was to obtain a list of the employees' names.
2. When the interviewer was ready to start interviewing in any department, it was recommended that he first go to the foreman in charge and make his presence known.
3. It was recommended that the interviewer select the man he wanted to interview because otherwise the supervisor might be tempted to give him all his "problem cases" first. However, the interviewer was to cooperate with the supervisor so that the operation of the department would be interfered with as little as possible.
4. The interviewer was to ask the supervisor's advice about where the employee should be interviewed—whether away from the job or on the job. (Subsequent experience showed that it was usually advisable to interview an employee away from his work. Thereafter it was recommended that the interviewer ask the department chief for a bench or desk where he could conduct the interviews without interruption.)
5. The interviewer was to make sure that the necessary arrangements were made for paying the employee his average earnings for the time consumed in the interview.
6. In his contacts with supervisors the interviewer was to be careful not to betray the confidence of any employee and to refrain absolutely from discussing the content of the interviews with the supervisors.
7. Only a few employees from any one location were to be interviewed on the

same day, so that the work of the department might go on normally and without undue confusion or curiosity.[3]

The following careful instructions with regard to approaching the employee and conducting the interview were also given the interviewers:

1. Whenever possible, the employee was to be formally introduced to the interviewer by the supervisor. Interviewers were not to interview employees whom they knew, because the acquaintance might influence the employees' comments.

2. When the interviewer and employee were seated and ready to proceed with the interview, the employee was to be told the interviewer's name again.

3. The interviewer was to explain to the employee the purpose of the interview, i.e., why any comments, either favorable or unfavorable, that the employee cared to make about his supervisors, working conditions, and job were being solicited.

4. The employee was to be told how the interviews would be used; for example, any complaints he had about working conditions would be investigated together with those of the other employees, and, as far as practicable, remedial action would be taken. The manner in which the material gathered from the interviews was to be used in supervisory training conferences was also to be explained.

5. The interviewer was to make clear to each employee that the interviews would be kept strictly confidential; i.e., the employee could tell the interviewer anything, no matter how bad it was, without getting in trouble himself or getting his supervisors or his co-workers in trouble. The interviewer was to explain that no names or company numbers would appear on any records and that the people who read the interviews or heard them read would not be told who the employee was or where he worked. Anything the employee said which might identify him with his supervisor would be deleted from his interview.

6. The employee was to be told that the company was as much interested in the things he liked as in those with which he was dissatisfied and which he thought needed to be corrected.

7. The interviewer was to take almost verbatim notes as the employee talked. He was to explain to the employee that he was writing down what was said word for word so that there would be no possible chance of misrepresenting or forgetting anything. (At first it was thought that taking notes might make the employee reluctant to talk, but this was found not to be true.)

8. The interviewer was to be sympathetic and a good listener and to let the employee know that he was really interested in his problems and complaints.

9. Strict care was to be taken to express no agreement or disagreement with the complaints the employee made. The interviewer was to let the employee know that he himself was in no position to judge the correctness or incorrectness of what the employee was saying.

10. The interviewer was not to inform the employee of the nature of the complaints made by other employees.

11. The interviewer was not to give the employee advice as to what he should do. In rare cases he might advise an employee to see his supervisors, or tell him about the various benefit plans, the Hawthorne evening school, or similar things. However, the interviewer was not to hesitate to offer encouragement to any employee if he thought it would do him good.

12. The interviewer was to write up the interview under six headings. The opinions of the employee were to be divided first into three categories: working conditions,

[3] *Ibid.*

supervision, and job. Each of these headings was to have two subclassifications, likes and dislikes.[4]

The reaction of the employees to this program was generally very favorable. A typical reaction was, "It is a good idea to interview the operators, as they may have something in their minds that they want to talk over with someone and this gives them a chance to do it." Another type of reaction was, "I never really thought this interviewing amounted to very much, but since you explained it to me I see where I can tell you something that may help in the supervisors' conferences." Other reactions were illustrated by still another type of response: "Now maybe some of the people will take a hint and learn that they have not been doing the right thing, especially some of these underhanded bosses. I hope they get their share of this."

Some employees attributed changes in conditions to the interviews when in reality the only change that occurred was in the attitude of the person interviewed. For example, one woman believed that the food in the cafeteria was better Some employees, on the other hand, believed that management would make no changes at all; a few were very suspicious.

Naturally, one should expect all types of reactions to an interviewing program, and this was no exception. The success or failure and the final total reaction depend entirely upon whether the employees understand and endorse the real purpose and whether the company's actions conform to it. Because in this instance they did, it was possible to obtain a tremendous amount of information that could not have been secured in any other way. Not only was employee reaction favorable, but the reaction of the supervisors was also. One quotation illustrates this: "Did you go to the last conference we had? Say, they are getting to be good. You know, I am getting a lot of help from them. I am learning to see the operator's viewpoint of things, and really believe I am learning to do my job much better by attending these conferences."

Another outstanding aspect of this part of the Hawthorne Studies was its flexibility. The program was expanded and the method of conducting interviews underwent drastic changes as the facts became known. Although originally the interviews were not meant to be of the questionnaire variety, there nevertheless was a typical way of beginning each interview and during its course specific points were covered by direct questions.

At conferences it was reported that employees often discussed irrelevant topics, that is, irrelevant from the point of view of the subject matter the interviewers were to cover—supervision, working conditions, and the job. Since this happened fairly frequently, it was decided that these discussions were not irrelevant at all. Analysis showed that direct questions put the employees in a "yes or no" frame of mind, often impeded the progress of the interviewer, and inhibited the spontaneous expression of their real convictions.

As a result, the interview procedure was changed to the indirect method. The employee had considerable freedom in choosing his own topic at the start of the interview and the interviewer allowed him to continue to talk. This change in technique resulted in an increase in the average length of the interview from a half-hour to one and one-half hours. The material on which the report was based also increased from two and a half to ten pages. When the program was suspended because of the depression, 21,126 employees had been interviewed.

The ramifications of the various phases of the interviewing program were many. The interviews had a real effect on the company's supervisor training courses. It

[4] *Ibid.*

provided management with a list of complaints about specific environmental conditions that could be investigated. Data concerning employee attitudes and opinions were made available on a large scale. It was also found that the employees benefited psychologically merely as a result of being interviewed.

However, certain perplexing problems arose as a result of the program. For example, employees reacted differently to the same surroundings. Thus some expressed satisfaction with the same thing that others considered unsatisfactory. Comments during the interview were expressed as either fact or sentiment, but often no distinction was made in the mind of the person being interviewed. In other words, the employees were reacting to a personal situation on the basis of their previous social conditioning. Further, they were responding as part of the social organization of the group in which they worked and in relation to their position in this group. The program showed that such items as wages, hours of work, and physical conditions could not be considered as factors in themselves. Instead they should be considered as carriers of social values. They could be understood only through the acquisition of information about the individual's position or status in the group in which he worked and in the company as a whole. Thus the meaning assigned to wages, hours of work, locker rooms, etc., varied according to the employee's position in the group, the group attitude toward the specific item, and his relations with people outside the job.

From this it followed that information should be obtained not only about the individual's attitudes and opinions but also about the social groups that existed. Evidence acquired in this phase of the study indicated that these groups, about which management knew nothing, could exercise considerable control over the work behavior of the individual members. Restricted output, which often occurs in industry, was found to exist in the plant. Hints indicated that it was probably the result of the formation of social groups and the resulting pressures. There was also some evidence of the development of informal personal leadership in these groups. The existence of this type of leadership was not recognized by management, even though it was likely to be as important as any other factor—in some cases more so—in the group's acceptance or rejection of management's regulations. It was in this setting that the fourth study was begun.

STUDY 4. BANK WIRING OBSERVATION ROOM

This part of the Hawthorne Studies attempted to obtain more exact information about social groups within the company. The preceding study had progressed from the proposed guided interview to a more intensive type of unguided interview, and then to a series of interviews with one person. In it, the emphasis was on obtaining information from large numbers of employees. The last phase of the program pointed to the need to go back to a study of the Relay Assembly Test Room type, in which information of an intensive nature would yield data on the social groups in existence. The reports of two of the interviewers will serve as a good introduction to the fourth of the Hawthorne Studies.

> They [the employees] firmly believe that they will not be satisfactorily remunerated for any additional work they produce over the bogey, or that if they do receive more money it could only be for a brief period, at the end of which the job would be rerated.

Because of the belief that rates many ultimately be lowered if output is too great, there seems to be a tacit agreement among the members of this group to limit their production to the bogey requirements on each operation. Seldom do they exceed the bogey by a large margin. Most of the work is turned out in the morning in order that they can "take it easy" during the latter part of the afternoon. When questioned as to whether or not their earnings would be greater if they turned out more work, they claimed that the difference, if any, would be negligible because the percentages made by the other groups tend to pull theirs downward. To this general scheme all their attitudes and behavior are related.

The leader in this group is one of the two group chiefs, undoubtedly a very significant factor in giving the group a strong feeling of security. This supervisor, A, was at one time on the bench in the same group which he now supervises, but he refused to allow the change to alter his relations with the men. From observing the group one can hardly draw a line between supervisor and employees. It is obvious that he is very popular with them; no one has any adverse criticism to make of him. He is very close to the men, keeping them well informed at all times as to the group standing in the department, i.e., relative percentages, rates, output, etc. When asked why they consider him a good supervisor, his men replied with such statements as: "He knows his stuff," "He's fair and impartial," "He'll go to hell for you and make sure you get plenty of work." In short, all their statements implied a firm conviction that this group chief would protect their interests. By way of contrast, while A was on sick leave, another supervisor, B, took over the group. Toward B the employees expressed strong antagonism. B is an older man, further removed from the interests and sentiments of his subordinates. He is not quite trusted by the men and commands very little respect. As one employee sized him up, "When he bawls you out, he is more nervous than you are." This group is only vaguely conscious of the other supervisors in the department, in fact, a confusion of the supervisory ranks is quite evident For example, C, a section chief, has held the same position for a number of years; but the men cannot figure out what position he holds in the department, who reports to him, or what his duties are.

An attitude common to this group, but existing in varying degrees of intensity, may be characterized as a lack of ambition and initiative and a complacent desire to let well enough alone. Most indifferent is their attitude toward advancement, referring, of course, to promotion or higher-grade work. Whereas it is usual in any group to find several employees striving to improve their position, here only one or two seem to be interested. The others merely say, "All we are here for is the old pay check." Sometimes they speak of the department as the "Old People's Home" because, quoting one man "The fellows get in here and don't seem to want to get away Take a fellow like me I have been on this job ten years. If I was transferred out, I would have to start all over again and I would have a pretty tough time."

In their group life they are congenial and happy-go-lucky. This is obvious not only during rest periods but also while they work. Side play is frequent, and good-natured bantering constantly flashes back and forth. During rest periods everyone either plays cards or stands by as an interested spectator, and in these games rivalry is always keen but congenial. Several of the newer men express the consensus of opinion by describing their associates as "a swell bunch of guys."[5]

Another investigator reporting on a different group tells a different story.

A says B gets mad because he (A) does too much work: "B sometimes tries to do as much as I do, and whenever he can't make it he gets mad and swears about it. Then he will go over to some of the others and say that I am trying to kill the bogey." The girl assemblers in the group tell A that he should not stand for the treatment he gets from the group chief. They tell him he does the most work and gets the least recognition.

[5] *Ibid.*

A mistrusts D because D represented himself as a sort of a supervisor to A and took the easy jobs when A first came to work here. He is not friendly with E because E does favors for everyone but him. His friends are an old man, G, and the girls. When B was called to the office because his production was too low, A told him it was his own fault. B then said, "What! Do you expect us to come down here and slave?"

B is 36 years of age, a rather stocky, well-built, athletic type. Although he supports his father and mother, their dependence apparently serves to accentuate his own manhood. He says that the supervisors are all satisfactory. He knows them well because he has worked here so long. He takes a senior position in the group and gets along well with everyone but A. His attitude toward A is indicated by the incessant "kidding" to which he subjects him. He attributes to himself all the best characteristics of virile manhood and attributes to A feminine characteristics. He says that A is an hermaphrodite. He demonstrated in the first interview how A swings his hips and carries himself like a woman. B thinks that A works hard because he is "dumb," and that nobody likes A because he does so much. He explains that A sits all by himself (in reality he sits next to B), and nobody will talk to him, so all he can do is work from the first whistle to the last. B was once offered a position as supervisor which he refused.[6]

The need for a more systematic inquiry resulted in the selection of 14 male operators who were to work under standard shop conditions. These workers were observed and interviewed over a period of six and a half months; the study was terminated when work ceased because of the depression. This group of men were reluctantly assigned to a separate room. By this time the researchers knew that such a change is often of importance; however, it made possible better control of the study. The observer was stationed in the room; he was to assume the role of a disinterested spectator but was not to set himself off from the group. He adhered strictly to the following rules: (1) Give no orders and in no way demonstrate authority. (2) Do not take part in arguments. (3) Do not enter into conversation or seem overanxious to hear about what is going on. (4) Never violate confidences of employees.

The observer was asked to note the formal organization of supervisor and employees, and also all informal groupings of the men. Further, he was to observe the interrelations of these two types of organizations. The interviewer did not enter the test room. His function was to gain insight into the workers' attitudes, thoughts, and feelings, whereas the observer was to describe the actual verbal and overt behavior of the group. Working thus together, these two were to gather data from this group concerning the department, the company, and the community.

The workers in the Bank Wiring Observation Room study consisted of three groups: nine wiremen, three soldermen, and two inspectors. Each did a specific task but necessarily collaborated with the others. This department was chosen because it met such criteria as (1) the sameness of the task; (2) exactly determinable output; (3) shortness of task (one minute required); (4) work pace determined by operator; (5) assurance of continued work; (6) the ease of removing the group as a unit from the department; (7) the experience of the operators. These criteria were similar to those used in the Relay Assembly study, but from this point on there was a difference.

The men were invited to cooperate in the study. The first week they worked or appeared to be working all the time. They were cautious toward the observer. When they complained to him about poor lighting, he told them that he had no authority

[6] *Ibid.*

and suggested that they refer all complaints to their supervisor. It was three weeks before the men started to relax and behave more as they did in their regular department. It was learned that these men did not think that either the group chief or the section chief had much authority. The foreman spent little time in the room so they were relatively free from authority.

The system of payment was a complicated wage incentive plan that had been instituted to promote efficiency by encouraging production; it was also believed to be a fair means of apportioning employee income. It was soon found that this wage plan was not working. The workers defined a day's work as the complete wiring of two units and either they stopped before quitting time or they paced themselves to last out the day. No uniform explanation or reason was forthcoming for this definition of a day's work by the men, but it completely invalidated the incentive plan, as the following conversations prove:

WORKER 2: (After claiming that he turned out more work than anyone else in the group.) They [his co-workers] don't like to have me turn in so much, but I turn it in anyway.

(In another interview.) Right now I'm turning out over 7000 a day, around 7040. The rest of the fellows kick because I do that. They want me to come down. They want me to come down to around 6600, but I don't see why I should. If I did, the supervisors would come in and ask me what causes me to drop like that. I've been turning out about that much for the last six months now and I see no reason why I should turn out less. There's no reason why I should turn out more either.

WORKER 3: No one can turn out the bogey consistently. Well, occasionally some of them do. Now since the layoff started there's been a few fellows down there who have been turning out around 7300 a day. They've been working like hell. I think it is foolishness to do it because I don't think it will do them any good, and it is likely to do the rest of us a lot of harm.

INTERVIEWER: Just how do you figure that?

WORKER 3: Well, you see if they start turning out around 7300 a day over a period of weeks and if three of them do it, then they can lay one of the men off, because three men working at that speed can do as much as four men working at the present rate.

INTERVIEWER: And you think that is likely to happen?

WORKER 3: Yes, I think it would. At present we are only scheduled for 40 sets ahead. In normal times we were scheduled for over 100. If they find that fewer men can do the work, they're going to lay off more of us. When things pick up they will expect us to do as much as we are now. That means they will raise the bogey on us. You see how it works?

INTERVIEWER: You say there is no incentive to turn out more work. If all of you did more work, wouldn't you make more money?

WORKER 4: No, we wouldn't. They told us that down there one time. You know, the supervisors came around and told us that very thing, that if we would turn out more work we would make more money, but we can't see it that way. Probably what would happen is that our bogey would be raised, and then we would just be turning out more work for the same money. I can't see that.

WORKER 5: There's another thing, you know the fellows give the fast workers the raspberry all the time. Work hard, try to do your best, and they don't appreciate it at all. They don't seem to figure that they are gaining any by it. It's not only the wiremen, the soldermen don't like it either. . . . The fellows who loaf along are liked better than anybody else. Some of them take pride in turning out as little work as they can and making the boss think they're turning out a whole lot. They think it's smart. I think a lot of them have the idea that if you work fast the rate will be cut. That would mean that they would have to work faster for the same money. I've never seen our rate cut yet, so I don't know whether it would happen or not. I have heard it has happened in some cases though.

WORKER 6: (Talking about a relative of his who worked in the plant.) She gets in here early and goes ahead and makes up a lot of parts so that when the rest of the girls start in she's already got a whole lot stacked up. In that way she turns out a great deal of work. She's money greedy. That's what's the matter with her and they shouldn't allow that. All she does is spoil the rate for the rest of the girls.

INTERVIEWER: How does she do that?

WORKER 6: By turning out so much. When they see her making so much money, they cut the rate.

WORKER 7: There's one little guy down there that turns out over 7000 a day. I think there's a couple of them. And we have to put up with it.[7]

The men devised various means of controlling production. Name calling and minor physical punishment were two of the more common ways of restricting output. Workers who produced too much were nicknamed "Slave," "Speed King," or "Phar Lap" (a champion race horse of that year). They were also "binged." A "bing" is a very hard blow on the muscles of the upper arm. The one who is hit never protests but is allowed to "bing" back.

The men's concept of their average daily production was reflected in rather constant weekly production figures. The men achieved this constancy by reporting more day-work allowances than they were entitled to. In addition they sometimes reported more—or in some cases less—production than they had actually turned out. The primary reason for this was to gain group acceptance.

Three men always reported more work than they actually produced and two reported less; the others varied their reports. A comparison of morning and afternoon production showed that the faster men slowed down in the afternoon, whereas the slower men worked at a more even pace. Briefly, the findings were that the men were restricting production in accordance with their definition of a working day, thus nullifying the validity of the wage incentive plan. Interpersonal relations apparently were more important than the wage incentive.

The group chief had certain difficulties In the first place he had to handle the day-work claims of the men. These claims were made to justify being paid at an hourly rate rather than on a production basis. Company rules allowed such claims, and they could be made for any number of reasons. The group chief had either to accept these claims as justifiable or be arbitrary in rejecting them. He chose to accept them and thereby gained the good will of his men. It would have been difficult to prove any of the reasons given as being incorrect or unwarranted.

Another problem was job trading. The only excuse for this was physical incapacity, as when a solderman developed a sore finger. Determining how sore a "sore" finger has to be made this a difficult claim to dispute. Thus the group chief was sympathetic to his men and steered a middle course, and he, in turn, was popular with them. During this study he was demoted because of business conditions and a group chief with greater seniority took over. The new one placed great stress on conduct and efficiency. The men thought that he was exercising more authority than was vested in him, and they did not admit his authority merely because he exercised it. Certainly the first group chief with his leniency received more cooperation than the second one.

The next representative of management was the section chief, and since he supervised a number of groups he was never in close contact with any one group at all times. His function was more managerial, and he was considered to be more "in the

[7] *Ibid.*

know" than the group chief. Even though the men argued freely with him, they regarded him as having more authority than the group chief. The assistant foreman, next in the management hierarchy, was listened to with respect, but the men never argued with him. If they disliked what he said, they waited until he left to voice their opinions. The relations of the assistant foreman to the group were pleasant. The foreman was called the "old man." When he came in, conversation stopped and no one knowingly broke any of the rules. The men showed apprehension while he was present.

Considering the management-employee situation, it was apparent that although communications traveled down in the form of orders, the two first-line supervisors were likely to be questioned. But there was a gap in communications on the way up from employee to foreman. Consequently the foreman and top management were unaware of the reasons for the failure of the financial incentive. The fact that the men reacted differently from the way it had been assumed they would made the incentive plan ineffectual and was something the supervisory organization could not remedy.

The relations among the employees were especially interesting. The men worked according to "their standard" of production, but in addition they talked, argued, played games, matched coins or indulged in other forms of gambling, formed cliques, took sides, traded jobs, shared candy, insulted one another by belittling nationality and religion, and helped one another in their work. They nicknamed each other "Runt," "Shrimp," "Jumbo," and "Goofy." Their conversation ranged from work to women to horse racing. In short, they did many things together, in addition to working.

The connector wiremen, even though their rates might be the same as the selector wiremen were paid, represented the "elite." Going on connectors was a step forward, whereas being put on selectors was regarded as a demotion. The wiremen occupied a social position above the soldermen. Job trading between them originated most often with a request from the wiremen. The soldermen wore goggles which they resented, and the wiremen demonstrated their superiority by expressing disapproval when the soldermen did not wear them. Lowest was the truckman, who transported materials. He was the butt of much horseplay.

The inspectors belonged to a different group. They were responsible to a different set of supervisors. They were not an integral part of the group and were considered outsiders.

A subtle manifestation of status appeared in the way the men dressed. The foreman and assistant foreman wore coats and vests. The section group chiefs wore vests but no coats. The men wore neither coats nor vests. When the men reported for interviews they did not put on their coats, but the inspectors put on both coats and vests.

During lulls in activity the men played games. It was interesting to note that two groups always formed. Group I consisted of four wiremen, a solderman, and an inspector. This group usually gambled. Group II, not as completely set, consisted primarily of one solderman and three or four wiremen. They preferred "binging." A solderman, a wireman, and the other inspector were isolates, that is, not in either group. These groups or cliques carried over from games to job trading, quarrels over opening and closing windows, and friendships and antagonisms. Furthermore, Group I regarded itself as the superior or "front-room" clique. They felt that their talks were on a higher plane, they ate chocolates rather than "junk," and they were less boisterous.

A diagrammatic summary of the internal organization is shown in Figure 10.3.

"There were four main determinants of clique membership: (1) you should not turn out too much work (rate busting); (2) you should not turn out too little work (chiseler); (3) you should not tell a supervisor anything that would harm an associate (squealer); (4) you should not act officiously (this applied to inspectors and group chiefs as well as workers).

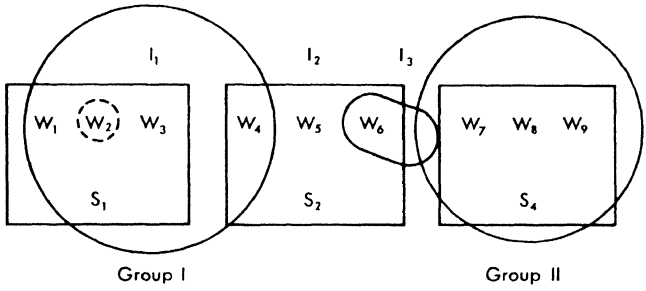


Figure 10.3. Internal organization, bank wiring observation room. (See p 311, footnote 2)

This intricate social organization served to protect the group both inside and outside. Control inside was obtained through ridicule, sarcasm, and "binging." Protection outside was afforded by excessive day-work claims and constancy of production. It has already been noted that management knew nothing about the group and its attitudes toward production and management rules until this phase of the study uncovered it. All companies, large and small, have such a setup and under usual conditions they have no way of knowing about it.

STUDY 5. PERSONNEL COUNSELING

The Personnel Counseling study was not begun until four years later, because of the depression. However, these four years allowed for an appraisal of the vast significance of this series of unique studies in industry. The Experiments on Illumination showed that there was nothing resembling a close relationship between a change in physical environment and production. The Relay Assembly Test Room with its continuous rise in production, regardless of the changes, showed the importance of employee attitudes toward job, supervisor, and home. The Mass Interviewing Program not only brought about improved methods of interviewing but also uncovered much data on specific attitudes, as when it showed that similar situations and conditions could be sources of satisfaction to some and dissatisfaction to others. The suggestion of a group formed by the workers led to the Bank Wiring Observation Room study, with its findings on the intricacies of informal organization and the ways in which this affects production and supervisory relations. It also showed the breakdown in the flow of communication from employee to management; management was free to give orders but completely in the dark as to how they were finally received.

Since all four studies indicate both a lack of accomplishment in really promoting employee relations and the importance of this work, the fifth study may be regarded

as the culmination of the experimenters' efforts to bring this important area into focus. There were two objectives. The first was to have a nonauthoritative and impartial agency interview employees to diagnose their problems and work with supervisors on their methods of supervision. The second was to improve the method of communication within the company, in view of the ample evidence indicating that it was lacking in certain situations when the social organization conflicted with the managerial organization. The plan was put into operation with the announcement that "personnel men" would be assigned to the departments and would be around to talk to the employees. The program, which was generally accepted, led to improvement in three fields: personal adjustments, supervisor-employee relations, and employee-management relations. In several cases there were obvious improvements in personal adjustments. Employees showed changes in personality, and freedom from anxiety and other forms of behavior that might be diagnosed as psychoneurotic. The personnel counselors were helpful in making the supervisors see their problems with less emotion and more understanding. Finally, the reports helped management to a better understanding of employee behavior and sentiment so that policies could be formulated that would result in less friction between management and workers.

IMPLICATIONS OF THE HAWTHORNE STUDIES

The Hawthorne Studies are significant because they represent an honest and concerted effort to understand employees, instead of approaching the problem only from the managerial point of view of increasing "efficiency" on an economic level. Moreover, they are interesting because they are a type of experiment which is rarely performed in industry. They do not try to define a response as the result of the introduction of change; they recognize that it is a result of the specific change *plus* the employee's attitudes, his social situation on the job, and his previous attitudes as determined by his personal history and background.

The Hawthorne Studies have been discussed fully because they are important to industrial psychology. They are not to be considered as the "ideal" experiment or the acme of perfection. There are many reasons for being critical and even for calling the researchers naive. For example, it may be that the end product, personnel counseling, is merely a substitute for a real need of employees to organize. It may be that the vast difference between the Bank Wiring Observation Room and the Relay Assembly Test Room results was caused by the fact that one group was male and the other female. Evidence from other sources indicates that women regard supervisors as more important than do men, and they are also more tolerant toward work conditions, possibly because they often look upon their job as a temporary stopgap until marriage or child-raising. Further criticisms are avoided lest the wrong impression be created. The Hawthorne Studies can be considered as the coming of age of industrial psychology just as Münsterberg's work might be considered its birth.

There are many reasons why this series of experiments deserves careful study. Although the studies are old when compared with contemporary literature, they are nevertheless as timely now as when they were first conducted. Many of the findings

tend to be rediscovered and in many instances the newer data are not so adequate —for example, on the importance of employee attitude and the existence of informal employee groups and their controls on production. Similarly, the relation of hours of the workweek to production and the fact that a shortening of the workweek does not necessarily curtail total production rings a modern bell. Also, one must mention that increasing the financial incentive does not always increase production. These are only a few of the findings that are still not believed in many quarters.

However, above and beyond the generalizations based upon the findings are the tremendous implications for research. Too often under the guise of research are the thinly veiled attempts to gather data that conform to existing knowledge and to prove the point that the researcher or investigator sets out to prove. Predetermined conclusions discourage the real search for knowledge. The Hawthorne Studies sought conclusions and the experimentors were not afraid to go in the directions indicated by the data. Their research was not "directed."

Important indicators of the value of any research undertaking are the questions raised as a result of the knowledge acquired. Meaningful questions must often await the answering of questions on a more simple level. Many college professors recognize the value of their presentation in relation to the questions asked as a result of what has been learned.

The significance of the Hawthorne Studies can be understood not only in relation to their findings but also because they are an outstanding example of research that was not steered to predetermined conclusions and because they raised questions that otherwise might not have been asked.

REFERENCES

Mayo, E. *The human problems of an industrial civilization.* Macmillan, New York, 1933.

National Research Council The Western Electric researches. In *Fatigue of workers.* Reinhold, New York, 1941.

Roethlisberger, F. J. *Management and morale.* Harvard University Press, Cambridge, Mass., 1943.

Roethlisberger, F J., and W. J Dickinson. *Management and the worker.* Harvard University Press, Cambridge, Mass., 1943.

Snow, C. E. A discussion of the relation of illuminating intensity to production efficiency. *Technical Engineering News,* November 1927.

Whitehead, T. N. *The industrial worker* Harvard University Press, Cambridge, Mass., 1933.

11 MOTIVATION AND WORK

Work is a highly complex phenomenon in our present stage of technological development. To understand work we need to recognize this, rather than assume an oversimplified push-button concept. Work may be a task, a duty, or an accomplishment. It may be mental, physical, or both. It may be repetitive or creative. Further, it may be drudgery or personally rewarding, and its results may be obvious or subtle. Its end product is often evaluated differently by different people. In other words, work takes on different shades of meaning, and most important is the intrinsic meaning that it has for the individual performer and the group with whom he identifies.

Regardless of its meaning, work cannot be considered apart from the individual who performs it. His motives, experiences, and social interrelations with his family, company, and community must always be considered.

To strip work of these attributes is to reduce it to the mechanics of an automatic machine, and this cannot be done if we are to understand its meaning. Work has an economic aspect and a mechanical aspect, but it also has a psychological aspect. Attempts to overlook this last characteristic result in an oversimplification leading to misunderstanding.

The view that specific incentives will encourage increases in production has not always been substantiated in fact, even though management has often attempted to spur production by such offerings and has often attributed production increase to them. Throughout the years production has increased for many reasons in addition to the particular incentives offered—and sometimes in spite of them.

Industry has generally overlooked the complexity of human motivation and has erroneously oversimplified a highly complex phenomenon. Since the psychologist is especially concerned with understanding an individual through his motives and has acquired a body of knowledge in this field that often differs from the layman's knowledge, it is necessary to review briefly, from the psychologist's point of view, what is known about motivation at the present time.

THE COMPLEXITY OF MOTIVATION

The one thing that psychologists know is that an individual rarely if ever behaves or responds in a situation as a result of a single motive. They clearly recognize the complexity of human behavior and understand that a person often does not know the true reason for his behavior. Because of this complex behavior, an individual is, in many instances, unpredictable.

A major error in industry has been the oversimplification of the concept of motivation. Too often since Taylor's time (see page 575) it has been assumed that the primary reason that people work is to make money. This is both absurd and unquestionably false. Man in industry is just as complex as man in any other phase of life, and any attempt to reduce his behavior to a single system of motivation must result in artificiality and narrowness. Man has many motives, and unless we recognize the part played by each one we cannot possibly begin to understand his behavior.

When the psychologist talks about motivation, he is concerned with studying the individual with respect to ever-changing physiological conditions and a multitude of previous experiences. Because of these two factors, physiological changes and previous experiences, it must be recognized that motivation may stem from within an individual or from factors acting on him from the outside. Furthermore, these two categories are not mutually independent; on the contrary, they interact at all times.

The simpler motives originating from physiological imbalance within an organism are generally known as hunger, thirst, oxygen need, the need for rest and sleep, curiosity, and sex. A good many motives act on the individual as a result of the multitude of his experiences; these are in large part determined by the social pressures of the society in which he lives. Although these social motives are sometimes given only vague and indirect expression, they nevertheless must be recognized as just as important as the so-called simpler motives if we are to be successful in understanding why man behaves the way he does. Gregariousness, self-assertiveness, acquisitiveness, the desire for prestige, and the desire to conform are only a few of the many social motives that act on man.

In the integration of the behavior of any normal individual at any one moment many motives are present. The particular behavior that results will depend in part upon the physiological imbalance of the individual and also upon the pressures of society. For example, a man may have a craving for steak, but if his religion has told him not to eat meat on Friday, he may not eat it. Thirst can be satisfied most easily by water; but the influence of advertising results in our drinking the many "colas," or even alcoholic beverages that are likely to affect our emotional behavior and that actually have little to do with satisfying the original thirst drive. While it is true that a certain amount of the sex drive is physical, the prohibitions, inhibitions, and repressions are determined so completely by the pressures of society as to lead some people to recommend that young people go in for athletics and take long walks rather than think about this drive.

Motivation is truly complex. Different forms of behavior are sometimes similarly motivated. A desire to be outstanding or to acquire prestige may cause one person to write a book; another may decide to achieve this by dressing effectively, another by marrying well, and still another by remaining single. The next person may decide to fulfill this desire by acquiring wealth. We can go down the list of all the different ways of behaving to achieve this one end. All these different forms of behavior may lead to the same amount of success by achieving the same end. The converse is also true. Different motives may sometimes result in one form of behavior. Thus a person may write a book because he wants to acquire wealth or gain prestige or have an opportunity for self-expression or creativeness.

To complicate matters further, people often do things without being aware of the basic motive or motives involved. The person who is fired may honestly believe that the foreman was biased, instead of realizing that the reason for dismissal is inefficiency. Giving "good" rather than "real" reasons for behavior is known as

rationalization, a dynamism of which there are many examples. Sublimation, projection, identification, and compensation are a few of the forms of behavior in which the real motive is unknown to the individual. The neurotic does not generally understand why he really is neurotic, although he may know that he is neurotic. A psychotherapist is needed to help him understand the basic motives compelling the particular form of aberrant behavior.

The vocabulary of motivation is large. Such terms as motive, purpose, desire, goal, preference, perception, attitude, and incentive all have their place. Regardless of the individual word applied at any given moment, a person is likely to have a number of motives, drives, incentives, desires, wishes, and purposes operating at the same time. Whether they are consciously present or are unconscious, many of them act upon the individual simultaneously. It is therefore safe to expect a normal person to be in conflict some of the time. A child trying to decide between candy and ice cream is in conflict. An employee who wants to tell the boss off but also wants to keep his job is in conflict. An employer who must decide whether to sell an oversupply of a commodity at a loss or hold it a little longer is likewise in conflict. Furthermore, a decision once made does not necessarily resolve the conflict.

Drives or motives vary in strength not only from one individual to the next but within the same individual at different times. Sometimes the immediate goal is at cross-purposes with the distant goal, even though both are desirable. Of course if we could know all there is to be known about the various motives that operate both within the individual and from the outside and, further, if we could know the relative strength of these drives at the particular moment, we could predict behavior accurately. But since present-day knowledge in psychology has not reached this stage, predicting behavior is extremely difficult. The factory worker may work overtime or not, depending upon whether he believes he needs a rest, has a date, wants to "get in good" with the supervisor, needs the money for food or clothing, or works just for the sake of earning money. Intimately tied up with motivation are man's various emotions; fear, anger, joy, or just no feeling at all will also enter into his decision on how to answer the boss's simple question: "Will you work overtime tonight?"

A man on a job never works in the vacuum of his job. Therefore, man's conflict in industry is just as real and just as complex as a conflict that a psychologist or psychiatrist studies in the office. The decision to strike, to quit the job, to ask for a salary raise, to talk back to the boss, or to argue with a coworker rarely if ever results from a single motive, regardless of the fact that the employee, when requested to explain his actions, may give a single reason. Known or unknown to him at that time are the consequences of satisfying the various physiological needs and the social drives operating in the situation and in him.

Young people can be expected to act differently from old people; if they act in the same way, it is usually because of different motives which occupy different relative positions. Married workers will act differently from unmarried workers. Men will act differently from women on a job because they have different motives, or motives of different relative strengths.

The production manager of a large factory employing young women once complained of the high turnover rate, despite the fact that job security was offered and the company had the proud record of never having dismissed a single employee for lack of work. In addition, it paid fair competitive wages and offered many inducements such as bonuses, music, good lunches for 35 cents, and other things. The

company was considering the possibility of introducing psychological tests to reduce the labor turnover. In this case it does not appear that such tests would decrease the turnover to any appreciable extent, although they might be useful in shortening the training period and thus be moderately helpful. Labor turnover here is a reflection not of the girls' ability or the company's practices; it is a function of the fact that this firm employs young women who have many reasons for leaving a job, including the possibility that they may have a greater desire to marry than to continue working. Perhaps hiring men or older women would reduce the turnover. This illustration is not to be interpreted as a bias for or against hiring young women. Since this factory is very successful despite its high turnover, it might be that a change in its personnel policies would create other problems such as modification of the wage structure or changes in the method of supervision.

MASLOW'S THEORY OF HUMAN MOTIVATION

One way of defining motivation is to talk in terms of a particular state of the individual—a state of *imbalance* or *disequilibrium*—as defining a motivated person. The person is motivated to correct any such imbalance, that is, he seeks some way of obtaining relief. Thus one might talk of a *motivational cycle* as a process having three distinct parts.

1. Need or motive: a state of physiological or psychological imbalance
2. Response of motivated behavior: an action directed toward alleviating the imbalance
3. Goal: that which has to be obtained to reduce the imbalance—the object of the motivated behavior

Psychologists interested in motivation have spent considerable time discussing different types of needs. Maslow (1943) has proposed an interesting theory concerning human needs and their effect upon human behavior. He suggests that human needs may be classified into five different groups or classes:

1. *Physiological needs*: These are the basic needs of the organism such as food, water, oxygen, and sleep. They also include the somewhat less basic needs such as sex or activity.
2. *Safety needs*: Here Maslow is talking about the need of a person for a generally ordered existence in a stable environment which is relatively free of threats to the safety of the person's existence.
3. *Love needs*: These are the need for affectionate relations with other individuals and the need for one to have a recognized place as a group member—the need to be accepted by one's peers.
4. *Esteem needs*: The need of a stable, firmly based self-evaluation. The need for self-respect, self-esteem, and for the esteem of others.
5. *Self-actualization needs*: The need for self-fulfillment. The need to achieve one's full capacity for doing.

The important thing about Maslow's Theory, however, is not so much his classification system as it is in the fact that he considers these five need classes to form a "hierarchy of needs" structure. That is, it proceeds from very basic needs (groups 1 and 2) to a cluster of higher, social needs (groups 3, 4, and 5).

The hierarchy concept is critical to Maslow, since his basic premises are that:

1. The behavior of any person is dominated and determined by the most basic groups of needs which are unfulfilled.
2. The individual will systematically satisfy his needs, starting with the most basic and moving up the hierarchy.
3. More basic need groups are said to be *prepotent* in that they will take precedence over all those higher in the hierarchy.

RESEARCH ON MASLOW'S MODEL

Although Maslow first proposed his hierarchy in 1943, it was only in recent years that industrial psychologists attempted to determine its appropriateness as a model for studying motivation in the business firm. The major contribution to the research on Maslow's model in an industrial setting has been by Porter at the University of California. He and his coworkers have published a number of research studies which are of sufficient importance that a brief summary of each will be presented.

STUDY 1 In his first study Porter (1961) administered a survey to 64 bottom-level managers (foremen) and 75 middle-level managers from three different companies. The questionnaire contained 15 items designed to provide information about five different motivational need classes which were derived from Maslow. The 15 items and their need categories were:

I. Security needs
   1. The *feeling of security* in my management position

II. Social needs
   2. The *opportunity* in my management position *to give help to other people*
   3. The *opportunity to develop close friendships* in my management position

III. Esteem needs
   4. The *feeling of self-esteem* a person gets from being in my management position
   5. The *prestige* of my management position *inside* the company (that is, the regard received from others in the company)
   6. The *prestige* of my management position *outside* the company (that is, the regard received from others not in the company)

IV. Autonomy needs
   7. The *authority* connected with my management position
   8. The *opportunity for independent thought and action* in my management position
   9. The *opportunity* in my management position *for participation in the setting of goals*
   10. The *opportunity* in my management position *for participation in the determination of methods and procedures*

V. Self-actualization needs
   11. The *opportunity for personal growth and development* in my management position
   12. The *feeling of self-fulfillment* a person gets from being in my management position (that is, the feeling of being able to use one's own unique capabilities and realize one's potentialities)
   13. The *feeling of worthwhile accomplishment* in my management position

Items specific to two or more need categories
   14. The *pay* for my management position
   15. The feeling of *being-in-the-know* in my management position

Note that Porter added an additional category of autonomy needs and also did not include any questions dealing with physiological needs. Each respondent was asked to indicate for each item:

a. *How much* of the characteristic *is there now* connected with your management position?
b. *How much* of the characteristic do you think *should be* connected with your management position?
c. *How important* is this position characteristic *to you*?

Thus, a typical item looked like this on the questionnaire:

The feeling of self-esteem a person gets from being in my management position:

a	How much is there now?	(min)	1	2	3	4	5	6	7 (max)
b	How much should there be?	(min)	1	2	3	4	5	6	7 (max)
c	How important is this to me?	(min)	1	2	3	4	5	6	7 (max)

Porter then defined *need fulfillment* as the difference between how much there should be and how much is now connected with management position [part (b) minus part (a)]. He then looked at both need fulfillment and importance scores as shown in Table 11.1.

TABLE 11.1 *Summary of Relative Need-fulfillment Deficiency and Need Importance Within Management Levels*

	Bottom Management		*Middle Management*	
Need Categories	*Relative Deficiency*	*Relative Importance*	*Relative Deficiency*	*Relative Importance*
Security	moderate	large	small	large
Social	small	moderate	small	moderate
Esteem	large	small	small	small
Autonomy	large	moderate	moderate	small
Self-actualization	large	large	large	large

SOURCE: L. W. Porter. A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, *45*, 1–10.

Porter's conclusions were as follows (1961, pp. 9–10):

1. The vertical location of management positions appears to be an important variable in determining the extent to which psychological needs are fulfilled.
2. The greatest differences in the frequency of need-fulfillment deficiencies between bottom- and middle-management positions occur in the esteem, security, and autonomy need areas. These needs are significantly more often satisfied in middle than in bottom management.
3. Higher order psychological needs are relatively the least satisfied needs in both bottom and middle management.

4. Self-actualization and security are seen as more important areas of need satisfaction than the areas of social, esteem, and autonomy by individuals in both bottom- and middle-management positions.
5. The highest-order need of self-actualization is the most critical need area of those studied, in terms of both perceived deficiency in fulfillment and perceived importance to the individual, in both bottom and middle management. This need is not perceived as significantly more satisfied at the middle-management level than at the bottom-management level.

STUDY 2 The second study in the series (Porter, 1962) was a much more comprehensive survey in that the questionnaire (minus items 14 and 15) was distributed to a nationwide sample of 6000 managers and executives. Usable returns were obtained from 1916 of these. Table 11.2 shows a summary of the findings.

Note that all the values in Table 11.2 are positive. In other words, in all cases the respondents seem to feel that there should be more opportunity for these needs to be fulfilled than actually existed. The vertical level of position within manage-

TABLE 11.2 *Ranks of Mean Deficiencies in Need Fulfillment for Five Need Categories Within Subgroups of Respondents: Five Management Levels by Four Age Groups*

		Need Categories				
Management Level	*Age Group*	*Security*	*Social*	*Esteem*	*Autonomy*	*Self-actualization*
President	20–34	—	—	—	—	—
	35–44	3	2	4	5	1
	45–54	4	3	2	5	1
	55+	5	2	4	3	1
Vice-President	20–34	5	4	3	2	1
	35–44	3	5	4	2	1
	45–54	3	5	4	2	1
	55+	3	5	4	2	1
Upper-Middle	20–34	5	4	3	2	1
	35–44	4	5	3	2	1
	45–54	4	5	3	2	1
	55+	3	5	4	2	1
Lower-Middle	20–34	4.5	4.5	3	2	1
	35–44	5	4	3	2	1
	45–54	4	5	3	2	1
	55+	4	5	3	1	2
Lower	20–34	5	4	3	2	1
	35–44	3.5	5	3.5	2	1
	45–54	—	—	—	—	—
	55+	—	—	—	—	—

SOURCE: L. W. Porter. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, *46*, 375–384.

ment had a strong relationship to the degree of perceived satisfaction of the three high order needs. Also, the needs of self-actualization and autonomy were consistently regarded as the least fulfilled needs at all levels of management.

STUDY 3 In the third study Porter (1963a) examined the same data obtained in Study 2, looking this time at need importance rather than need fulfillment. He found that higher level managers placed more emphasis upon self-actualization and autonomy needs than did lower level managers. There were no significant differences due to managerial level for the other needs. (See Table 11.3.)

TABLE 11.3 *Ranks of Mean Importance for Five Need Categories Within Each Subgroup of Respondents*

		Need Categories				
Management Level	*Age Group*	*Security*	*Social*	*Esteem*	*Autonomy*	*Self-actualization*
President	20-34	—	—	—	—	—
	35–44	3	4	5	2	1
	45–54	3	4	5	2	1
	55+	3	4	5	2	1
Vice-President	20–34	4	5	3	2	1
	35–44	5	4	3	2	1
	45–54	3.5	3.5	5	2	1
	55+	4	3	5	2	1
Upper-Middle	20–34	5	4	3	2	1
	35–44	5	4	3	2	1
	45–54	3	4	5	2	1
	55+	3	4	5	2	1
Lower-Middle	20–34	5	4	3	2	1
	35–44	5	3	4	2	1
	45–54	3	4	5	2	1
	55+	3	2	5	4	1
Lower	20–34	3	5	4	2	1
	35–44	4	3	5	2	1
	45–54	—	—	—	—	—
	55+	—	—	—	—	—

SOURCE: L. W. Porter. Job attitudes in management: II. Perceived importance of needs as a function of job level. *Journal of Applied Psychology*, 1963, *47*, 141–148.

STUDY 4 All the above studies examined either need fulfillment or need importance as a function of a person's *vertical* location in an organization. In his fourth study Porter (1963b) looked at these two motivational constructs as a function of *horizontal* differences among managerial positions, *viz.*, line versus staff jobs. He divided his managers into three groups: line managers, combined line and staff managers, and staff managers.

His results showed the following:

1. Line managers perceived greater need fulfillment than did staff managers.
2. Largest need-fulfillment differences occurred in the need for esteem and for self-actualization.
3. Line and staff managers did not differ on the importance they attached to each type of need except for the autonomy need, which staff managers found more important.
4. Differences due to horizontal organizational structure are smaller than those due to vertical organizational structure.

STUDY 5 In the next study, Porter (1963c) examined need fulfillment and need importance as influenced by the size of the organization to which a manager belonged. Here he found that at the lower levels of management smaller company managers were more satisfied (had greater perceived need fulfillment) than were large company managers. However, just the *reverse* was found to be true at higher levels of management! Here Porter found that managers in large companies were more satisfied than small company managers. Size of company was found to have no bearing upon perceived need importance.

These findings are extremely interesting and important. As Porter states (1963c, p. 387):

> There are good reasons for presuming that organizational level might have an interaction effect on size in relation to job attitudes. For example, a worker at the bottom of a large organization has a much larger superstructure of organization levels and of sheer numbers of people above him than does a similar worker in a small company. In effect, the worker in the large company has more bosses above him and has less absolute influence on his work environment than does the worker in the small company. However, at the other end of the hierarchy—top management—the picture should be reversed. A top manager in a large company controls or "bosses" more people than a top manager in a smaller organization, and hence has (or should have) more absolute influence in the work situation. To the extent that this analysis of the interaction of size of organization and level of position within the organizational hierarchy is correct, it would lead to the following hypothesis: the higher the organizational level, the relatively more favorable will be the job attitudes of individuals in large organizations compared with those of individuals in small organizations.

SUMMARY OF MASLOW'S THEORY

The work of Porter in applying the Maslowian need structure as a frame of reference for furthering our understanding of management in industry cannot be overemphasized. It marks one of the most meaningful developments in many years. Hopefully, Porter or others will extend the research to types of organizations other than business firms. For example, one might hypothesize that different kinds of organizations such as military, educational, religious, etc., might produce quite different findings than were obtained with the business-firm people.

VROOM'S THEORY OF HUMAN MOTIVATION

Recently Vroom (1964) has proposed a theory of human motivation which, although quite general in form, takes as its context the individual in the world of work. It is probably the first such attempt at formal motivation "model building"

in industrial psychology. Vroom has drawn very heavily on the work of the great social psychologist, Kurt Lewin. Like Lewin, Vroom uses the basic concept of *valence* as a key notion. He defines valence as "the attractiveness of a goal or outcome." Another definition is "the *anticipated satisfaction from an outcome*" (Vroom, 1964, p. 15).

Vroom then outlines two propositions (Vroom, 1964, pp. 17–18).

> *Proposition 1.* The valence of an outcome to a person is a monotonically increasing function of the algebraic sum of the products of the valences for all other outcomes and his conceptions of its instrumentality for the attainment of these other outcomes.
>
> *Proposition 2.* The force on a person to perform an act is a monotonically increasing function of the algebraic sum of the products of the valences of all outcomes and the strength of his experiences that the act will be followed by the attainment of these outcomes.

Proposition 1 may be thought of as follows: The desire (valence) for any particular objective (outcome) on the part of an individual is directly related to the likelihood (instrumentality) that the objective will in turn lead to other subsequent objectives of given desirability (valence). Consider a young man who joins the Navy right out of high school. His valence for joining the Navy might be thought of as being a sum of his feelings about all the outcomes (both good and bad) that are apt to result from this action, each being weighted by its anticipated probability. Thus, he knows he will have to postpone his college education ($p = 1.00$), a fact which may disappoint him. Yet he also knows he may have more pleasant duty than if he were to be drafted ($p = 0.80$), a fact which may please him very much. The combination of an 80 percent chance for something of very high valence contrasted to a 100 percent chance for something of moderate negative valence may result in an overall positive feeling or valence about joining the Navy at this time in his life.

Proposition 2 simply states that the greater the valence of any outcome, the more apt is a person to take action. In the case of our young man who intuitively (or even unconsciously) "computes" his valence toward joining the Navy, we assume that the higher the overall valence of this act the greater the force acting upon him to join.

EVIDENCE FOR VROOM'S MODEL

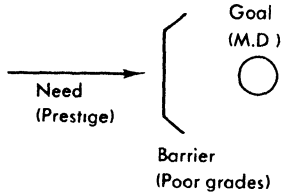
In the following chapter considerable research is examined relevant to the specific topic of job satisfaction and its relationship to job performance. Can the data in that section provide any support for Vroom's Model? Vroom suggests that job satisfaction is a reflection of how desirable a person finds his job—thus, it is a measure of a person's valence for his work situation. Now his model would predict (Proposition 2) that the force to remain on a job should be directly related to the valence of his job. The two measures which are appropriate to testing this hypothesis are (1) turnover and (2) absenteeism. As the data in the next chapter indicate, there does indeed appear to be a moderate relationship between job satisfaction measures and these two variables.

The model is more difficult to apply in the case of the relationship between job satisfaction and work productivity. While it is true that high productivity assures a worker that he will not be fired (and thus removed from a high valence situation),

most workers have no difficulty in being productive enough to avoid this. Thus, one might expect job satisfaction (valence) to have a less clear relationship to work productivity measures than to turnover and absenteeism measures. This appears to be the case, as the data in the next chapter show no consistency in the relationship between satisfaction and productivity measures across a wide range of situations.

MOTIVATION AND FRUSTRATION

Earlier the motivational cycle was defined as a process beginning with a need followed by a response directed toward a goal object. Very often we see situations in life when a goal that would ordinarily satisfy a need is blocked for some reason. For example, consider a person who has a high need for prestige and desires to become a surgeon (the goal) as a means of satisfying the need. Suppose, however, his grades are not good enough for medical school. These grades form a *barrier* around the goal.



When a person is blocked from a desired goal in this fashion we say that he has been *frustrated*. Frustration is a common event in our daily lives, and it is an important concept in trying to understand the behavior of workers. The strength of frustration in a situation is related to the magnitude of the need that is being thwarted.

When a person is frustrated he can respond in either of two ways:

1. *Adaptive responses:* The person may find some new and acceptable way of reducing the need—that is, an acceptable substitute goal which is attainable.
2. *Maladaptive responses:* The person may continue trying to reach the unattainable goal, or he may give up trying to reach any goal whatsoever. One of the standard maladaptive responses to frustration is that of *aggression*.

FRUSTRATION-AGGRESSION HYPOTHESIS

First proposed by Dollard and others at Yale (Dollard *et al.*, 1939), the frustration-aggression hypothesis states that "*aggression is always a consequence of frustration*. More specifically, the proposition is that the occurrence of aggressive behavior always presupposes the existence of frustration and, contrariwise, that the existence of frustration always leads to some form of aggression" (Dollard *et al.*, 1939, p. 1). This hypothesis has been subject to some revision in recent years, due to a body of research evidence which has been accumulated. In its current form it is generally

accepted that aggression is typically produced by frustration, but that being frustrated does not necessarily result in an aggressive response. Responses to frustration are now considered to be of four basic types:

withdrawal responses	attack responses
limitation responses	substitution responses

THE IMPORTANCE OF FRUSTRATION IN INDUSTRY

How important is the concept of frustration in a work situation? Does the work situation necessarily have to be a source of frustration to the worker? Certainly everyone finds his job frustrating at times—indeed, many people might even say that it is the very fact that they do get frustrated which makes the job worth having. To them, the challenge of finding a way to deal effectively with whatever barriers confront them as they strive for a particular goal is what provides the "spice" to their job. They actually look forward to meeting these obstacles, and if they did not occur they would probably find their work very drab and dull. The point is that one can make a good case that for some people it may be that the lack of frustration is the most frustrating situation of all!

Evidence for this notion is actually rather substantial, particularly in the domain of tasks which are very repetitive and unchallenging to the worker. Workers typically find such tasks exceedingly unrewarding, dull, and hence very frustrating. Scott (1966) has recently reviewed this data and has suggested that *activation theory* is a very appropriate model for understanding behavior in this kind of work environment. Briefly, activation theory states that the human organism needs stimulation and variety in its environment; without this, motivation will suffer and frustration may result. To the extent, then, that barriers and obstacles to goals provide variety and stimulation to the worker, they may actually tend to reduce the overall frustration experienced.

GOALS

If the structure of the work situation is such that it is not possible for an employee to select goals or to obtain goals which are necessary to satisfy his needs, then frustration-instigated behavior is likely to occur. If these responses are maladaptive in nature, then the productivity of the worker (and perhaps also that of his coworker) is likely to be affected.

It is therefore important that a system of goals or rewards be available to the employee that will be satisfactory in satisfying whatever needs he brings into the work situation. Eaton (1952) has listed seven different hypotheses about the origin of frustration in a work situation:

1. The worker is frustrated by the insignificance of his group
2. The worker is frustrated by the absentee ownership of the production of his work.
3. The worker is frustrated by unfulfilled expectations of upward mobility which attend his labor.
4. The worker is frustrated by his lack of a defined role and by the many alternatives available in his work.
5. The worker is frustrated by the changing technology and conditions of his work.
6. The worker is frustrated by the isolation of his work within the community.
7. The worker is frustrated by the economic insecurity of his work.

JOB ENLARGEMENT

One technique which has been suggested for reducing the number of frustration-producing situations inherent in many jobs is that of *job enlargement.* The basic notion of job enlargement is that jobs should be designed to use the full capacity of an individual. That is, he should have the opportunity to achieve the maximum in terms of need fulfillment. In a sense, job enlargement is diametrically opposed to the trend in industry toward job specificity which has been occurring during the past several decades, and it will be interesting to observe the effect and impact of this technique as it becomes more widely adopted.

Another argument in favor of the job enlargement notion is that it also follows from activation theory as outlined by Scott (1966) that an "enlarged" job should provide greater variety and stimulation to the worker. One would, therefore, predict that enlarging a job should result in less frustration and greater work motivation. However, the notion of job enlargement, while it has merit, should be viewed with some caution, as it can be overdone. If a job is made too complex, it may overburden the worker to the point that he cannot fulfill its requirements—a situation where a great deal of frustration may result. Figure 11.1 shows how job largeness

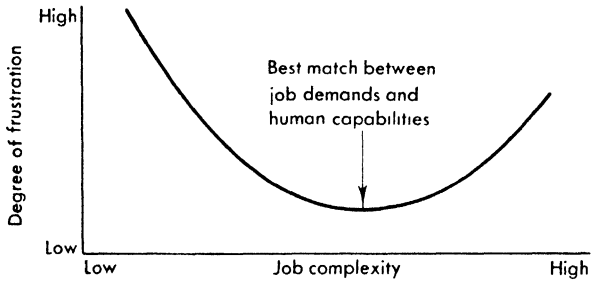


Figure 11.1. Hypothesized relationship between frustration and job complexity.

may be related to frustration. Highest frustration is probably associated with tasks that are either very simple or very complex, and the least frustration is experienced with those tasks which best fit the capacity limits of the individual.

FINANCIAL INCENTIVES AS A MOTIVATING FORCE

In view of the complexity of motivation in man, it can be definitely stated that industry has overemphasized the importance of financial incentives. Some people believe that money is the only incentive in industry, and many believe that it is the most important incentive (see Figure 11.2). This false premise and oversimplification has caused many of the best laid incentive plans to fail. For example, after a six-week drive the Wiremold Company of Connecticut concluded that awarding war bonds to employees for perfect attendance was not the solution to the absenteeism problem. In announcing the discontinuation of the plan, the president of the firm stated that the plan had failed to influence the habits of the employees because most of those who had been irregular in attendance before the award system was inaugurated were still irregular.

A bus slowdown in New York City some years ago was attributed by the third

vice-president of the drivers' union to the drivers' refusal to work overtime. He said that the men were not seeking more money; they were tired, and they thought the company should hire additional drivers. Their average earnings were $3200 a year, an income which, even then, could not be considered very high.

Ever since Taylor's time "experts" have concocted one wage incentive system after another, until there are literally hundreds of such systems. One of the most comprehensive reviews of wage incentive plans was offered by Lytle (1938), who

TABLE 11.4 *Classification of Financial Incentive Plans by Production Earning Characteristics*

(A definite quality, quantity standard must be established and enforced as a prerequisite to any of these plans except I-1.)

Class I. *Employer takes all gain or loss*
1. Time hour, week, or any straight salary rate; not an extra-financial incentive
2. Standard time using two rates, one either side of task; a two-zone multiple time plan
3. Multiple time: arithmetic steps in rate between production zones (sometimes called standard time plan)
4. Multiple time: geometric steps in rate between production zones

Class II. *Employee takes all gain or loss*
5. Piece or straight commission rate, this subdivides into punitive, basic, and high
6. Taylor (multiple piece rate or multiple commission)
7. Merrick (multiple piece rate or multiple commission)
8. Gantt (combination of No. 1 and No. 5 with step between). (Without step, it would be called piece rate with guarantee, Manchester, Standard Hour, 100% Premium, or Haynes Manit. All five have identical earning curves)

Class III. *Gain shared between employer and employee* but day wage guaranteed, excepting in Barth and "one-third premium" form of Halsey
9. Halsey
10. Diemer
11. Baum
12. Bedaux, Dyer, Keays-Weaver, K.I.M., Shanley, and Stevens
13. Ficker Time } (awkward and unsound)
14. Ficker Piece }
15. Sherman Individual-Group (awkward and unsound)
16. Rowan, Mansfield, and Bayle
17. Barth (particularly good for beginners)

Class IV. *Empiric Location of Points Between the Two Variables*
18. Emerson
19. Wennerlund (piece work or commission above 100% production)
20. Knoeppel
21. Bigelow
22. Bigelow-Knoeppel
23. Parkhurst
24. Ernst and Ernst
25. Sylvester

SOURCE: C. W. Lytle. *Wage incentive methods*, rev. ed. Ronald, New York, 1938.

believed that such plans were important in any cost production problem. According to him, incentive payments have two advantages: (1) an increase in production per unit and (2) an increase in employee earnings. He believed that the advantages of a properly selected, well-installed, and ably managed wage payment plan accrue annually to employees and employers. Of course, Lytle did not say that these advantages "accrue mutually and fairly," for they do not. As Mark Spade says in his brilliant book, *How To Run a Bassoon Factory*, wage incentive systems are a "means of paying employees more—but not so much."

From the point of view of the industrial psychologist, it is extremely difficult to differentiate among the various financial incentive systems. In some respects, the only essential difference is who obtains the consulting fee. Lytle has attempted to classify all financial incentive plans on the basis of production-earning characteristics. His results, presented in Table 11.4, give an idea of the numerous systems that have been promoted.

The plans given in Table 11.4 are really only of historical interest. While during the 1930s and 1940s there was great interest in such plans, very few industrial firms actually employ such incentive system today. Most wage systems are under union control, nearly all are hourly, and the sales profession is probably the one remaining job which does consistently use an incentive pay scheme of one sort or another. In modern industry the emphasis has shifted away from financial rewards to reward systems based upon other values. Not only have financial incentive systems diminished in importance, but they do not work as effectively as they are assumed to. The Bank Wiring Observation Room experiment in the Hawthorne plant (see Chapter 10) showed clearly that restricted rather than unlimited production characterized the workers. There is no doubt that employees and employers need money, if for no other reason than that it is the recognized medium of exchange. However, just how much a person needs is a subject for much speculation. Money is a necessary thing, but only in some instances does it operate as an incentive. Some people can just get along on $5,000 a year, others on $10,000, and still others on $20,000. Although the low-income group does not appreciate the financial problems of the high-income group, both have such problems—problems that from their respective points of view are of the greatest intensity to both groups. It is not a question of which is right or wrong. Money is needed for food, clothing, shelter, medical care, education, luxuries, social position, and power. Even when people desire money out of proportion to the amounts specifically desired by others, the existence of their need cannot be denied. The power drive, and the belief that money is a source of power, can be just as real a need to one person as the need for money to obtain food is to others. Furthermore, the amount needed for food varies according to size of family, living habits, etc. We cannot talk of the need for subsistence on an absolute basis because too many variables are involved.

Stockford and Kunze (1950) report a study in which the results suggest that the value of wages is relative and not absolute. They compared the attitudes of workers and starting salaries in relation to salary on the previous job. Figure 11.2 presents the results, which indicate that employees with the same starting salary may have favorable or unfavorable sentiment toward the company. "Unfavorable" starting wages (that is, less than previous wage) exert a persistent and detrimental effect upon sentiment toward company, employment stability, and performance. The negative attitude is significantly more intense than the positive attitude resulting from "favorable" starting wage rates.

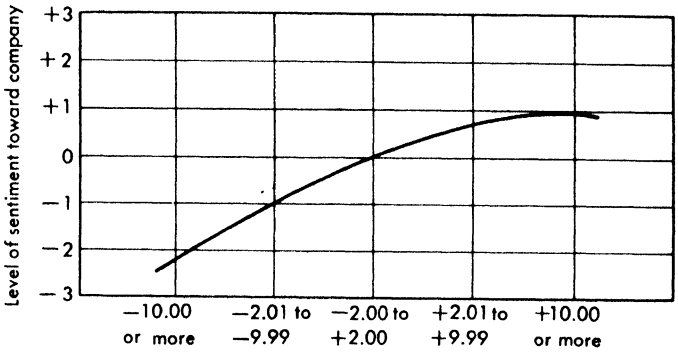


Figure 11.2. Relationship between level of sentiment and starting wage rates related to previous salary. (From L. O. Stockford and K. R. Kunze. Psychology and the pay check. *Personnel*, 1950, 27, 129–143.)

THEORIES CONCERNING MONEY AS AN INCENTIVE

Are there any theoretical foundations in psychology for the role of money as an incentive for the worker? Opsahl and Dunnette (1966) have listed five such theories which attempt to explain the effect of money on job performance.

1. *Money as a Generalized Conditioned Reinforcer.* This theory states that money, because it is associated with more basic reinforcers (need satisfiers), acquires the status of a secondary reinforcer. Studies by Wolfe (1936) and Cowles (1937) demonstrated that poker chips acquired reinforcement value when they could be exchanged for food—that is, subjects in their experiments worked as hard to get chips which could "buy" food as they had previously worked for the food itself.

2. *Money as a Conditioned Incentive.* According to this point of view, money acquires an incentive value due to continued pairing with other incentives of a more basic nature. While there is some laboratory data to indicate this kind of conditioning can occur, applying this paradigm to the actual work situation is difficult. For example, in our society what are the other, more basic incentives with which money is paired?

3. *Money as an Anxiety Reducer.* The notion here is that the absence of money is generally associated with anxiety reactions in our society, and therefore lack of money becomes a conditioned stimulus for an anxiety response. Money, in turn, therefore, serves as a reducer of anxiety.

4. *Money as a "Hygiene Factor."* In the chapter on job satisfaction (Chapter 12) the Herzberg, Mausner, and Snyderman notion of "satisfiers" and "dissatisfiers" as being two distinct classes of motivators is discussed in some detail. For them, money is a "hygiene factor" which serves primarily as a dissatisfier—that is, it causes dissatisfaction when absent, but contributes little to satisfaction when present. However, there is some question as to whether their data actually support their hypothesis (Opsahl and Dunnette, 1966).

5. *Money as a Means of "Instrumentality."* Earlier in this chapter, the motivational postulates of Vroom's theory of motivation were briefly outlined. In Vroom's model money acquires valence due to its perceived instrumentality for obtaining other desired outcomes. For example, if money were perceived as being instrumental to the positive goal of security, then money would acquire a positive valence. Then the likelihood of a person acting in a way to acquire money would be a function

of his need for security multiplied by the expectancy that a particular money-seeking act would result in a specified monetary return.

SUMMARY OF THE ROLE OF MONEY

In their review on the role of financial compensation in industrial motivation, Opsahl and Dunnette (1966, p. 115) state:

> Very little is known about the behavioral laws regulating the effectiveness of incentives. We continue to dole out large sums of money under the guise of "incentive pay" without really knowing much about its incentive character. We do not know, for instance, the nature of the effect of a pay raise or the length of time before that effect occurs; or, for that matter, how long the raise may be effective. Nor do we know the optimal reinforcement schedule to be used in giving salary increases for obtaining desired changes in job behavior. A simple monitoring of work outputs on jobs where amount of production is under the direct control of the employee and where it is easily assessed, may provide valuable information here. Such knowledge would have important implications for how often and in what amounts incentive raises should be built into the compensation package.

Thus, the role of money as a reward in modern industry continues to be poorly understood in spite of the fact that our economy seems to be based upon a monetary reward system. Although the use of formal incentive pay systems has diminished in recent years in favor of other nonfinancial incentives, money units will probably always be exchanged for work units. Thus, a better knowledge of the dynamics of this exchange process is certainly needed.

KINDS OF INCENTIVES

Incentives may be classified "financial" and "nonfinancial," but it seems more desirable to classify them as "competitive" and "cooperative." In both education and industry, the values of incentives based upon cooperation are beginning to be realized. Organizing people into meaningful groups and having them work together toward a common goal results in more learning at the educational level and greater production at the industrial level.

The term *incentive* is used for a spurring force that is introduced as a means of achieving a goal. An incentive bolsters activity in the direction of the goal. An individual may introduce his own incentives, in which case they are extremely meaningful. However, in industry the incentive is often used as a means of spurring on an employee toward someone else's goal. He may not be interested in either reduced costs or increased production, but he may be persuaded to achieve this goal by an incentive. When the incentive results in satisfaction for the employee and the achievement of the goal results in equal satisfaction for the employer, the plant is being successfully operated. The major difficulty with financial incentives has been the fact that the additional money has not meant as much to the employee as the employer assumed it would. The outstanding reason for employee objections to such plans has been the fear of a rate cut. This fear is present, regardless of whether it is justified or not. There is also the fear of a layoff; and so the worker does not respond to the magic of a financial incentive to the extent that its proponents would have us believe.

Both competition and cooperation may be regarded as incentives. Competition calls for one individual to do better than the next individual. Cooperation calls for people to contribute equal and maximum efforts toward a common goal. Competition and cooperation are not mutually exclusive, especially when individuals cooperate in groups to compete with other groups. Industry in the past has placed too much emphasis upon competition; only recently have the advantages of cooperative behavior as an incentive been recognized. The use of cooperation as an incentive holds considerable promise of success.

RESEARCH ON INCENTIVES

BLUM AND RUSS STUDY

In the mind of the employee a minimum of five drives are likely to operate on the job. In alphabetical order they are advancement, hours of work, salary, security, and supervisor relationships. A study conducted by Blum and Russ (1942) attempted to determine the relative importance of these five incentives. The data were obtained from a group of 286 gainfully employed people ranging in age from 17 to 60 years, and in occupation from unskilled manual labor to the professions. There were 181 men, 72 of whom were married, and 105 women, 26 of whom were married. All the subjects lived in the New York City area.

The questionnaire used in the study is shown in Figure 11.3. An examination of it reveals that each of the five incentives is compared with the other four. Actually, this results in ten comparisons. Each time one incentive is preferred to the other in the comparison, a score of one is assigned to it; in other words, a maximum of ten points is distributed among the five incentives. Consider, for example, the following score:

Advancement	3
Security	3
Salary	2
Supervisor	1
Hours of work	1

This score means that the individual checked salary twice out of a possible 4 comparisons, advancement three times out of a possible 4 comparisons, etc.

The subjects in this study were told to answer the questionnaire as it applied to their own feelings rather than to any general attitude. The purpose of the questionnaire was not discussed with any subject prior to his completing it.

Although the number of subjects was small and therefore the interpretations and conclusions may be in error because of the limited sampling, repetitions of this questionnaire have yielded similar results. This study's results are presented not so much to indicate absolute findings, as to illustrate the possible uses to which this questionnaire can be put.

The total sample was divided into thirds on the basis of when the completed

Sex _____ Age _____ Married or single _____

Title of present job _____

Circle number applying:

How many jobs have you held?

1 2 3 5 8 10 More than 10

How many years have you been employed?

1 2 3 5 8 10 More than 10

How many years with present employer?

1 2 3 5 8 10 More than 10

How many employees in present firm?

5 10 25 50 100 More than 100

In the following list, please check one item in each pair which you consider more important to you in a job. There is no correct or model answer. Each answer is indicative of your attitude.

_____ Receive more pay and have an insecure job, or
_____ Receive less pay and have a secure job.

_____ Have a friendly supervisor and work more hours, or
_____ Have an unfriendly supervisor and work fewer hours.

_____ Receive more pay and have no advancement possibilities, or
_____ Receive less pay and have advancement possibilities.

_____ Have a secure job and work more hours, or
_____ Have an insecure job and work fewer hours.

_____ Have a friendly supervisor and no advancement possibilities, or
_____ Have an unfriendly supervisor and have advancement possibilities.

_____ Receive more pay and have an unfriendly supervisor, or
_____ Receive less pay and have a friendly supervisor

_____ Have a secure job and no advancement possibilities, or
_____ Have an insecure job and advancement possibilities

_____ Receive more pay and work more hours, or
_____ Receive less pay and work fewer hours.

_____ Have a secure job and an unfriendly supervisor, or
_____ Have an insecure job and a friendly supervisor

_____ Work fewer hours and have no advancement possibilities, or
_____ Work more hours and have advancement possibilities

Interviewer _____

Figure 11.3. Attitude questionnaire. (From M. L. Blum and J. Russ. A study of employee attitudes toward various incentives. *Personnel*, 1942, *19*, 438–444.)

questionnaires were returned by the interviewers. From Table 11.5 it appears that not only are the ranks of the various incentives similar, but also only slight differences in the averages exist.

The attitude of men and women toward the various incentives is shown in Table 11.6 as an average score on each incentive for the two groups. In addition, the

number of times a particular incentive was preferred is shown as a percentage of the total obtainable for that incentive. An analysis of these figures indicates that the two sexes tend to agree as to the relative importance of the various incentives. The single exception is the women's rating of supervisor ahead of salary; for the men this is reversed. Both groups rate advancement and security in first and second place and agree that hours of work are least important. The men consider advancement and salary much more important as incentives than the women do. But supervisor relationship is more important for the women than for the men. Security and hours of work are not statistically differentiated within these groups.

A further analysis was made on the basis of marital status (Table 11.7). Both the married men and the single men attach approximately the same importance to salary, supervisor, and hours. Security is more important to the married men and advancement is more important to the single men. The women show no statistically significant differences among the various incentives with the single exception of hours of work; the married women attach greater importance to working fewer hours

TABLE 11.5 *Average Score of First and Third Groups*

	Salary	*Security*	*Supervisor*	*Hours*	*Advancement*
First third	1.64	2.75	1.36	0.82	3.43
Third third	1.76	2.66	1.31	0.82	3.48

TABLE 11.6 *Scores and Percentages for Each Sex on the Questionnaire*

	Men		*Women*	
	Score	*Percent of Total Score Attainable*	*Score*	*Percent of Total Score Attainable*
Salary	1.84	46	1.43	39
Security	2.77	69	2.86	72
Supervisor	1.33	33	1.79	45
Hours of work	0.59	15	0.86	21
Advancement	3.47	87	3.06	76

TABLE 11.7 *Attitude Toward Various Incentives According to Marital Status*

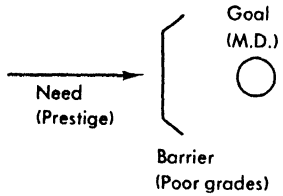
	Men		*Women*	
	Married	*Single*	*Married*	*Single*
Salary	46%	46%	34%	36%
Security	76	65	65	73
Supervisor	32	34	51	45
Hours of work	13	16	29	18
Advancement	83	89	71	78

SOURCE, TABLES 11.5–11.7: M. Blum and J. Russ. A study of employee attitudes toward various incentives. *Personnel*, 1942, *19*, 438–444.

most workers have no difficulty in being productive enough to avoid this. Thus, one might expect job satisfaction (valence) to have a less clear relationship to work productivity measures than to turnover and absenteeism measures. This appears to be the case, as the data in the next chapter show no consistency in the relationship between satisfaction and productivity measures across a wide range of situations.

MOTIVATION AND FRUSTRATION

Earlier the motivational cycle was defined as a process beginning with a need followed by a response directed toward a goal object. Very often we see situations in life when a goal that would ordinarily satisfy a need is blocked for some reason. For example, consider a person who has a high need for prestige and desires to become a surgeon (the goal) as a means of satisfying the need. Suppose, however, his grades are not good enough for medical school. These grades form a *barrier* around the goal.



When a person is blocked from a desired goal in this fashion we say that he has been *frustrated*. Frustration is a common event in our daily lives, and it is an important concept in trying to understand the behavior of workers. The strength of frustration in a situation is related to the magnitude of the need that is being thwarted.

When a person is frustrated he can respond in either of two ways:

1. *Adaptive responses:* The person may find some new and acceptable way of reducing the need—that is, an acceptable substitute goal which is attainable.
2. *Maladaptive responses:* The person may continue trying to reach the unattainable goal, or he may give up trying to reach any goal whatsoever. One of the standard maladaptive responses to frustration is that of *aggression*.

FRUSTRATION-AGGRESSION HYPOTHESIS

First proposed by Dollard and others at Yale (Dollard *et al.*, 1939), the frustration-aggression hypothesis states that "*aggression is always a consequence of frustration*. More specifically, the proposition is that the occurrence of aggressive behavior always presupposes the existence of frustration and, contrariwise, that the existence of frustration always leads to some form of aggression" (Dollard *et al.*, 1939, p. 1). This hypothesis has been subject to some revision in recent years, due to a body of research evidence which has been accumulated. In its current form it is generally

accepted that aggression is typically produced by frustration, but that being frustrated does not necessarily result in an aggressive response. Responses to frustration are now considered to be of four basic types:

withdrawal responses	attack responses
limitation responses	substitution responses

THE IMPORTANCE OF FRUSTRATION IN INDUSTRY

How important is the concept of frustration in a work situation? Does the work situation necessarily have to be a source of frustration to the worker? Certainly everyone finds his job frustrating at times—indeed, many people might even say that it is the very fact that they do get frustrated which makes the job worth having. To them, the challenge of finding a way to deal effectively with whatever barriers confront them as they strive for a particular goal is what provides the "spice" to their job. They actually look forward to meeting these obstacles, and if they did not occur they would probably find their work very drab and dull. The point is that one can make a good case that for some people it may be that the lack of frustration is the most frustrating situation of all!

Evidence for this notion is actually rather substantial, particularly in the domain of tasks which are very repetitive and unchallenging to the worker. Workers typically find such tasks exceedingly unrewarding, dull, and hence very frustrating. Scott (1966) has recently reviewed this data and has suggested that *activation theory* is a very appropriate model for understanding behavior in this kind of work environment. Briefly, activation theory states that the human organism needs stimulation and variety in its environment; without this, motivation will suffer and frustration may result. To the extent, then, that barriers and obstacles to goals provide variety and stimulation to the worker, they may actually tend to reduce the overall frustration experienced.

GOALS

If the structure of the work situation is such that it is not possible for an employee to select goals or to obtain goals which are necessary to satisfy his needs, then frustration-instigated behavior is likely to occur. If these responses are maladaptive in nature, then the productivity of the worker (and perhaps also that of his coworker) is likely to be affected.

It is therefore important that a system of goals or rewards be available to the employee that will be satisfactory in satisfying whatever needs he brings into the work situation. Eaton (1952) has listed seven different hypotheses about the origin of frustration in a work situation:

1. The worker is frustrated by the insignificance of his group.
2. The worker is frustrated by the absentee ownership of the production of his work.
3. The worker is frustrated by unfulfilled expectations of upward mobility which attend his labor.
4. The worker is frustrated by his lack of a defined role and by the many alternatives available in his work.
5. The worker is frustrated by the changing technology and conditions of his work.
6. The worker is frustrated by the isolation of his work within the community.
7. The worker is frustrated by the economic insecurity of his work.

JOB ENLARGEMENT

One technique which has been suggested for reducing the number of frustration-producing situations inherent in many jobs is that of *job enlargement*. The basic notion of job enlargement is that jobs should be designed to use the full capacity of an individual. That is, he should have the opportunity to achieve the maximum in terms of need fulfillment. In a sense, job enlargement is diametrically opposed to the trend in industry toward job specificity which has been occurring during the past several decades, and it will be interesting to observe the effect and impact of this technique as it becomes more widely adopted.

Another argument in favor of the job enlargement notion is that it also follows from activation theory as outlined by Scott (1966) that an "enlarged" job should provide greater variety and stimulation to the worker. One would, therefore, predict that enlarging a job should result in less frustration and greater work motivation. However, the notion of job enlargement, while it has merit, should be viewed with some caution, as it can be overdone. If a job is made too complex, it may overburden the worker to the point that he cannot fulfill its requirements—a situation where a great deal of frustration may result. Figure 11.1 shows how job largeness

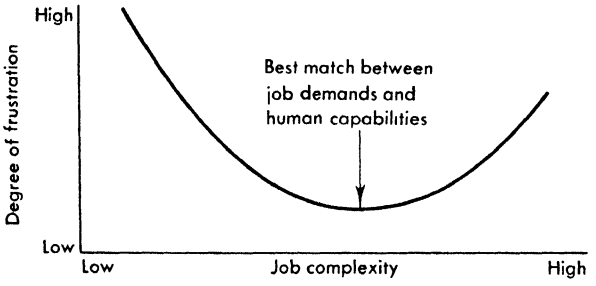


Figure 11.1. Hypothesized relationship between frustration and job complexity.

may be related to frustration. Highest frustration is probably associated with tasks that are either very simple or very complex, and the least frustration is experienced with those tasks which best fit the capacity limits of the individual.

FINANCIAL INCENTIVES AS A MOTIVATING FORCE

In view of the complexity of motivation in man, it can be definitely stated that industry has overemphasized the importance of financial incentives. Some people believe that money is the only incentive in industry, and many believe that it is the most important incentive (see Figure 11.2). This false premise and oversimplification has caused many of the best laid incentive plans to fail. For example, after a six-week drive the Wiremold Company of Connecticut concluded that awarding war bonds to employees for perfect attendance was not the solution to the absenteeism problem. In announcing the discontinuation of the plan, the president of the firm stated that the plan had failed to influence the habits of the employees because most of those who had been irregular in attendance before the award system was inaugurated were still irregular.

A bus slowdown in New York City some years ago was attributed by the third

vice-president of the drivers' union to the drivers' refusal to work overtime. He said that the men were not seeking more money; they were tired, and they thought the company should hire additional drivers. Their average earnings were $3200 a year, an income which, even then, could not be considered very high.

Ever since Taylor's time "experts" have concocted one wage incentive system after another, until there are literally hundreds of such systems. One of the most comprehensive reviews of wage incentive plans was offered by Lytle (1938), who

TABLE 11.4 *Classification of Financial Incentive Plans by Production Earning Characteristics*

(A definite quality, quantity standard must be established and enforced as a prerequisite to any of these plans except I-1.)

Class I. *Employer takes all gain or loss*
1. Time: hour, week, or any straight salary rate; not an extra-financial incentive
2. Standard time using two rates, one either side of task; a two-zone multiple time plan
3. Multiple time: arithmetic steps in rate between production zones (sometimes called standard time plan)
4. Multiple time: geometric steps in rate between production zones

Class II. *Employee takes all gain or loss*
5. Piece or straight commission rate; this subdivides into punitive, basic, and high
6. Taylor (multiple piece rate or multiple commission)
7. Merrick (multiple piece rate or multiple commission)
8. Gantt (combination of No. 1 and No. 5 with step between). (Without step, it would be called piece rate with guarantee, Manchester, Standard Hour, 100% Premium, or Haynes Manit. All five have identical earning curves)

Class III. *Gain shared between employer and employee* but day wage guaranteed, excepting in Barth and "one-third premium" form of Halsey
9. Halsey
10. Diemer
11. Baum
12. Bedaux, Dyer, Keays-Weaver, K.I.M., Shanley, and Stevens
13. Ficker Time } (awkward and unsound)
14. Ficker Piece }
15. Sherman Individual-Group (awkward and unsound)
16. Rowan, Mansfield, and Bayle
17. Barth (particularly good for beginners)

Class IV. *Empiric Location of Points Between the Two Variables*
18. Emerson
19. Wennerlund (piece work or commission above 100% production)
20. Knoeppel
21. Bigelow
22. Bigelow-Knoeppel
23. Parkhurst
24. Ernst and Ernst
25. Sylvester

SOURCE: C. W. Lytle. *Wage incentive methods,* rev. ed. Ronald, New York, 1938.

believed that such plans were important in any cost production problem. According to him, incentive payments have two advantages: (1) an increase in production per unit and (2) an increase in employee earnings. He believed that the advantages of a properly selected, well-installed, and ably managed wage payment plan accrue annually to employees and employers. Of course, Lytle did not say that these advantages "accrue mutually and fairly," for they do not. As Mark Spade says in his brilliant book, *How To Run a Bassoon Factory*, wage incentive systems are a "means of paying employees more—but not so much."

From the point of view of the industrial psychologist, it is extremely difficult to differentiate among the various financial incentive systems. In some respects, the only essential difference is who obtains the consulting fee. Lytle has attempted to classify all financial incentive plans on the basis of production-earning characteristics. His results, presented in Table 11.4, give an idea of the numerous systems that have been promoted.

The plans given in Table 11.4 are really only of historical interest. While during the 1930s and 1940s there was great interest in such plans, very few industrial firms actually employ such incentive system today. Most wage systems are under union control, nearly all are hourly, and the sales profession is probably the one remaining job which does consistently use an incentive pay scheme of one sort or another. In modern industry the emphasis has shifted away from financial rewards to reward systems based upon other values. Not only have financial incentive systems diminished in importance, but they do not work as effectively as they are assumed to. The Bank Wiring Observation Room experiment in the Hawthorne plant (see Chapter 10) showed clearly that restricted rather than unlimited production characterized the workers. There is no doubt that employees and employers need money, if for no other reason than that it is the recognized medium of exchange. However, just how much a person needs is a subject for much speculation. Money is a necessary thing, but only in some instances does it operate as an incentive. Some people can just get along on $5,000 a year, others on $10,000, and still others on $20,000. Although the low-income group does not appreciate the financial problems of the high-income group, both have such problems—problems that from their respective points of view are of the greatest intensity to both groups. It is not a question of which is right or wrong. Money is needed for food, clothing, shelter, medical care, education, luxuries, social position, and power. Even when people desire money out of proportion to the amounts specifically desired by others, the existence of their need cannot be denied. The power drive, and the belief that money is a source of power, can be just as real a need to one person as the need for money to obtain food is to others. Furthermore, the amount needed for food varies according to size of family, living habits, etc. We cannot talk of the need for subsistence on an absolute basis because too many variables are involved.

Stockford and Kunze (1950) report a study in which the results suggest that the value of wages is relative and not absolute. They compared the attitudes of workers and starting salaries in relation to salary on the previous job. Figure 11.2 presents the results, which indicate that employees with the same starting salary may have favorable or unfavorable sentiment toward the company. "Unfavorable" starting wages (that is, less than previous wage) exert a persistent and detrimental effect upon sentiment toward company, employment stability, and performance. The negative attitude is significantly more intense than the positive attitude resulting from "favorable" starting wage rates.

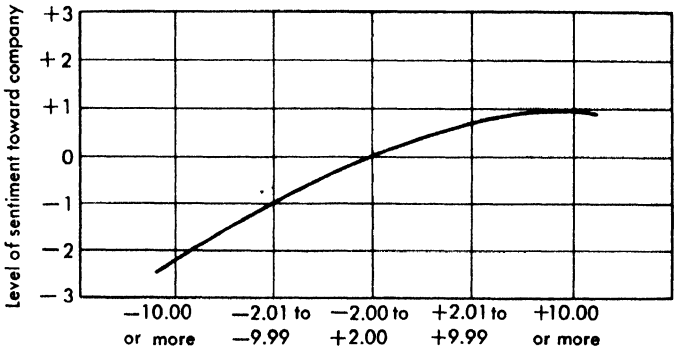


Figure 11.2. Relationship between level of sentiment and starting wage rates related to previous salary. (From L. O. Stockford and K. R. Kunze. Psychology and the pay check. *Personnel*, 1950, 27, 129–143.)

THEORIES CONCERNING MONEY AS AN INCENTIVE

Are there any theoretical foundations in psychology for the role of money as an incentive for the worker? Opsahl and Dunnette (1966) have listed five such theories which attempt to explain the effect of money on job performance.

1. *Money as a Generalized Conditioned Reinforcer.* This theory states that money, because it is associated with more basic reinforcers (need satisfiers), acquires the status of a secondary reinforcer. Studies by Wolfe (1936) and Cowles (1937) demonstrated that poker chips acquired reinforcement value when they could be exchanged for food—that is, subjects in their experiments worked as hard to get chips which could "buy" food as they had previously worked for the food itself.

2. *Money as a Conditioned Incentive.* According to this point of view, money acquires an incentive value due to continued pairing with other incentives of a more basic nature. While there is some laboratory data to indicate this kind of conditioning can occur, applying this paradigm to the actual work situation is difficult. For example, in our society what are the other, more basic incentives with which money is paired?

3. *Money as an Anxiety Reducer.* The notion here is that the absence of money is generally associated with anxiety reactions in our society, and therefore lack of money becomes a conditioned stimulus for an anxiety response. Money, in turn, therefore, serves as a reducer of anxiety.

4. *Money as a "Hygiene Factor."* In the chapter on job satisfaction (Chapter 12) the Herzberg, Mausner, and Snyderman notion of "satisfiers" and "dissatisfiers" as being two distinct classes of motivators is discussed in some detail. For them, money is a "hygiene factor" which serves primarily as a dissatisfier—that is, it causes dissatisfaction when absent, but contributes little to satisfaction when present. However, there is some question as to whether their data actually support their hypothesis (Opsahl and Dunnette, 1966).

5. *Money as a Means of "Instrumentality."* Earlier in this chapter, the motivational postulates of Vroom's theory of motivation were briefly outlined. In Vroom's model money acquires valence due to its perceived instrumentality for obtaining other desired outcomes. For example, if money were perceived as being instrumental to the positive goal of security, then money would acquire a positive valence. Then the likelihood of a person acting in a way to acquire money would be a function

of his need for security multiplied by the expectancy that a particular money-seeking act would result in a specified monetary return.

SUMMARY OF THE ROLE OF MONEY

In their review on the role of financial compensation in industrial motivation, Opsahl and Dunnette (1966, p. 115) state:

> Very little is known about the behavioral laws regulating the effectiveness of incentives. We continue to dole out large sums of money under the guise of "incentive pay" without really knowing much about its incentive character. We do not know, for instance, the nature of the effect of a pay raise or the length of time before that effect occurs; or, for that matter, how long the raise may be effective. Nor do we know the optimal reinforcement schedule to be used in giving salary increases for obtaining desired changes in job behavior. A simple monitoring of work outputs on jobs where amount of production is under the direct control of the employee and where it is easily assessed, may provide valuable information here. Such knowledge would have important implications for how often and in what amounts incentive raises should be built into the compensation package.

Thus, the role of money as a reward in modern industry continues to be poorly understood in spite of the fact that our economy seems to be based upon a monetary reward system. Although the use of formal incentive pay systems has diminished in recent years in favor of other nonfinancial incentives, money units will probably always be exchanged for work units. Thus, a better knowledge of the dynamics of this exchange process is certainly needed.

KINDS OF INCENTIVES

Incentives may be classified "financial" and "nonfinancial," but it seems more desirable to classify them as "competitive" and "cooperative." In both education and industry, the values of incentives based upon cooperation are beginning to be realized. Organizing people into meaningful groups and having them work together toward a common goal results in more learning at the educational level and greater production at the industrial level.

The term *incentive* is used for a spurring force that is introduced as a means of achieving a goal. An incentive bolsters activity in the direction of the goal. An individual may introduce his own incentives, in which case they are extremely meaningful. However, in industry the incentive is often used as a means of spurring on an employee toward someone else's goal. He may not be interested in either reduced costs or increased production, but he may be persuaded to achieve this goal by an incentive. When the incentive results in satisfaction for the employee and the achievement of the goal results in equal satisfaction for the employer, the plant is being successfully operated. The major difficulty with financial incentives has been the fact that the additional money has not meant as much to the employee as the employer assumed it would. The outstanding reason for employee objections to such plans has been the fear of a rate cut. This fear is present, regardless of whether it is justified or not. There is also the fear of a layoff; and so the worker does not respond to the magic of a financial incentive to the extent that its proponents would have us believe.

Both competition and cooperation may be regarded as incentives. Competition calls for one individual to do better than the next individual. Cooperation calls for people to contribute equal and maximum efforts toward a common goal. Competition and cooperation are not mutually exclusive, especially when individuals cooperate in groups to compete with other groups. Industry in the past has placed too much emphasis upon competition; only recently have the advantages of cooperative behavior as an incentive been recognized. The use of cooperation as an incentive holds considerable promise of success.

RESEARCH ON INCENTIVES

BLUM AND RUSS STUDY

In the mind of the employee a minimum of five drives are likely to operate on the job. In alphabetical order they are advancement, hours of work, salary, security, and supervisor relationships. A study conducted by Blum and Russ (1942) attempted to determine the relative importance of these five incentives. The data were obtained from a group of 286 gainfully employed people ranging in age from 17 to 60 years, and in occupation from unskilled manual labor to the professions. There were 181 men, 72 of whom were married, and 105 women, 26 of whom were married. All the subjects lived in the New York City area.

The questionnaire used in the study is shown in Figure 11.3. An examination of it reveals that each of the five incentives is compared with the other four. Actually, this results in ten comparisons. Each time one incentive is preferred to the other in the comparison, a score of one is assigned to it; in other words, a maximum of ten points is distributed among the five incentives. Consider, for example, the following score:

Advancement	3
Security	3
Salary	2
Supervisor	1
Hours of work	1

This score means that the individual checked salary twice out of a possible 4 comparisons, advancement three times out of a possible 4 comparisons, etc.

The subjects in this study were told to answer the questionnaire as it applied to their own feelings rather than to any general attitude. The purpose of the questionnaire was not discussed with any subject prior to his completing it.

Although the number of subjects was small and therefore the interpretations and conclusions may be in error because of the limited sampling, repetitions of this questionnaire have yielded similar results. This study's results are presented not so much to indicate absolute findings, as to illustrate the possible uses to which this questionnaire can be put.

The total sample was divided into thirds on the basis of when the completed

Sex _____ Age _____ Married or single _____

Title of present job _____

Circle number applying:

How many jobs have you held?

1 2 3 5 8 10 More than 10

How many years have you been employed?

1 2 3 5 8 10 More than 10

How many years with present employer?

1 2 3 5 8 10 More than 10

How many employees in present firm?

5 10 25 50 100 More than 100

In the following list, please check one item in each pair which you consider more important to you in a job. There is no correct or model answer. Each answer is indicative of your attitude.

_____ Receive more pay and have an insecure job, or
_____ Receive less pay and have a secure job.

_____ Have a friendly supervisor and work more hours, or
_____ Have an unfriendly supervisor and work fewer hours.

_____ Receive more pay and have no advancement possibilities, or
_____ Receive less pay and have advancement possibilities.

_____ Have a secure job and work more hours, or
_____ Have an insecure job and work fewer hours.

_____ Have a friendly supervisor and no advancement possibilities, or
_____ Have an unfriendly supervisor and have advancement possibilities

_____ Receive more pay and have an unfriendly supervisor, or
_____ Receive less pay and have a friendly supervisor.

_____ Have a secure job and no advancement possibilities, or
_____ Have an insecure job and advancement possibilities.

_____ Receive more pay and work more hours, or
_____ Receive less pay and work fewer hours.

_____ Have a secure job and an unfriendly supervisor, or
_____ Have an insecure job and a friendly supervisor

_____ Work fewer hours and have no advancement possibilities, or
_____ Work more hours and have advancement possibilities.

Interviewer _____

Figure 11.3. Attitude questionnaire. (From M. L. Blum and J. Russ. A study of employee attitudes toward various incentives. *Personnel*, 1942, *19*, 438–444.)

questionnaires were returned by the interviewers. From Table 11.5 it appears that not only are the ranks of the various incentives similar, but also only slight differences in the averages exist.

The attitude of men and women toward the various incentives is shown in Table 11.6 as an average score on each incentive for the two groups. In addition, the

number of times a particular incentive was preferred is shown as a percentage of the total obtainable for that incentive. An analysis of these figures indicates that the two sexes tend to agree as to the relative importance of the various incentives. The single exception is the women's rating of supervisor ahead of salary; for the men this is reversed. Both groups rate advancement and security in first and second place and agree that hours of work are least important. The men consider advancement and salary much more important as incentives than the women do. But supervisor relationship is more important for the women than for the men. Security and hours of work are not statistically differentiated within these groups.

A further analysis was made on the basis of marital status (Table 11.7). Both the married men and the single men attach approximately the same importance to salary, supervisor, and hours. Security is more important to the married men and advancement is more important to the single men. The women show no statistically significant differences among the various incentives with the single exception of hours of work; the married women attach greater importance to working fewer hours

TABLE 11.5 *Average Score of First and Third Groups*

	Salary	*Security*	*Supervisor*	*Hours*	*Advancement*
First third	1.64	2.75	1.36	0.82	3.43
Third third	1.76	2.66	1.31	0.82	3.48

TABLE 11.6 *Scores and Percentages for Each Sex on the Questionnaire*

	Men		*Women*	
	Score	*Percent of Total Score Attainable*	*Score*	*Percent of Total Score Attainable*
Salary	1.84	46	1.43	39
Security	2.77	69	2.86	72
Supervisor	1.33	33	1.79	45
Hours of work	0.59	15	0.86	21
Advancement	3.47	87	3.06	76

TABLE 11.7 *Attitude Toward Various Incentives According to Marital Status*

	Men		*Women*	
	Married	*Single*	*Married*	*Single*
Salary	46%	46%	34%	36%
Security	76	65	65	73
Supervisor	32	34	51	45
Hours of work	13	16	29	18
Advancement	83	89	71	78

SOURCE, TABLES 11.5–11.7: M. Blum and J. Russ. A study of employee attitudes toward various incentives. *Personnel*, 1942, *19*, 438–444.

than do the single women. The married men emphasize security more than the married women do, and security is more important to the single women than to the single men. The married women attach greater importance to the supervisor than do the married men; this difference is seen also in comparing the single men and women. Attitude toward salary is unaffected by marriage, a finding similar to that reported by Cole (1940). The married men care least about the number of hours worked and the married women care most; no difference is observed between the single men and women. Advancement is more important among the single men than among the married men; this is also true for women.

Inasmuch as attitude toward work may be a product of age, the two groups were separated into the following age categories: less than 20, 20 to 29, and 30 and above. Table 11.8, which presents the results of this analysis, indicates that the group over 30 is less interested in advancement than are the younger groups; this is true for both sexes. This older group also places greater emphasis on security than do the younger groups and attaches less importance to the supervisor relationship. The women, as they grow older, are more interested in working fewer hours. It is to be noticed that salary remains relatively constant for both sexes regardless of age.

TABLE 11.8 *Attitude as Affected by Age Toward Various Incentives*

	Men			*Women*		
	Less Than 20 Years	*20–29 Years*	*30 Years and Above*	*Less Than 20 Years*	*20–29 Years*	*30 Years and Above*
Salary	43%	47%	45%	39%	39%	37%
Security	72	65	78	71	70	77
Supervisor	38	32	33	46	48	42
Hours of work	10	17	13	17	20	27
Advancement	88	89	81	78	78	67

SOURCE: M. Blum and J. Russ. A study of employee attitudes toward various incentives. *Personnel*, 1942, *19*, 438–444.

The data on the number of jobs held and the years employed were analyzed; the results were very similar to those obtained on the basis of age. Naturally, the older a person is, the greater the opportunity for more jobs and longer service with an employer. However, a much larger group of subjects would be required before it could definitely be determined whether there actually is this close relationship among jobs held, years employed, and age.

The study also analyzed the attitudes of the two groups in relation to the total number of employees in the firms where they worked. A comparison of the attitudes of people who work for a firm with less than 5 employees and those in companies with 100 or more employees shows that the latter rate advancement as much more important than do the former. The supervisor is much more important in small firms. The other three incentives are rated alike by both groups.

To summarize, this study measured the attitudes of employees toward five incentives by the paired-comparison technique. It found that advancement and security were the two most important incentives, with hours of work the least important. Salary was rated third by the men and fourth by the women. Supervisor relationship was ranked fourth by the men and third by the women. Married men considered

security more important and advancement less important than did the single men. The married women were more interested in working fewer hours than were the single women. The significance of advancement as an incentive decreased with age for both sexes and security became more important with age, but advancement and security were considered more important than salary. The attitudes of employees changed depending on such major classifications as sex, marital status, and age.

JURGENSEN STUDY

Jurgensen (1947) had a group of 150 female and 1189 male applicants for positions at the Minneapolis Gas Light Company complete a job preference questionnaire. Each subject was asked to rank ten items in order of preference. The mean rank for each of the items is presented in Table 11.9.

TABLE 11.9 *Job Preferences of Men and Women Applicants*

	Mean Rank	
	Men (1189)	*Women (150)*
Security	3.1	4.7
Advancement	3.4	4.5
Type of work	3.7	2.8
Company	4.6	5.0
Coworkers	6.1	5.6
Pay	6.3	6.5
Supervisor	6.3	5.2
Hours	6.9	6.3
Working conditions	7.3	6.2
Benefits	7.3	8.2

SOURCE: C. E. Jurgensen. Selected factors which influence job preferences. *Journal of Applied Psychology*, American Psychological Association, 1947, *31*, 553–563.

Jurgensen found that job preferences were affected more by extent of education than by most other variables. Advancement became more important and security less important as education increased. Jurgensen concludes that management and labor leaders err in statements and demands with regard to job applicants' preferences. According to his findings, wages, hours, and working conditions are not so important as is generally thought, and type of work is more important than is ordinarily stated.

STAGNER STUDY

Stagner (1950) reports a study based upon 7000 employees of a nationwide corporation. The results are presented in Table 11.10. After reviewing various researches, Stagner concludes that "neither executives nor workers are concerned about pay as such, except when economically pinched. At other times they prefer ego-satisfactions such as prestige, power, recognition, security, and individual treatment.

TABLE 11.10 *Ratings of Importance by 7000 Workers on Various Job Factors*

	Percent of Workers Including This Item in the First Five	*Percent of Workers Choosing This Item as First Choice*
A steady job	61.9	36.1
Pay rate	52.6	7.2
A chance to get ahead	41.9	6.9
A square boss	39.6	4.8
Working on the job you prefer	35.3	15.2
Credit for the job you do	29.6	2.2
Vacations and holidays	21.5	0.4
Friendly working companions	21.3	0.7
Medical and health facilities	20.8	0.6
Pension	9.7	7.1

SOURCE: R. Stagner. Psychological aspects of industrial conflict, II, Motivation. *Personnel Psychology,* 1950, *3,* 1–16.

JONES AND JEFFREY STUDY

Jones and Jeffrey (1964) reported a study in which they evaluated preferences among electrical workers for alternate forms of job compensation. They considered each of four different compensation dimensions:

1. Hourly wage versus weekly wage
2. Merit incentive versus no merit incentive
3. Piece work incentive versus no piece work incentive
4. Regular pay rate versus higher pay rate

They found that for nonunion workers the weekly wage, the merit incentive, the piecework incentive, and the higher wage were all significantly preferred. With union members, however, the no-merit incentive choice was significantly preferred, as was a piece rate and higher pay rate. The interesting aspect of the data from union members was the fact that they appeared willing to take jobs with no merit incentive and regular pay in preference to jobs having merit incentives and high pay—they would forfeit money to stay clear of a merit system! This would certainly seem to argue that the union, as an organization, provides the worker with some form of "group protection" or security that is worth a sacrifice in actual money units. He no longer has to bargain for himself, but is part of a larger body which bargains in much broader terms.

NEALEY STUDY

In another recent study, Nealey (1964) has examined worker preferences for different employee benefit programs. He had 1133 members of an electrical trade union paired—compare the following six employee benefit options:

1. The company will pay the cost of an additional $50 a month pension to be added to my retirement benefits.
2. I will get a 6 percent raise.
3. The normal workweek shall be cut to 37½ hours without any reduction in weekly earnings.
4. The company will pay the entire cost of full hospital insurance for myself and my family.
5. It will be agreed that all regular employees must be members of the union. (The union shop will be put into effect.)

6. I will have three weeks' paid vacation a year in addition to my present vacation; the extra vacation to be taken when I choose.

The six plans were chosen so as to be of approximately equal cost to the employer. The relative preferences for the six plans are shown in Figure 11.4.

In general the workers preferred the hospital insurance option and the union shop and were quite unimpressed with the enticement of a 37½-hour week. Nealey also examined preferences as a function of age of the worker. These data are given in Figure 11.5.

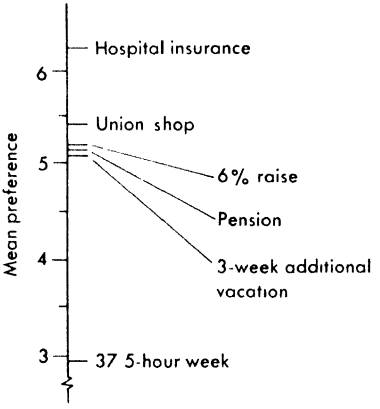


Figure 11.4. Relative preference for pay, union shop, and four fringe benefits. (From S. M. Nealey. Determining worker preference among employee benefit programs. *Journal of Applied Psychology*, 1964, 48(1), 7–12.)

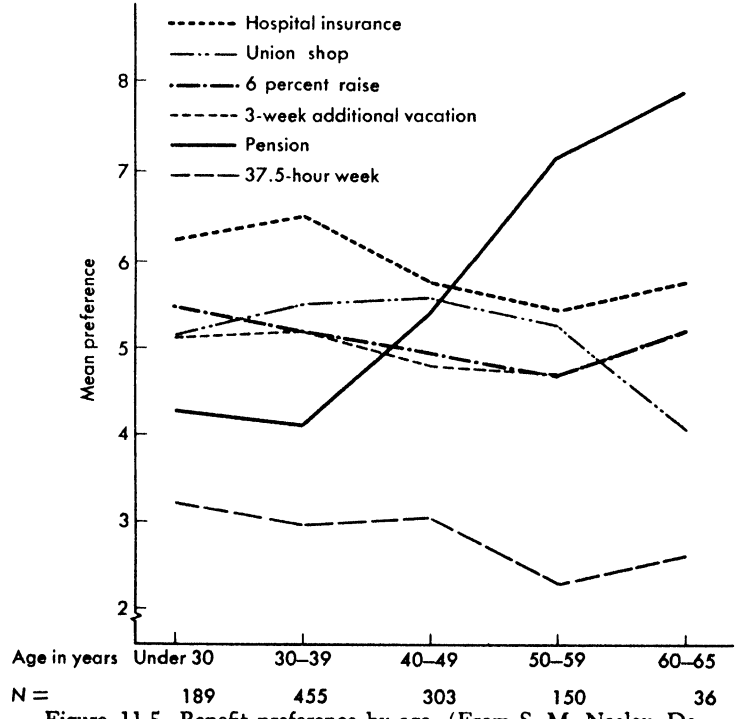


Figure 11.5. Benefit preference by age. (From S. M. Nealey. Determining worker preference among employee benefit programs. *Journal of Applied Psychology*, 1964, 48(1), 7–12.)

Some rather predictable patterns emerge in terms of the age factor. Note particularly the tremendous increase in preference for the pension plan as one moves into the older age groups. Note also that the union shop is judged most desirable (relative to the other plans) among middle-aged workers, and that the 3-week additional vacation was most preferred among younger workers, systematically becoming less preferred as one moves through the older groups.

DUDYCHA-NAYLOR STUDY

In a study designed to examine what job characteristics were of primary importance to today's college students, Dudycha and Naylor (1966) found that the students in their sample gave the following rank order of job traits:

1. Pay
2. Opportunity for growth and advancement
3. Security
4. Fellow employees
5. Working conditions
6. Supervisor-employee relations

These were students who for the most part had never held a regular job and who were indicating what things they felt would be important when they did start choosing between job opportunities. It is interesting to note the high value placed upon pay by this group.

SUMMARY OF FINDINGS

A pause to compare the results of the above research is desirable. Although different methods were used and the studies included different complexions of sampling, the results tend to be more similar than dissimilar. The findings substantiate the view that salary is not the most important factor on a job, nor does it operate as the all-powerful incentive. It follows that appealing to stronger motives for work will result in more highly motivated workers.

The views of Harold Ruttenberg (1941) are interesting in this connection. Although unions are often believed to have as their single purpose the obtaining of more money for their members, they may satisfy the employee's tremendous need for self-expression. Ruttenberg, who was formerly Research Director of the Steel Workers Organizing Committee, emphasizes this need when he says: "The urge for a means of self-expression is usually present in every individual in an industrial plant, and consciously or unconsciously (usually the latter) each individual constantly seeks some way to express himself." In his experience in industry he has found the need for self-expression to be a basic incentive, as important as the desire for economic betterment and personal security, and in some instances even more important.

The Survey Research Center of the University of Michigan has conducted some studies on the relationship between productivity, supervision, and employee morale. In connection with the study done in an insurance company it stated (1948, p. 10):

> People are more effectively motivated when they are given some degree of freedom in the way in which they do their work than when every action is prescribed in advance. They do better when some degree of decision-making about their jobs is possible than

when all decisions are made for them. They respond more adequately when they are treated as personalities than as cogs in a machine. In short, if the ego motivations of self-determination, of self-expression, of a sense of personal worth can be tapped, the individual can be more effectively energized. The use of external sanctions, of pressuring for production, may work to some degree, but not to the extent that the more internalized motives do. When the individual comes to identify himself with his job and with the work of his group, human resources are much more fully utilized in the production process.

MASLOW'S THEORY AND FINANCIAL INCENTIVES

How does the fact that pay seems to be unimportant as a motivator fit with the theory of motivation proposed by Maslow? The first response might be that pay satisfies only the lower order needs such as the physiological and safety needs. Therefore it has little or no motivating effect upon the typical work situation where these basic needs are satisfied already. Recently this position has been severely criticized by Lawler, who has collected data from large numbers of executives which indicate that salary is indeed very important to these men. He suggests that pay is a unique incentive, one that is able to satisfy *all* levels of needs.

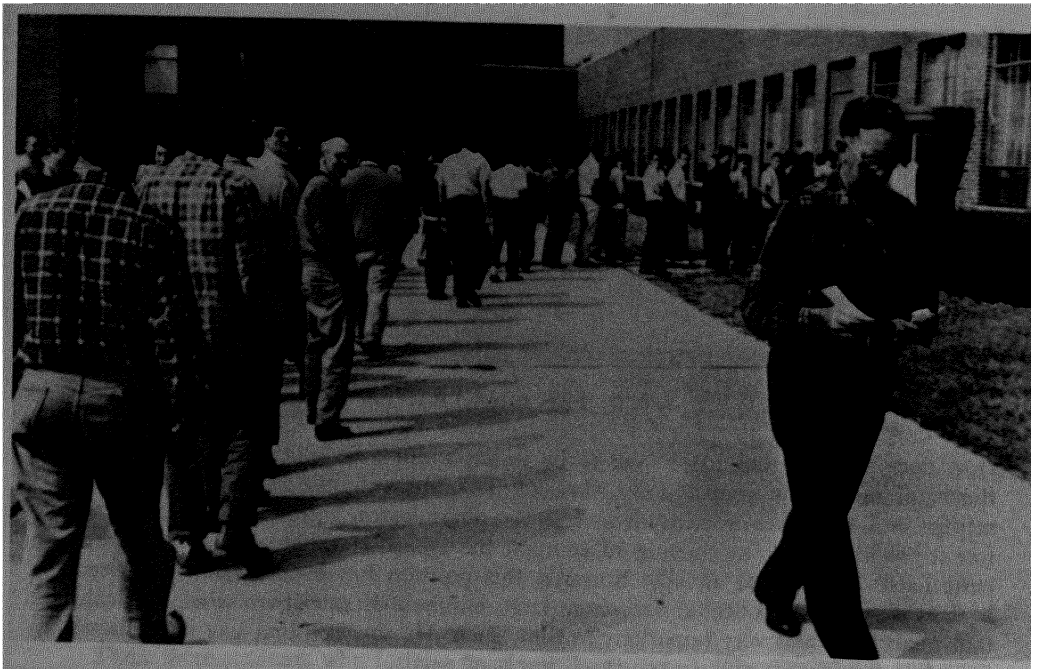
For example, Lawler (1965) cites one executive as saying, when asked why salary was important to him, "It is just like bridge—it isn't any fun unless you keep score." Thus this manager was not interested in pay in terms of what goods and services it could buy, but rather in terms of its being a symbol or yardstick of success.

Lawler proposed that one of the major reasons that one fails to find a pay incentive being more important in the studies that have been carried out is that in many instances pay is not related to merit in the actual job situation. He cites evidence that indicates that many managers simply do not feel that their job performance will affect their pay. In one survey he found that when managers were asked what factors they felt determined their pay, they responded by indicating that training and experience were the most important factors. Is there any reason to believe that pay can ever be an incentive for better job performance if the system in turn does not use job performance as a criterion for determining pay? Lawler does not think so, and your authors would tend to agree.

WHY PEOPLE WORK

Life, whatever it is, can be characterized as activity, and during the course of it people are continually active, even when asleep (everyone turns and moves during a normal night's sleep). Work is a form of activity that has social approval and satisfies a real need of the individual to be active. To produce, to create, to gain respect, to acquire prestige, and incidentally to earn money—these are some of the reasons that people work. The paycheck must mean many different things to different workers.

In our society there are few instances of forced idleness; but when they occur, they are usually characterized by unpleasantness. People in jails or hospitals and the



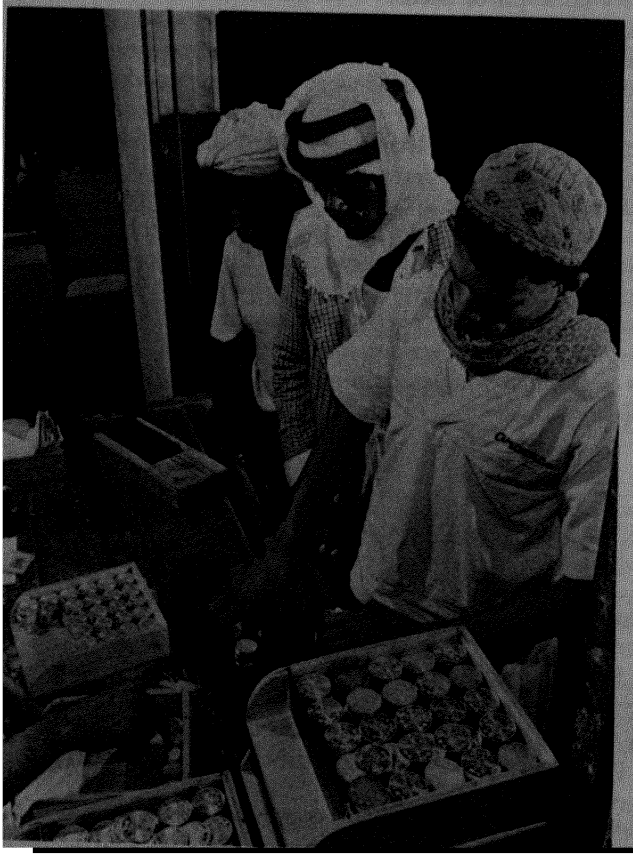


Figure 11.6. Do people work f money only? Top: Workmen at th General Motors Boxwood Road assem bly plant near Wilmington, Delawar collect their paychecks covering wag earned before they and other Gener Motors workers went on strike. (Wid World.) Bottom: Arabs fingerprint th payroll receipt and pick up their silv riyals. (Courtesy Standard Oil Co [N.J.].)

unemployed are examples of forced idleness. Almost everyone in jail wants to get out; even the hospital patients who have all their wants attended to and who receive tremendous attention from attractive nurses, besides their family's overabundant concern, still yearn to get back on the job. Similarly, the average worker waits an entire year for his two-week vacation and plans how he will do nothing but rest. In some cases the vacation results in a need for a real vacation, but the employee handles this somehow when he is back on the job—within a month he is making plans for resting up on his next vacation.

In 1935 Sidney Roslow conducted a survey on the attitude of 98 relief workers toward work relief or home relief. Although the money received by these two groups of people on relief differed, by no stretch of imagination could either be regarded as financially affluent. Everyone in the entire group preferred work relief to home relief because it allowed him to maintain his self-respect. Only 3 percent based their preference on the fact that work relief paid more; 43 percent preferred it because of the effect it had on their morale, and 50 percent preferred it because they were paid for working or they preferred working to being idle.

Many theories have been advanced to explain work. The Gilbreths believed that work could be made pleasant. Henry Ford, on the other hand, believed that it was basically unpleasant and that therefore the hours of work should be as short as possible and the pay as high as possible.

One author has been successful in stumping his classes over a period of years by requesting them to differentiate between work and play. Is a major league baseball player, or a man who plays this game on his only day off, working or playing? Is a business executive entertaining an important customer at a nightclub at work or at play? Is an amateur fisherman who gets up at 3 A.M. to go fishing at work or at play?

Nor does an attempt to differentiate between work and play on a monetary or voluntary basis help. Is the volunteer nurse at work or at play? Forced or required activity for which one does not receive pay may be work, but it also may be play. When a psychologist mows the lawn under a hot sun because the lawn needs attention, he is a gardener; but is it fair to say that the gardener and the psychologist are both working? Of course there are differences between work and play, but in many instances the only difference is in the attitude toward the activity or the label attached to it.

We must recognize the important values which are attached to the many needs expressed and fulfilled in work. Mastery of the job and the importance of self-expression are two such values. An example is this little story told by Bob Burns:

> If you want a job done well, give it to a man who takes pride in his work. You may have to wait a little longer for it, but when you get it, it'll be done right. We've got an old tailor down home that has the reputation of being the best pants maker in seven counties, but you can't rush him.
>
> A traveling man heard of his reputation one time, and while he was in town he ordered a pair of pants from this old tailor. The salesman left town and came back in two weeks and called up the tailor and the tailor says, "Your pants ain't done yet but I'm working on 'em." A month later the salesman came back to town and his pants wasn't ready yet.
>
> Finally on his fourth trip he dropped around to the tailor's shop and found the old tailor had just finished the last stitch. The salesman was pretty mad. He turned to the tailor and he says, "It took six days to make the world and it took you eighty days to make these pants!" The old tailor said, "Yes, but just take a look at the world and then take a look at them pants!"

Few people are able to continue a job that they regard as meaningless or that they really know they cannot do well. Near a certain station of a railroad there is a gatetender whose job is to lower the gates when trains pass his station. It is with considerable pride that this man, day in and day out, lowers the gates and then stands where he can see the engineer. His behavior is so routinized that in many respects it is compulsive. When a train is a certain distance away he starts to wave with a large circular motion, and the expression on his face changes from pleasure to delight and finally to sheer ecstasy. When the engineer comes alongside him his hand is upraised and his face is beaming; anyone can see that this man enjoys his job. Apparently the engineers enjoy his antics too, because they toot their whistles out of all proportion to safety needs; each one responds to the greeting in his own way. This man is very different from most gatetenders, who regard each passing train as an interference with peaceful slumber and each engineer as a nuisance.

Tell the man who is polishing automobiles or drawing intricate plans for a house that he has done an excellent job and he will find reward in his work. Mastery of the job and pride in successful accomplishment are real reasons why people work and work hard. This applies to all jobs that are recognized as jobs, regardless of where they stand in the social hierarchy. The busboy hopes some day to be a waiter; therefore he likes to work with a waiter who will give him a chance to learn. The plumber's helper wants to be a master plumber, and the college instructor wants to be a full professor.

Although there are not many case studies of the average, typical, normal worker, the few that have been made indicate that the financial incentive does not play an overwhelming role in a man's working life. Prestige, social acceptance, pride in work, self-expression, and the many related social drives are all equal to financial incentive or even more meaningful. An interesting book edited by Jack Common (1938)[1] describes the life and work history of seven ordinary Englishmen: a plasterer, steel worker, unemployed man, gas worker, market man, blast-furnace man, and a railroad worker. To read these industrial autobiographies of average-typical-menial-marginal workers is to become convinced that money is not the most important incentive to a worker. A workingman does not separate his living from his working life. He is a human embodying many motives and conflicts, and his behavior is as unpredictable on the job as it is in life. These seven people have written autobiographies emphasizing their work experiences and attitudes. Although each story is very different from the others, all have a common core in that they describe part of that large group of people known as employees. Wisely enough, no attempt has been made to type workers on the basis of these seven autobiographies.

Jack Hilton, the plasterer, is a man who takes great pride in his work. As he puts it: "Plastering is essentially human and it makes its tradesmen human. Our material comes from Mother Earth and we use it with skill and muscle. Relative to men on other occupations we may appear primitive. That is our good fortune. . . . We get the satisfaction that we have really done something; something that a machine hasn't done; something that has called for greater patience, ingenuity and skill than a four-minute job."

According to Hilton, the plasterer is proud of his long trade-union history. He is proud of doing a job well, is skillful, and knows how to be adaptable. Furthermore, his work, his employers, and his fellow employees give him variety. A job

[1] Jack Common (ed.). *Seven shifts.* E. P. Dutton & Co., Inc., New York, copyright 1938, renewal © 1966 by Jack Common. By permission of E. P. Dutton and Secker & Warburg Limited, London.

usually lasts from six to ten weeks, and when it is unpleasant he has the consolation of knowing it will be over soon.

In speaking of his work's disadvantages, Hilton mentions the time lost because of slack periods and bad weather. He also dislikes the long and costly trips to and from a job. The common menace faced by all plasterers is the lime in the materials which sometimes gets into the eyes.

The plasterer has a higher status than the laborer, and he shows this superiority in many ways. With respect to the plasterer's relation to the guffer (boss), Hilton writes, "I can work harder for a guffer that is a decent sort and leaves the job to me than for one who is never satisfied and is always coming around trying to pull a bit more out of me." At times the men will compete vigorously with one another and work at a terrific pace. They shout, "Stuff, stuff, stuff" and use it up as fast as the laborers can bring it. Finally one of the men will capitulate, whereupon they return to a more normal pace in which they do not try to "murder themselves."

Hilton is not necessarily a typical plasterer, but he is typical of a man who is proud of his work.

James Stirling is a very different person. He had a job in a steel mill and was preoccupied with the dangers in the rollers and other machinery in the mill. He had a strong dislike for the fumes, dirt, and primitive work conditions. Finally thrown out of work because of a slack, he held a series of jobs such as knocking bricks from a kiln, working with a barrow, operating a crane, and doing clerical work.

Stirling is rather bitter about the depression and its effect, and also about management experts. He writes: "Trainees drift through large workshops in two years or so, spending a month here, six weeks there, going over every department, learning 'workshop control,' developing themselves to be managers sometime in the future. They are probably decent lads, but they don't know a damn thing when it is all over that this or that scrub worker at fifty shillings a week cannot do with his eyes shut and without thinking."

Will Oxley, the unemployed man, tells a vivid and lurid story of what goes on inside a man when he has no work over a long period.

Herbert Mannion has a job in a "gas works" but apparently his basic motive is the desire for drink; work is merely an interlude between consuming a pint or so. He is pessimistic and somewhat fatalistic about his work: "So it goes on—until a slump sends me to the dole and the Means Test, accident to the hospital or early old age dumps me on the scrap-heap. Whichever of these finishes my working career, I shall end up with nothing in the bank and not so much as a thank you from the people who have profited by my labour. You all come into the world with nothing and you can't take anything out. Well, that's a working lad's life, anyway."

Simon Blumenfeld, who operates a stall in a market, is aware of the ills suffered by minority groups. He tells about the practical jokes that are played on these people and their ways of adjusting to the situation. He is aware of the "social and racial groupings" that gravitate to certain occupations as the line of least resistance. He does not consider the life of a small trader a happy one. According to him, "If the small trader could be profitably absorbed in industry, this would probably be a good thing for everyone concerned. He works longer hours than most of our craft unions would allow, and becoming more and more the rule, draws less money than the lowly salaried worker at the end of the week." He must hang on because,

even if he was once a craftsman, he cannot go back to his trade, and a job as a shop assistant is closed to him because these positions are usually filled by young people.

J. H. Watson, who works in a blast furnace, is a "strong and cursing man" who has little respect for anyone who does not sweat while he works. In describing the process that leads finally to the tapping of the molten iron, he writes, "No doubt this sounds very dull but I can assure those people who have never seen a blast furnace that I have worked on one for a dozen years and assisted at two or more tappings a shift, and I still get a feeling of awe, sometimes a mood of exhilaration, when we tap." He is dissatisfied with his inadequate and primitive living quarters and is concerned about the many evil manifestations of the power of money. He believes that the man who lives next to a cesspool is more precious than the man who condemns him to live there. He considers the laborer an outcast who smells a bit, but he regards the remainder of society as dependent upon his existence.

T. A. McCullock, a fireman on a railroad, counts himself one of the lucky ones because he has had no unemployment, receives steady pay, and each year is given three free passes for travel. He admits that he gets tired of the monotonous regularity of railway work, but he still believes the day he applied for the job was a lucky one.

He is most concerned with what the other workers think. Warnings and fines hurt because their recipient feels that his prestige among his fellow workers is lowered. As McCullock says, "The worker always has his pride in being equal to his job. The company, of course, is just his natural enemy. He could evade the company's watch on him and be comfortably inefficient. But he can't escape his mate's opinion of him. His mates don't object to him drinking while on duty; they don't care if he steals from the company. Damn good workers do both."

McCullock is critical of leadership but believes that it does no good to blame it because the self-styled leaders have no clear mandate. He thinks that the general view of employees is that the dumber they are the better, and that brains and imagination are discouraged. "Most of the time we'd be content to let the self-styled clever folk carry on, provided they left us a loop hole to be free in. For it's a big job to be responsible for, and we know too much about jobs in general to rush at one before we have some idea of how it should be done."

Another illustration is afforded by quoting a delightful essay in a column by William Chapman White. The man he refers to is probably not too highly motivated by any incentive which is extraneous to his way of life.

Bird of $5 Passage[2]

He came to the front door and said: "Your garden could use a good weeding, if you'll excuse my saying so." No one argued that. "I'm a gardener," he added, "and I'm looking for work by the day. How about it?"

He was a tall cadaverous man past sixty, as thin as a flower stalk. To judge from his appearance he wouldn't last an hour in the hot sun. He wore the oldest of felt hats and a coat sweater over a flannel shirt that had been washed so often that it had faded to the palest gray. In spite of that costume, he managed an air of elegance, slightly bored, the air of a man who had been everywhere and seen everything and wasn't the slightest impressed.

[2] Copyright, 1954, *New York Herald Tribune*.

Because the weeds were high he was hired Not only did he survive the first four hours in the sun but he did as much work as any two men might be expected to do. The weeds were out, and neatly raked soil showed where he had passed.

As he rested for a few minutes in the afternoon he pointed to a robin. "I'm like one of those birds. North in summer, as far north as I can get. I spend my winters in Florida——" and he made that sentence sound as if he traveled there by his own ninety-foot yacht.

"I don't go to Florida directly, but I stop here and there, the same way I come north. I work at a place until I have $5 for bus fare and $10 cash in my pocket, then I move on. I keep out of the big cities—that's where you get into trouble. I work a little while in Jersey or Maryland and then in Virginia, the Carolinas, and on to Florida, and I keep going as far as I can get, right to Key West. In the spring I head north the same way. I used to go east and west the way I now go north and south, but that's too long a distance at my age. It takes too long to get where I'm going, even though where I'm going's never very important."

He refused a cigarette. "Cuts my wind for my journeying, I find And, mister, I've journeyed. There's no place in America I haven't been You name any town and I'll tell you the name of its main street. Bismarck, North Dakota? That would be Main Street. Dover, New Hampshire? Central Avenue. Corpus Christi, Texas? Broadway."

He picked up the rake and looked over the garden "I'm a gardener because I like plants," he continued "But I've been many other things When I hit a place I look for gardens, and if no one wants me I take any job I can find I guess I've had more jobs in my time than most anybody. I've tended horses on a fancy ranch in California and took care of a sick man for three days in Bisbee, Arizona, while his wife went off to care for a sick mother. I've sold tickets to a roadside zoo below Palm Beach and worked with a popcorn stand in a carnival until the wife of the owner got mad at him one day and slipped in cold cream instead of butter "

He grinned at that. "I was companion to a sort of sick man once, and the family paid me $300 to stay with him, but he paid me $500 the next day to quit, so I quit, but I paid back that first $300, of course. I've helped trap eels, and I worked for a lady once in Jersey tending a little lawn where she docked me a dollar if she found a weed.

"I've never owned a thing in my life but my clothes, and never wanted to. I guess it all doesn't add up to much, but it adds up as much for me as for any bird like that robin, except they come north to raise a family, and I've not even bothered with that and don't miss it none."

After the third day of work he asked to be paid. He said as he left. "See you in the morning."

He didn't show up the next day nor ever again. His story would be unfinished if a neighbor hadn't happened on the end of it.

"I saw him on the bus north the other night," the neighbor said. "He attracted my attention because I heard him say at the bus ticket window: 'Mister, how far north can I go for five dollars? That's where I want a ticket to.' "

This gardener and the previously mentioned seven ordinary working-class men are just as human as everyone else. They differ in their jobs and their motives and their way of life. None of them believe that they are dumb; they are critical, sometimes even contemptuous, of those who "run the show." They do not really want to take over the functions of these others, but they do want a chance for self-expression and they feel a crying need to be understood.

The case study approach to the problem of the motivation of people in industry can lead to productive results and better understanding. Although efficiency systems often assume that there are shortcuts in motivating people, there are none that are worthwhile.

AN OVERALL VIEW OF MOTIVATION AND WORK

Morse, Weiss, and Griggs (1954) report after a survey that work has a double function. Gainful employment enables people to get money to support their families and themselves. Work also relates to society. It gives people a feeling of "place" or "role." In other words, work not only allows a person to exist but also tends to stabilize his place in society.

The place or role that a worker perceives is determined not only by individual values, drives, motives, and sentiments but also by the manner in which the worker relates to his group. A man at work, regardless of his job level, is part of a social structure in and outside the plant, factory, or office. He is a member of an informal group of colleagues or coworkers and possibly a member of a formal group. He is also part of the company structure. Some organizations already recognize this and encourage feelings of group identification with the company.

Relations with coworkers, supervisors, and community may often influence what motivates a person to produce or restrict production, remain on the job or leave, or absent himself from work. These social forces may be as strong or stronger than certain isolated incentives offered by management which are interpreted by the worker as "out of context."

Too often the motive to increase production is management's and not the worker's. When the incentive to increase production is perceived differently by labor, then it does not work or it is only temporarily effective.

A bowling team may promote socialization. The employer who believes that employees will produce more because they can bowl may be making an unwarranted assumption. The workers may accept the bowling opportunity as offered and believe that it has nothing to do with producing more. Chances are if management indicated that only those who produced more could bowl, many workers would not accept the recreational activity. In other words, bowling may be interpreted by both employer and employee as a good-will gesture. This narrow illustration may well serve as the nub of the problem. Management's goal and employees' goal may be perceived differently, and when this happens, the two groups will also perceive a specific incentive differently, as operating toward each one's particular goal.

To understand employee and employer effectively, we must start with a knowledge of the complexity of motivation and the ways in which motives differ for different people. Age, education, marital status, and the host of personal characteristics also contribute to the different meanings that certain incentives have.

An understanding of motivation will contribute knowledge to the way a person perceives his role in society. By further understanding his attitudes, job satisfaction, and the way he relates to his work group, we can then begin to approximate a meaningful perspective on the variety of problems confronting man in his workaday world.

It is misleading to talk about motivating workers without specifying the source as well as the goal of the motivation. Optimally effective motivation may require a unity of employer and employee sources and goals of motivation. This unity may exist only in Utopia. A most effective worker motivation may be impossible if the goals of employer and employee differ. This would mean that motivating employees according to employer goals may not always work. If the prime consideration is to motivate employees most effectively, then management may have to relinquish some of its own goals.

It is highly probable that maximum job satisfaction may interfere with maximum potential production. It is possible that maximum satisfaction cannot occur when one works intensely and at a forced pace. It may well be that when top production is desired, job satisfaction at the moment may be less than optimal.

Goals must be perceived as short-term and long-term. A short-term goal of high production may be achieved under certain conditions of motivation and at the same time vitiate a long-term goal of high production. Chances are the "gimmicks" fail in the long run because an achievement of a short-term goal was mistaken for that of a long-term goal. One of the major difficulties with industrial psychology is that the measures of production, training effectiveness, turnover, etc., may be indicative of immediate effect rather than long-term accomplishment. What is needed is more clear-cut understanding of long-term goals and objectives and how to achieve them.

Short-term goals are most likely to be effective when they reinforce and lead to the satisfaction of long-term goals as well. Short-term goals may be only temporarily effective when they have no relationship to long-term goals. Under certain circumstances, the short-term goal may be in conflict with the long-term goal and yet operate for a short time. The success of short-term goals may need continuing and ever-increasing reinforcement. Some of the difficulties with incentives in industry is that the long-term goal of the employer is perceived as a short-term goal by the employee. After the employee has experienced the clash between this goal and his own long-term goal, the short-term goal is no longer effective.

No attempt will be made to list short-term and long-term goals of employers and employees. Such a listing would be erroneous and would limit thinking and experimentation. Probably long-term goals vary from person to person and even in the same person from time to time. Short-term goals have much less stability and need more immediate satisfactions.

Possibly an example will illustrate the point of view. A person has a long-term goal to acquire money. Short-term goals of money are then meaningful. Such a person is likely to save not spend. Another individual has as his long-term goal relaxation and recreation as offered by boating, golfing, country-club life, and cruises. The short-term goal of money is effective only as a means of having enough to take a cruise, and the opportunity to make money at the same time the vacation is desired may be rejected.

On a somewhat different economic level, one man may want to work and save, but another may only want to meet expenses. Offering him more money for working overtime may be meaningless to him. The girl who works in a factory and is looking forward to getting married may be temporarily moved to produce more units, but will definitely prefer to stay out late on a date and not go to work the next day if she feels that the date is fun and may lead into marriage and that she is not in danger of losing her job or can get another one if she does. Endless examples can be offered. The important point is that meaningful research is needed. To date we have concerned ourselves more with ideal planning so that we can isolate variables; in so doing, we have lost track of the complexities and realities of motivation and work. The next step is good experimentation conforming to reality rather than the laboratory.

Motives and incentives when understood in an individual, in his group, and in the reality of the work situation can lead to describing behavior and predicting the success of short-and-long-term goals. When management clearly states its goals, deals openly and honestly with its employees, and tries to ascertain their goals, then

employees and employers can work together. Each will give a little, but in return each will be likely to gain more.

This chapter may be somewhat disappointing to those who believe that the psychologist should know all about motivation and hence should be able to explain motives and predict behavior. In many respects one of the great differences between the psychologist and the layman is that the psychologist is much more reluctant to make predictions about behavior.

Motivation is usually a complex process. An individual's needs are determined in part by certain physiological imbalances within him, in part by his previous experiences, and in part by the interaction of these two. Industry has tended to oversimplify the worker's motives by attaching too much importance to financial incentives. Many systems of incentive wage payments have been proposed. They overlap in many ways, and all of them are faulty because they exclude the many other motives that operate in man while he is at work, and they refuse to recognize that incentive systems themselves tend to cause fear of rate cutting or dismissal in his mind. The same starting wage reflects different attitudes in relation to the previous salary. This indicates that wages and their meaning are relative rather than absolute.

A study that attempted to discover the relative value of five common incentives on a job found differences based on sex, age, and marital status, but never found that salary was the most important of the incentives. Other studies similarly find that salary is not always regarded by employees as the most important factor on a job.

Recent research indicates the value of cooperative incentives. The common goal of a united group can be a more powerful incentive than that of a number of people competing with one another for individual goals which, in the final analysis, they believe to be of little worth. People work because of pride in successful accomplishment, mastery of the job, and other equally potent social pressures. In many instances work is not as dissimilar to play as is ordinarily presumed. Of value in understanding motives as they operate on the job are biographical sketches and case histories, for they make possible a keener insight into the motives of people in industry. Such studies are more valuable than the assumptions of the industrial engineer or "expert" who believes that people work only for the money.

Possibly the best means of understanding worker motivation is to consider the social meaning of work. In this respect, short-term goals and long-term goals of employees and employers may affect production variously. Accordingly, giving attention to the manner in which incentives are perceived is preferable to assuming that an incentive means the same thing to all.

REFERENCES

Blum, M., and J. Russ. A study of employee attitudes towards various incentives. *Personnel*, 1942, *19*, 438–444.

Cole, R. J. A survey of employee attitudes. *Public Opinion Quarterly*, 1940, 4, 497–506.

Common, J. (ed.). *Seven shifts*. Dutton, New York, 1938.

Cowles, J. T. Food tokens as incentives for learning by chimpanzees. *Comparative Psychology Monographs*, 1937, *14*, 1–96.

Dollard, J., W. Doob, N. E. Miller, O. H. Mower, and R. R. Sears. *Frustration and aggression.* Yale University Press, New Haven, Conn., 1939.

Dudycha, A., and J. C. Naylor. The effect of variations in the cue R matrix upon the obtained policy equations of judges. *Educational and Psychological Measurement,* 1966, 26, 583–604.

Eaton, W. H. Hypotheses related to worker frustration. *Journal of Social Psychology,* 1952, 35, 59–68.

Jones, L. V., and T. E. Jeffrey. A quantitative analysis of expressed preferences for compensation plans. *Journal of Applied Psychology,* 1964, 48 (4), 201–210.

Jurgensen, C. E. Selected factors which influence job preferences. *Journal of Applied Psychology,* 1947, 31, 553–563.

Lawler, E. E. The mythology of management compensation. Mimeographed. Yale University, New Haven, Conn., 1965.

Lytle, C. W. *Wage incentive methods,* rev. ed Ronald, New York, 1938.

Maslow, A. H. A theory of human motivation. *Psychological Review,* 1943, 50, 370–396.

Morse, N. C., R. Weiss, and R. Griggs. Attitudes toward work. Unpublished study, Institute for Social Research, University of Michigan, Ann Arbor, 1954.

Nealey, S. M. Determining worker preferences among employee benefit programs. *Journal of Applied Psychology,* 1964, 48 (1), 7–12.

Opsahl, R. L., and M. D. Dunnette. The role of financial compensation in industrial motivation. *Psychological Bulletin,* 1966, 66, 94–118.

Porter, L. W A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology,* 1961, 45, 1–10.

Porter, L. W. Job attitudes in management I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology,* 1962, 46, 375–384.

Porter, L. W. Job attitudes in management. II. Perceived importance of needs as a function of job level. *Journal of Applied Psychology,* 1963(a), 47, 141–148.

Porter, L. W. Job attitudes in management: III. Perceived deficiencies in need fulfillment as a function of line versus staff type of job *Journal of Applied Psychology,* 1963(b), 47 (4), 267–275.

Porter, L. W. Job attitudes in management. IV. Perceived deficiencies in need fulfillment as a function of size of company. *Journal of Applied Psychology,* 1963(c), 47 (6), 386–397.

Ruttenberg, H. Self-expression and labor unions. In National Research Council. *Fatigue of workers.* Reinhold, New York, 1941

Scott, W. E. Activation theory and task design. *Organizational Behavior and Human Performance,* 1966, 1, 3–30.

Stagner, R. Psychological aspects of industrial conflict II. Motivation. *Personnel Psychology,* 1950, 3, 1–16.

Stockford, L. O., and K. Kunze. Psychology and the pay check. *Personnel,* 1950, 27, 129–143.

Survey Research Center, University of Michigan. Selected findings from a study of clerical workers in the Prudential Insurance Company of America. *Survey Research Center Study #6,* University of Michigan, Ann Arbor, 1948.

Vroom, V. H. *Work and motivation.* Wiley, New York, 1964.

Wolfe, J. B. Effectiveness of token rewards for chimpanzees. *Comparative Psychology Monographs,* 1936, 12, 1–72.

12 JOB SATISFACTION

Job satisfaction is the result of various attitudes possessed by an employee. In a narrow sense, these attitudes are related to the job and are concerned with such specific factors as wages, supervision, steadiness of employment, conditions of work, advancement opportunities, recognition of ability, fair evaluation of work, social relations on the job, prompt settlement of grievances, fair treatment by employer, and other similar items.

However, a more comprehensive approach requires that many additional factors be included before a complete understanding of job satisfaction can be obtained. Such factors as the employee's age, health, temperament, desires, and level of aspiration should be considered. Further, his family relationships, social status, recreational outlets, activity in organizations—labor, political, or purely social—contribute ultimately to job satisfaction.

In short, job satisfaction is a general attitude which is the result of many specific attitudes in three areas, namely, specific job factors, individual characteristics, and group relationships outside the job.

CONFUSION OF TERMINOLOGY

The methods employed in measuring attitudes were presented in a preceding chapter. We emphasized there that the studies reported illustrated the various methods and that the findings were of secondary importance. The main reason for this is the great confusion in the use of the term "job satisfaction" and in the factors that contribute to it. Reviewing the many studies in the area almost leads to the conclusion that job satisfaction is anything that an author measures when he thinks that he is measuring "job satisfaction." Too few experimenters in this field have been concerned with either the reliability or the validity of their measures.

Some studies have dealt only with attitudes on some specific job factors but have been called job satisfaction studies. Others have investigated other job factors and have also been called job satisfaction research. Still others have considered the individual factors and the job factors, and a few have attempted to measure parts of all three areas. It is no wonder that conflicting results are found in the literature on this subject. No one study is necessarily more in error than another, but most are incomplete. For example, an organization that has a good personnel program and offers steady work, good supervision, etc., may nevertheless find that a particular worker has little job satisfaction because of his capabilities in connection with his particular job, family problems, or unhappiness about his failure in union activities.

One additional point must be made about the confusion among the terms "employee attitude," "job satisfaction," and "industrial morale." Although in many instances they are used interchangeably, they are not synonymous. An "attitude" is not "job satisfaction," although it may contribute to job satisfaction since the latter is comprised of a number of attitudes. Similarly, job satisfaction is not the same as industrial morale, although it may contribute to morale.

An attitude of an employee can be considered as a readiness to act in one way rather than another in connection with specific factors related to a job.

Job satisfaction is the result of various attitudes the employee holds toward his job, toward related factors, and toward life in general. Industrial morale is a by-product of a group and is generated by the group. It has four determinants: feeling of group solidarity; need for a goal; observable progress toward the goal; and individual participation in meaningful tasks necessary to achieving the goal. Industrial morale may be defined as the possession of a feeling, on the part of the employee, of being accepted by and belonging to a group of employees through adherence to common goals and confidence in the desirability of these goals.

Authors and experimenters have measured employees' reactions and have sometimes labeled them job satisfaction, sometimes morale, and at other times merely attitudes. There were similar practices years ago in the field of psychological testing. Thus an author of a test named it on the basis of what he believed it was measuring; "Technical Information" and "Teaching Aptitude" are examples. At the present time no one would dare propose a psychological test—and hope to keep his professional reputation—without establishing the validity of the test to prove that it measured what it was supposed to. Unfortunately, determining the "validity" of an attitude scale is a more complex matter.

REASONS FOR JOB SATISFACTION RESEARCH

Why seek information about job satisfaction? For an industrial organization the consequences are extremely important. By discovering attitudes on factors related to the job, a firm can correct certain bad situations and thereby improve the job satisfaction of its staff. From this point of view it would be justified in being concerned only with this area and neglecting the individual "ego" and its employees' adjustment to groups outside the plant. However, an industrial organization can benefit materially if it knows what individual attitudes contribute to job satisfaction. For one thing, applying this knowledge will result in better selection procedures. This is a broad implication as far as job satisfaction is concerned, and even though most industrial and business organizations feel that it is not their problem, it nevertheless exists. It is also directly related to vocational guidance, schools and colleges, public and private employment agencies, etc. Society as a whole must face it realistically.

FACTORS MEASURED IN JOB SATISFACTION

Because of the ramifications of the problems connected with job satisfaction, surveys and studies have been carried out on a community-wide basis, in specific professional or occupational groups and, of course, in specific industrial establishments.

TABLE 12.1 *Responses on Hoppock's Questionnaire*

1. Choose ONE of the following statements which best tells how well you like your job. Please place a check mark in front of that statement:

Response	*Frequency*	*Percentage*
I hate it	5	2
I dislike it	6	2
I don't like it	34	11
I am indifferent to it	29	9
I like it	194	63
I am enthusiastic about it	27	9
I love it	14	5
Total	309	101

2. Check one of the following to show HOW MUCH OF THE TIME you feel satisfied with your job:

Response	*Frequency*	*Percentage*
All of the time	128	41
Most of the time	83	27
A good deal of the time	24	8
About half of the time	29	9
Occasionally	16	5
Seldom	14	5
Never	15	5
Total	309	100

3. Check ONE of the following which best tells how you feel about changing your job:

Response	*Frequency*	*Percentage*
I would quit this job at once if I could get anything else to do	39	13
I would take almost any other job in which I could earn as much as I am earning now	13	4
I would like to change both my job and my occupation	20	7
I would like to exchange my present job for another job in the same line of work	12	4
I am not eager to change my job, but I would do so if I could get a better job	130	43
I cannot think of any jobs for which I would exchange mine	54	18
I would not exchange my job for any other	37	12
Total	305	101

4. If you could have your choice of all the jobs in the world, which would you choose? (Check one):

Response	*Frequency*	*Percentage*
Your present job	145	48
Another job in the same occupation	48	16
A job in another occupation	107	36
Total	300	100

SOURCE: R. Hoppock. *Job satisfaction.* Harper & Row, New York, 1935, 250–252. Copyright, 1935, by National Occupation Conference; renewed 1963 by Robert Hoppock. Reprinted by permission of Harper & Row, Publishers.

5. Check one of the following to show how you think you compare with other people:

Response	*Frequency*	*Percentage*
No one likes his job better than I like mine	59	19
I like my job much better than most people like theirs	50	16
I like my job better than most people like theirs	35	11
I like my job about as well as most people like theirs	114	37
I dislike my job more than most people dislike theirs	25	8
I dislike my job much more than most people dislike theirs	5	2
No one dislikes his job more than I dislike mine	18	6
Total	306	99

6. Which gives you more satisfaction? (Check one)

Response	*Frequency*	*Percentage*
Your job	200	66
The things you do in your spare time	102	34
Total	302	100

7. Have you ever thought seriously about changing your present job?

Response	*Frequency*	*Percentage*
Yes	122	39
No	159	51
Omitted	28	9
Total	309	99

8. Have you ever declined an opportunity to change your present job?

Response	*Frequency*	*Percentage*
Yes	81	26
No	196	63
Omitted	32	10
Total	309	99

9. Are your feelings today a true sample of the way you usually feel about your job?

Response	*Frequency*	*Percentage*
Yes	267	86
No	19	6
Omitted	23	7
Total	309	99

HOPPOCK STUDY

One of the early community-wide surveys was conducted by Hoppock in the town of New Hope, Pennsylvania (1935). Eighty-eight percent of the 351 employed adults answered the lengthy questionnaire. Table 12.1 presents the results.

These results indicate that 15 percent of the sample had negative attitudes, or job dissatisfaction. This early finding is quite similar to the preponderance of evidence that has since been reported. Robinson and Hoppock have collated the data on 191 assorted studies reporting percentages of job dissatisfaction (1952). The median figure is 18 percent dissatisfied.

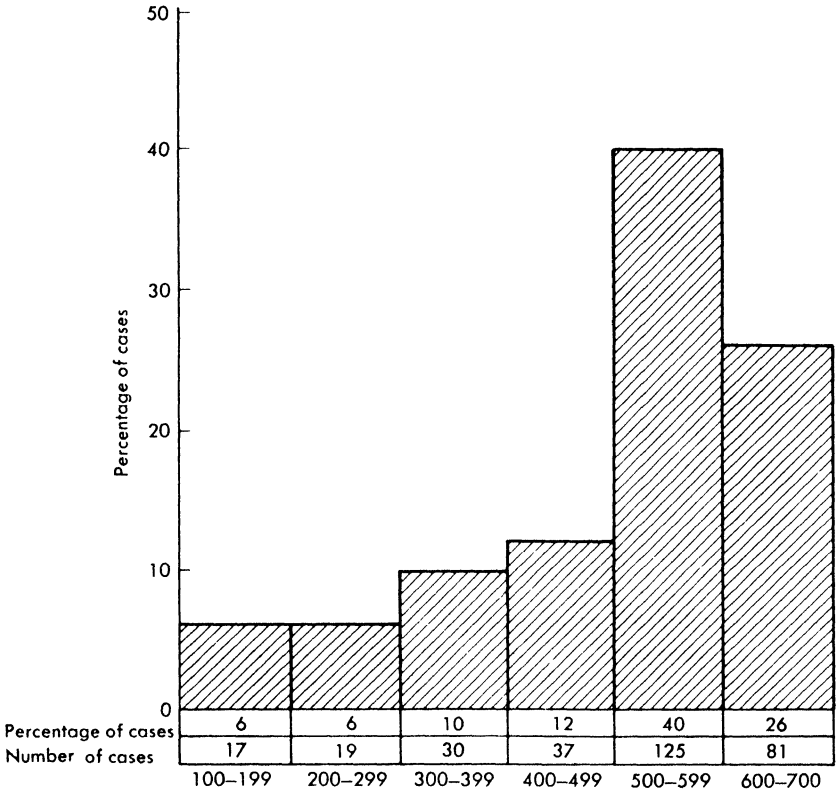


Figure 12 1. Distribution of job satisfaction indices, 309 employed adults. An index of 100 indicates extreme dissatifaction; 700 indicates extreme satisfaction; 400 indicates indifference. (From R. Hoppock. *Job satisfaction*. Harper & Row, New York, 1935, 253. Copyright 1935 by National Occupation Conference, renewed 1963 by Robert Hoppock. Reprinted by permission of Harper & Row, Publishers.)

These figures are quite different from the "common sense" view in which workers are considered as robots doomed to dissatisfaction in their work because of industrial mechanization and other self-attributed causes.

Hoppock asked 36 nationally prominent personnel officers to estimate percentages of workers who were dissatisfied. The answers ranged from 0 to 80 percent, the average being 49 percent. As Hoppock points out, if numbers from 0 to 100 were

put in a hat, the average of the numbers drawn would be 50. Hence it may well be that the 49 percent is as void of meaning as chance itself.

The view that is clearly taken is that the majority of the gainfully employed tend to have job satisfaction or are at least neutral. Only a small percentage have job dissatisfaction.

An index of job satisfaction was computed; the results are shown in Figure 12.1. A breakdown according to occupational classification indicates that the average index is lowest for the unskilled category and highest for the professional category; this is shown in Table 12.2.

This survey indicated that there is considerably more job satisfaction than dissatisfaction when all the persons who are gainfully employed are included in the survey.

Hoppock also conducted a survey on people in one occupation, namely, teaching. Five hundred teachers from 51 urban and rural communities in the northeastern United States estimated their job satisfaction on four attitude scales. By combining these scales, a measure of job satisfaction was obtained. Of this group the 100 most satisfied and 100 least satisfied were asked about 200 questions. A comparison of their answers differentiated the satisfied from the dissatisfied teachers in the following areas:

1. The satisfied showed fewer indications of emotional maladjustment.
2. The satisfied were more religious.
3. The satisfied enjoyed better human relationships with superiors and associates.
4. The satisfied were teaching in cities of over 10,000 population.
5. The satisfied felt more successful
6. Family influence and social status were more favorable among the satisfied.
7. The satisfied "selected" their vocations.
8. Monotony and fatigue were reported more frequently by the dissatisfied.
9. The satisfied averaged 7.5 years older

One interesting finding is that the difference in average salaries between the two groups was not statistically significant.

Table 12.2 *Job Satisfaction Indices of Five Occupational Groups—New Hope, Pennsylvania, 1933*

Occupational Classification	*Number of Cases*	*Range of Indices*	*Mean Index*
1. Unskilled manual	55	100–650	401
2. Semiskilled	74	125–650	483
3. Skilled manual and white-collar	84	125–675	510
4. Subprofessional, business, and minor supervisory	32	250–700	548
5. Professional, managerial, and executive	23	300–700	560

Source: R. Hoppock. *Job satisfaction.* Harper & Row, New York, 1935, 255. Copyright 1935 by National Occupation Conference; renewed 1963 by Robert Hoppock. Reprinted by permission of Harper & Row, Publishers.

One brief comment is necessary about this survey. No attempt was made to measure the proficiency of the teachers, and therefore it is not known whether those who were dissatisfied were less "good" than the satisfied group. This survey also brings out the point that job satisfaction and vocational interest are not identical.

For example, 84 percent of the dissatisfied teachers answered "yes" to the question "Is your work interesting?"

In the epilogue to his study on job satisfaction, Hoppock proposes the following six major components of job satisfaction:

1. The way the individual reacts to unpleasant situations
2. The facility with which he adjusts himself to other persons
3. His relative status in the social and economic group with which he identifies himself
4. The nature of the work in relation to the abilities, interests, and preparation of the worker
5. Security
6. Loyalty

These six items are not of the minute and specific character measured in many studies on job satisfaction: possibly that is what is wrong with these studies. Hoppock's approach is to be commended for this reason: He is aware of the real factors which contribute to job satisfaction and does not get lost in the petty details. Job satisfaction is an important generalized attitude in an individual, not a specific attitude about specific job factors. The idea that it is related to the individual's emotional adjustment suggests that those who are unstable emotionally may have considerably more difficulty adjusting themselves to a job and may therefore be dissatisfied with it.

FACTOR ANALYSIS STUDIES

One of the more elegant ways of assessing the basic underlying factors in job satisfaction is through the statistical technique of factor analysis. Vroom (1964), in surveying all such studies, lists the different dimensions which have been found in these different studies, as shown below:

Attitude Dimension	*Study*
1. Attitudes toward the company and company management	Wherry, 1954; Ash, 1954; Dabas, 1958; Roach, 1958; Twery, Schmid, and Wrigley, 1958; Kahn, 1960; Harrison, 1961
2. Attitudes toward promotional opportunities	Harrison, 1961; Kendall, Smith, Hulin, and Locke, 1963
3. Attitudes toward job content	Baehr, 1954; Ash, 1954; Roach, 1958; Kendall, Smith, Hulin, and Locke, 1963
4. Attitudes toward supervision	Baehr, 1954; Ash, 1954; Dabas, 1958; Roach, 1958; Twery, Schmid, and Wrigley, 1958; Kahn, 1960; Harrison, 1961; Kendall, Smith, Hulin, and Locke, 1963
5. Attitudes toward financial rewards	Wherry, 1954; Ash, 1954; Dabas, 1958; Roach, 1958; Kahn, 1960; Harrison, 1961; Kendall, Smith, Hulin, and Locke, 1963
6. Attitudes toward working conditions	Wherry, 1954; Dabas, 1958; Harrison, 1961
7. Attitudes toward coworkers	Roach, 1958; Twery, Schmid, and Wrigley, 1958; Kendall, Smith, Hulin, and Locke, 1963

RELATIVE IMPORTANCE OF DIFFERENT ASPECTS OF JOB SATISFACTION

Given that there are a number of different aspects which can contribute to the overall satisfaction or dissatisfaction of employees, the next question would seem to be that of their relative importance.

Herzberg, Mausner, Peterson, and Capwell (1957) report data compiled from 16 different studies and involving over 11,000 employees which gives an indication of how workers rank different factors in terms of their importance. Figure 12.2 presents

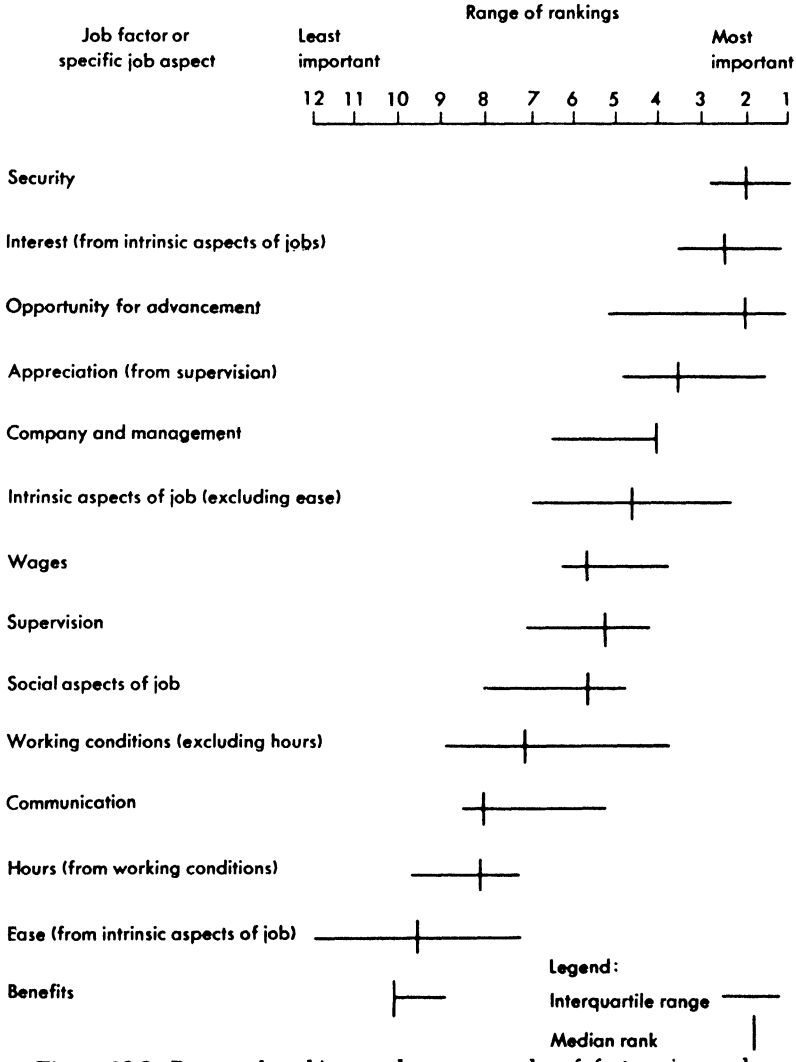


Figure 12.2. Range of rankings and average ranks of factors in employee attitudes (compiled from sixteen studies, including over 11,000 employees). (From F. Herzberg *et al. Job attitudes: Review of research and opinion.* Psychological Services of Pittsburgh, Pittsburgh, Pa., 1957.)

data showing the median rank for each factor plus a measure of the variability in ranks for that factor observed across all the studies. (This measure of variability is the "interquartile range.")

The figure shows quite dramatically that security is the most important single job variable. (Note also how consistently it was rated high.) Opportunity for advancement rates a close second. Factors judged as being of least importance were job benefits and ease of work.

While these data may be used as an approximate indication of the *overall* importance of various job factors, it is very important to keep in mind that this ranking is apt to be quite different for any particular class or group of workers. Herzberg *et al.* point this out quite clearly. For example, with people at higher occupational and/or educational levels, intrinsic aspects of the job go up in importance, while security drops off considerably. This is no doubt due to the greater "marketability" possessed by people in the higher occupational categories.

An excellent illustration of how job satisfaction is influenced by group variables is presented by Hulin and Smith (1964). They were interested in exploring whether men would differ significantly from women (in the same plant) in terms of their job satisfaction. They obtained measures of five separate aspects of job satisfaction from 295 male workers and 163 female workers drawn from four different plants. The data were analyzed with respect to the mean job satisfaction for the male and female workers. Analyses indicated that in three plants the female workers were significantly less satisfied than their male counterparts ($p < 0.05$) while in the fourth plant there was no significant difference. A test on the relative size of the differences indicated that the ordering of the differences in satisfaction level was somewhat consistent across the four samples ($p < 0.01$). The data can be seen even more clearly by examining Table 12.3 (a positive value indicates greater satisfaction on the part of men).

TABLE 12.3 *Mean Differences*

	Area of Job Satisfaction					
Company	*Work*	*Pay*	*Promotions*	*Supervision*	*People*	*p*
I, Plant A	0.88	1.28	8.64	5.08	5.10	<0.05
I, Plant B	−1.88	−3.28	2.92	0.14	−1.18	*ns*
II	2.80	−0 48	4.91	1.10	6.37	<0.05
III	0.00	−2.00	6.72	2.37	1.86	<0.05

SOURCE: C. L. Hulin and P. C. Smith. Sex differences in job satisfaction. *Journal of Applied Psychology*, April 1964, *48* (2), 88–92.

JOB SATISFACTION AND JOB BEHAVIOR

So far we have learned something about the dimensions of job satisfaction, their relative importance, and how that importance varies as a function of the particular reference group. As yet we have not considered the degree to which job satisfaction is related to other aspects of work behavior. Is there any relationship between how successfully a worker performs and the degree to which he is satisfied with the

various aspects of his job? As we shall soon see, the answer to that question is not yet clearly established.

Vroom (1964) has done an excellent job of examining the relationship between job satisfaction and various aspects of job behavior, and perhaps summarizing his findings is the best way of giving the reader an overview. Vroom categorizes studies in terms of which job behaviors are correlated with job satisfaction. Specifically, he groups them into studies of turnover, absenteeism, accidents, and job performance.

SATISFACTION AND TURNOVER

Of the seven studies examined by Vroom which related job satisfaction to turnover, all indicated a negative relationship. That is, the higher a worker's satisfaction, the less apt he was to leave the job.

SATISFACTION AND ABSENTEEISM

The results of ten studies in this grouping were somewhat equivocal. Four studies tended to support the notion of a negative relationship between the amount of job satisfaction and the degree of work absenteeism. However, three studies did not support this premise, and three others indicated that the magnitude of an absenteeism-satisfaction correlation can be demonstrated to be a function of such other variables as the type of absenteeism measure used and the sex of the worker.

SATISFACTION AND ACCIDENTS

Very few data are available which relate accident data to job satisfaction. Vroom reports only two studies, one of which found substantial negative relationships and another which found no relationships at all. Accidents as a criterion measure usually leave much to be desired since there is considerable evidence that most accidents are caused simply by chance factors.

SATISFACTION AND JOB PERFORMANCE

While each of the above categories was concerned with job-related behavior and job satisfaction, none of them is directly concerned with actual job performance, that is, how well the worker actually accomplishes his assigned task. For example, a worker may be absent a great deal but still do his job quite well. Does job satisfaction have any relationship to actual job performance? The available evidence seems to indicate that no such relationship exists. This was first brought dramatically into focus by Brayfield and Crockett (1955) who examined all research relating job satisfaction to job performance up to that time and concluded that there was virtually no evidence of any relationship between these two variables.

This, of course, is a rather critical finding for those who support the general "human relations" notion that a satisfied worker is a more productive worker.

Vroom, in his 1964 book, brings the research in this category up-to-date from the earlier Brayfield and Crockett paper.[1] Table 12.4 is a table which Vroom prepared to illustrate the general pattern of findings between job performance and job satisfaction. He found a median correlation of 0.14, with a range of 0.86 to −0.31. There seems little doubt that there is at best only a small relationship existing between these variables.

[1] Another rather extensive review on the topic can be found in Herzberg, Mausner, Peterson, and Capwell (1957).

TABLE 12.4 *Correlational Studies—Job Satisfaction and Job Performance*

Author	*Type of Analysis*	*Population*	*Correlation*	*Type of Criterion of Productivity*	N
Baxter cited in Brayfield & Crockett (1955)	Ind.	Insurance agents	0.33 0.26	Ratings Objective	233
Bellows cited in Brayfield & Crockett (1955)	Ind.	Air force control tower operators	0.005	Ratings	109
Bernberg (1952)	Ind.	Hourly paid workers	0.05	Ratings	890
Brayfield cited in Brayfield & Crockett (1955)	Ind.	Female office employees	0.14	Ratings	231
Brayfield & Mangelsdorf cited in Brayfield & Crockett (1955)	Ind.	Plumbers' apprentices	0.203	Ratings	55
Brayfield & Marsh cited in Brayfield & Crockett (1955)	Ind.	Farmers	0.115	Ratings	50
Brody (1945)	Ind.	Production employees on piece work	0.68	Objective	40
Fleishman, Harris & Burtt (1955)	Gr.	Work groups in an equipment manufacturing plant	−0.31	Ratings[a]	58
Gadel & Kriedt	Ind.	IBM operators	0.08	Ratings	193
Giese & Ruter (1949)	Gr.	Departments in mail-order company	0.19	Objective	25
Hamid (1953)	Ind.	Insurance agents	0.22	Objective	552
Heron (1954)	Ind.	Bus drivers	0.308	Objective[b]	144
Lawshe & Nagle (1953)	Gr.	Departments in an office	0.86[c]	Ratings	14
Lopez (1962)	Ind.	Administrative-technical personnel	0.12	Ratings	124
Mann, Indik, & Vroom (1963)	Gr.	Truck drivers—large work groups	0.14 −0.21	Ratings Objective	28
Mann, Indik, & Vroom (1963)	Gr.	Positioners—small work groups	0.18 0.02	Ratings Objective	24
Mossin (1949)	Ind.	Female sales clerks	−0.03	Ratings	94
Sirota (1958)	Ind.	Employees in an electronics firm	0.11	Ratings	377
Sirota (1958)	Ind.	Supervisors in an electronics firm	0.13	Ratings	145
Vroom (1960a)[d]	Ind.	Supervisors in a package delivery company	0.21	Ratings	96

[a] Ratings of foremen's proficiency.

[b] A composite criterion based on five objective measures and one supervisory rating. Correlation reported is between job satisfaction and value to the employer with age partialled out.

[c] Correlation reported is between attitude toward the supervisor and productivity.

[d] These data were not reported in the original publication.

Source: V. H. Vroom, *Work and motivation*, Wiley, New York, 1964.

SUMMARY OF JOB SATISFACTION RELATIONSHIPS

In summarizing the research relating job satisfaction to job behavior variables, Vroom (1964, p. 186) draws a number of conclusions which are worthy of repeating here.

1. There is a consistent negative relationship between job satisfaction and the probability of resignation. This relationship appears when scores on job satisfaction are obtained from individuals and used to predict subsequent voluntary dropouts and when mean scores on job satisfaction for organizational units are correlated with turnover rates for these units.
2. There is a less consistent negative relationship between job satisfaction and absences. This relationship appears to emerge most consistently with measures of unexcused absences, and when frequency of absence rather than actual days lost are used.
3. There is some indication of a negative relationship between job satisfaction and accidents. However, the number of existing studies of this relationship is too small to permit any firm conclusions.
4. There is no simple relationship between job satisfaction and job performance. Correlations between these variables vary within an extremely large range and the median correlation of 0.14 has little theoretical or practical importance We do not yet know the conditions which affect the magnitude and direction of relationships between satisfaction and performance. Obtained correlations are similar for analyses based on individuals and groups and do not seem to depend to any appreciable extent on the occupational level of the subjects or on the nature of the criterion (objective or ratings) employed.

THEORIES OF JOB SATISFACTION

Several different theories exist concerning the dynamics of job satisfaction and its general impact upon worker behavior. A brief mention of the more prominent of these seems highly appropriate.

MASLOW'S THEORY

The need hierarchy notion of Maslow which was outlined in detail in Chapter 11 is also relevant to the present topic. Jobs which are able to satisfy more of the Maslow needs would be jobs which would result in greater satisfaction on the part of the employee.

VROOM'S THEORY

Like the Maslow model, Vroom's Valence-Force theory discussed in Chapter 11 also is relevant here. In Vroom's model, job satisfaction reflects valence of the job for its incumbent. From proposition 2 it would therefore follow that the strength of the force on a worker to remain on his job is an increasing function of the valence of his job. Thus satisfaction should be negatively related to turnover and absenteeism, which it seems to be. Whether or not this valence should also lead to greater production while on the job is less clear, however—a point which again seems to be upheld by the highly ambiguous existing data.

STOGDILL'S THEORY

As we have pointed out several times, one of the major sources of frustration in job satisfaction research is that there does not seem to be any clearly defined rela-

ance. Stogdill (1959), extremely concerned with this, decided that it was time to stop trying to view satisfaction as a "causer" of job performance (that is, as an input variable). Instead, he felt it much more appropriate to view the individual in terms of the context of the total organization.

In an extremely well-documented and well-organized book, he proceeded logically to integrate prior research findings. His conclusions were that the "output" of organizations are group integration, production, and morale. Further, satisfaction of individual expectations results in group integration and cohesiveness but is *not* necessarily related to production. Instead, both morale and production are a function of group structure. Therefore morale and production will only be related to satisfaction when the conditions which lead to high morale and production are also those which lead to the reinforcement of worker expectations.

The general notion of job satisfaction being an output or dependent variable has also been suggested by other workers in this area, most notably in the work of Katzell, Barrett, and Parker (1961) and of Locke, Smith, Kendall, Hulin, and Miller (1964).

HERZBERG'S THEORY (A CURRENT CONTROVERSY)

No theory of job satisfaction has received as much attention or has been subjected to as much criticism as has the model proposed by Herzberg, Mausner, and Snyderman (1959). Because of its increasing popularity among personnel managers and because of its somewhat uncertain status as an accurate model for job satisfaction, we shall examine the Herzberg Model in somewhat more detail than was done with the preceding models.

BASIC STUDY The original study of Herzberg, Mausner, and Snyderman was concerned with an investigation into the causes of job satisfaction and dissatisfaction of engineers and accountants. Their method was quite simple. They interviewed each person individually. Each worker was asked to describe, in detail, times when he felt exceptionally good or exceptionally bad about his job. The interviews were then "content analyzed" to see (1) what kind of things were mentioned when people described the times they were very satisfied with their job, (2) what kinds of things were mentioned when people were describing times they were very dissatisfied with their job, and (3) if the kinds of things described in these two different circumstances would be different.

The results of their study did indicate that things which were associated with high satisfaction ("satisfiers") were somewhat different from the things which were associated with situations of low satisfaction ("dissatisfiers"). They found that the descriptions of good periods included such things as achievement, recognition, advancement, responsibility, etc. All of these things seemed to relate to the actual content of the job—they were therefore called *content factors*. Descriptions of bad work periods seemed to be filled with items dealing with company policy, supervision, salary, and working conditions. These items seemed to relate to the context in which a person performed his task and were therefore referred to as *context* factors. Figure 12.3 shows the relative frequencies that these various kinds of factors were mentioned.

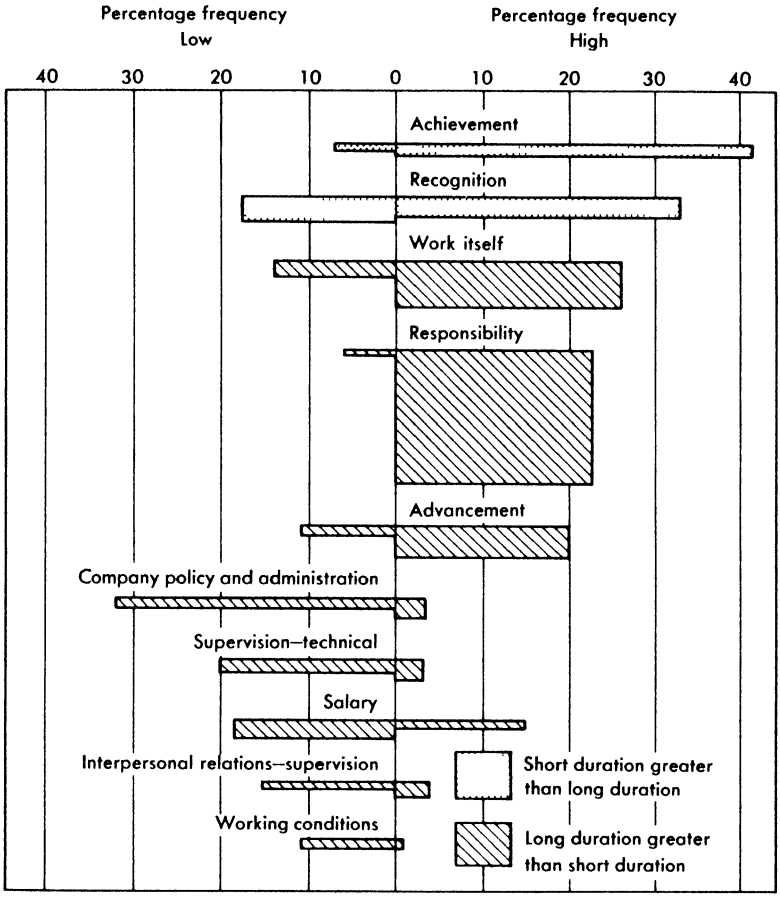


Figure 12.3. Comparison of satisfiers and dissatisfiers. (From F. Herzberg *et al. The motivation to work.* Wiley, New York, 1959, 481.)

THE MODEL Given the findings described above, Herzberg, Mausner, and Snyderman postulated two general classes of work variables—*satisfiers* and *dissatisfiers.* Satisfiers are those things which lead to job satisfaction, they are generally job content factors or "motivators." Dissatisfiers are those things which result in job dissatisfaction; they are generally job context or "hygiene" factors. To quote the authors (1959, p. 114):

> Man tends to actualize himself in every area of his life, and his job is one of the most important areas. The conditions that surround the doing of the job cannot give him this basic satisfaction; they do not have this potentiality. It is only from the performance of a task that the individual can get the rewards that will reinforce his aspirations. It is clear that although the factors relating to the doing of the job and the factors defining the job context serve as goals for the employee, the nature of the motivating qualities of the two kinds of factors are essentially different. Factors in the job context meet the needs of the individual for avoiding unpleasant situations. In contrast to this motivation by meeting avoidance needs, the job factors reward the needs of the individual to reach his aspirations. These effects on the individual can be conceptualized as actuating approach rather

than avoidance behavior. Since it is in the approach sense that the term motivation is most commonly used, we designate the job factors as the "motivators," as opposed to the extra-job factors, which we have labelled the factors of hygiene. It should be understood that both kinds of factors meet the needs of the employee; but it is primarily the "motivators" that serve to bring about the kind of job satisfaction and, as we saw in the section dealing with the effects of job attitudes, the kind of improvement in performance that industry is seeking from its work force.

Because the model postulates these two general classes of variables—one class which can satisfy but not dissatisfy, and one which causes dissatisfaction but not satisfaction—the model has been designated as Herzberg's Two-Factor Theory.

MAJOR CRITICISMS OF THE MODEL In the process of the study the authors had to make the implicit assumption that the people being interviewed had both the ability and the desire to report *accurately* on the conditions which make them satisfied or dissatisfied with their jobs. If this is not so, then their results may be only an artifact of the "set" people carried into the interview situation. For this reason, the study has been severely criticized. The major objection is that when a person is asked to tell about something good, he is apt to attribute the causes of these to his own accomplishments and achievements (content items); we all like to satisfy our need for esteem. Similarly, when someone is asked to tell about an unpleasant or dissatisfying work experience, he is more apt to blame others for this (context items) than he is to blame himself. Many attribute Herzberg's findings to this simple tendency on the part of people.

EVIDENCE BEARING ON THE MODEL Since the original publication of the model, a number of studies have attempted to provide evidence concerning its validity. Unfortunately they have tended to be contradictory in nature. Porter (1966, p. 411) has recently provided a brief but comprehensive summary of all such research.

> First, the evidence in favor of the theory Myers (1964) found that, for a representative sample of employees (both salaried and hourly paid) in a manufacturing company, Herzberg's two distinct types of attitudes emerged. Satisfaction was related to intrinsic work factors and dissatisfaction to extrinsic factors. No information was provided concerning the relations of attitudes to performance. Likewise, Schwartz, Jenusaitis and Stark (1963), in a study of lower-level supervisors, found the appropriate relationship between satisfaction-dissatisfaction and intrinsic-extrinsic factors but they also reported no evidence concerning attitude-performance relationships. Other studies with similar results are those of Friedlander and Walton (1964) and Lodahl (1964). An additional study by Friedlander (1964) was interpreted by the author as providing partial confirmation of the theory, although this interpretation seems open to question.
>
> Evidence negative to Herzberg's theory is provided in studies by Dunnette (1963), Ewen (1964), Friedlander (1963), Kornhauser (1965), and Wernimont (1964). All of these investigations seem to support Dunnette's conclusion that. ". . . the two-factor notion of job satisfaction is an oversimplified representation of the motivational milieu of the world of work." From his reading of the recent literature, this reviewer would have to agree strongly with this statement. Factors involved in feelings of satisfaction and dissatisfaction do not appear to divide as neatly as was the case with Herzberg's original study.

SOME GENERAL COMMENTS ON JOB SATISFACTION

Anyone who has ever held a job knows that if he does not get along with his coworkers the job is unsatisfactory. The most rugged, aggressive, and independent individual will not be satisfied at work if he cannot get along with the working

group. Furthermore, adjustment to people on the outside affects a person's adjustment to his job. A secretary who is not getting along with her boyfriend or a husband who has had an argument with his wife is likely suddenly to find some dissatisfaction with the job which did not exist before; this usually disappears just as suddenly, after the "kiss-and-make-up" stage.

In our culture, as in many others, there is a strong desire to be approved and respected by others, especially one's friends. If an individual is to have job satisfaction, he must feel that he is on a par with his friends. College graduates consciously or unconsciously sacrifice money when they enter the professional rather than the business field. They are reluctant to choose selling as an occupation because they feel they are "too good" for it. The fact that they eventually sell professional services does not bother them, whereas selling insurance or hats usually does. The factory worker whose friends also work in a factory can feel he is as good as they are; he will have greater job satisfaction than the factory worker whose friends are mainly office workers.

Job dissatisfaction may well be the result of a lack of vocational guidance. A person who is "too good" or "not good enough" for a job, in terms of his abilities and interests, is not likely to be satisfied with his job. During the depression years college graduates were available at a "dime a dozen." Department stores discovered this and proceeded to hire them at almost this rate. But they soon had to discontinue this policy, because the young men and women were "too good" for the job and left in droves, despite the fact that there were few positions open. People are not capable of working for any length of time at a job which they feel is below them. This is true also of people who do not possess the necessary abilities. A college professor who cannot keep up with his students is likely to be a "sad sack." He may come to the conclusion that his students ask questions only to prove how smart they are. Certain police departments face the peculiar problem of having officers who are less capable than the patrolmen. Friction results. The same phenomenon appears in all types of business organizations.

Security contributes to job satisfaction, but we must remember that security is social as well as economic. More important, security is relative, not absolute. Resignations occur in the civil service even though the positions carry tenure; furthermore, some people holding these jobs report greater insecurity than some industrial employees who have no guarantee of continued employment. Of course, it may be that insecure people seek civil service jobs and that their basic insecurities prevent their feeling any great job security. Adventurous and dangerous jobs are sometimes considered "secure" even when life itself may be in the balance. A physician is secure in his position and cannot be prevented from continuing in it (barring a few legal exceptions), but he has no guarantee that people will call him. The point is that security is a relative, not an absolute, concept.

Knutson (1952) makes the interesting point that what some writers refer to as security or insecurity is defined by others as adjustment or maladjustment, high or low morale, and the like. He states:

> Personal security is a "dynamic" concept. People seek security within their various areas of striving: by achieving it, they become secure in some areas; by failing to achieve it, they become insecure in other areas. The areas of personal striving are interrelated and overlapping, and some may be of greater importance than others to the security of the individual. Feelings of security or insecurity may tend to become generalized throughout the personality structure, or they may be projected from one area of striving into others. Furthermore, the biological structure or physical or social environment of an individual,

by limiting or influencing the character and direction of his motivations, may have a direct bearing upon the nature of his securities. In view of these conditions or possibilities, there appears to be a general level or status of security, as well as levels of security within the individual areas of striving.

A person's feelings of security or insecurity within any area of striving involve his own subjective evaluation of his success, satisfaction, and surety or confidence with respect to the carrying out of his purposes in past and present situations and group relationships, also, his expectations, hopes, fears, or uncertainties with respect to the carrying out of his purposes and aspirations in future situations and group relationships.

Feelings of security within any functional area vary on a continuum, so that security status within any area of striving or within any situation may be considered a matter of degree rather than a matter of absolute.

Knutson found that the patterns of security of very similar occupational groups are more alike than those of more distantly related occupational groups. This finding would lead to the speculation that job satisfaction may not be a unitary trait but that it may exist in some people when they are employed in certain occupations and not in others.

Loyalty is not to be overlooked in job satisfaction. The feeling that the worker has for his firm or boss, or both, contributes to job satisfaction. Although it is not likely to be as strong as the "school spirit" evidenced by loyal sons of Alma Mater, it is present in employees. It is the factor that leads people to sacrifice their own needs and purposes. It is the factor that prompts service beyond the line of duty in relation to job requirements.

To understand job satisfaction better, we must take into consideration the opportunities for satisfaction that the job itself affords the individual as well as the broader opportunities it offers him. The job satisfaction index rises with occupational level. Kornhauser (1940) reports findings which have a bearing on this point. He asked four different economic groups seven questions dealing with personal satisfaction, and in each instance he found that the higher income groups indicated greater personal satisfaction. The questions and answers (given in percentages) are presented in Table 12.5 (modified from his findings).

Each of these questions is directly or indirectly related to job satisfaction. Questions 2 and 4 are the only ones to which the majority of the lowest income group gave favorable answers, and yet they are of the essence of job satisfaction. This study is cited to show that if the economic range is great enough, salary may become a significant factor, but that with any one occupational group (for example, Hoppock's teachers) or within a limited range, it is a minor factor.

Morse (1953) has published a book entitled *Satisfactions in the White-Collar Job.* It is based upon the interviewing of 742 clerical workers in a large insurance company. In addition, 73 first- and second-line supervisors were also interviewed.

As a result of this study, the hypothesis was advanced that satisfaction depends basically upon what an individual wants from the world and what he gets. Morse developed four indices of job satisfaction. The first is related to *content* of the job; the second is related to *identification with the company;* the third is related to *financial and job status satisfaction;* and the fourth is related to *pride and group performance.* While it must be stated that an author is as an author does, the point of view expressed in this text is to refer to job satisfaction as the resultant of the sum total of individual attitudes, and to refer to morale as a group factor. Morse apparently considers job satisfaction and morale as interchangeable concepts. To

TABLE 12.5 *Comparison of Groups in Personal Satisfaction Questions (Percent of Workers Answering Positively)*

	Annual Salary (in 1937)			
	Over $5000	*$2000–$5000*	*$1000–$2000*	*Under $1000*
Do you feel that your children have as much opportunity as they should have?	83%	60%	46%	39%
Do you like the kind of work you do?	95	91	81	67
Do you feel that your pay is fair?	90	69	53	34
Would you say that you are treated well by the people you work for?	96	90	86	77
Do you feel that there is any danger of losing your job?	89	80	72	49
Do you feel that you have as much opportunity to enjoy life as you should have?	82	69	55	36
Do you feel that you have a good chance to get ahead in life and become fairly well off?	96	78	69	43

SOURCE. A. Kornhauser Analysis of class structure of contemporary American society. In G. W. Hartmann and T. M. Newcomb (eds.). *Industrial conflict A psychological interpretation.* Cordon Co., New York, 1940, 242.

be more correct, she considers morale to be a combination of three of her indices of job satisfaction: job content, company involvement, and pay and job status. However, at another point, she describes the degree of group spirit as equal to morale. While it is not important for present purposes to become involved in the problem of semantics, it nevertheless is important for an author to be consistent in definition as well as in application.

One of the underlying generalizations offered by Morse is that level of satisfaction is a combination of both level of aspiration or need-tension level and amount of return from environment. Satisfaction exists when these two are in line, and job dissatisfaction exists when the return from the environment is much less than the need level of the individual. Morse found that productivity and employee satisfaction did not necessarily go hand in hand. This is similar to other findings of the Michigan group on section gangs on a railroad and factory workers in a tractor plant. All three studies find that high producers do not differ significantly from low producers in overall satisfaction with their employment.

Only one who is naive or who makes unguarded and unwarranted assumptions would expect to find job satisfaction and production related in a cause-and-effect relationship. This view is what might be called "unseasonable sweetness"—that is, all good things should be expected to go together. Since job satisfaction is a good thing and so is high production, the two ought to go together. High production, however, may be an employer's goal and job satisfaction may be an intrinsic quality in an employee. Many instances can be cited from everyday experience where people who are very job-satisfied are not necessarily high producers. It may well be best to regard job satisfaction as an entity in itself, which may or may not be related to

production. Just as one might hope to find that beauty and intelligence go together, one finds that sometimes they do not, although sometimes they do.

Schaffer (1953) has proposed the hypothesis that "Overall satisfaction will vary directly with the extent to which those needs of an individual which can be satisfied in a job are actually satisfied; the stronger the need, the more closely will job satisfaction depend on its fulfillment."

Schaffer investigated 12 needs, using as his sample 72 employed men, most of whom were in the professional and semiprofessional occupational groups. He found that two clusters appeared. One contained needs which were passive or hostility-restraining in nature; the other contained the assertive, aggressive needs. A high correlation existed between the person's two strongest needs and overall satisfaction. Among the strongest needs reported in this sample were creativity and challenge, mastery and achievement, and helping others.

Morse's work and Schaffer's work seem to point to the view that job satisfaction depends upon the fulfillment of personal needs. Such fulfillment may not at all depend upon productivity as measured by the number of units turned out, and therefore the expectation that increasing job satisfaction increases productivity may be a "Pollyanna" approach to the reality that does exist.

Weitz (1952) raises an interesting point that a measure of general satisfaction of an individual should be taken in relation to that individual's job satisfaction. He offers a number of hypotheses such as "those who have high general dissatisfaction scores and a large number of job dissatisfactions are less likely to quit than those who have low general dissatisfaction scores and a large number of job dissatisfactions; other combinations are obvious, such as high general dissatisfaction–low job dissatisfaction (less likely to quit), etc."

Granted that we should know with what job satisfaction correlates. However, while we may hypothesize, we should not glibly expect job satisfaction to correlate with arbitrary and assumed external criteria. We may expect job satisfaction to correlate with length of employment and production, but under certain conditions this expectation may be false. Both production and length of employment may be employers' goals and cited as measures of success. However, these two criteria are not correlated to any perfect extent with each other. Blum's finding (1940) that production and length of employment correlated + 0.44 would indicate that each is measuring a somewhat different aspect of success on the job.

Severin (1952) points out that care should be used in substituting one criterion for another since the median correlation of job performance criteria with various other criteria is about + 0.30. It is clear that criteria correlate with one another variously, and some are more equivalent to each other than are others.

In a market of many jobs, the more productive workers may leave to take better jobs and the poorer ones will stay on. Job satisfaction may therefore contribute to the less productive workers remaining, while it may be that the more productive workers, regardless of job satisfaction, take advantage of better job opportunities.

Whether job satisfaction is related to termination may depend to a large extent on employment conditions and the availability of other jobs. Therefore, the correlation between job satisfaction and termination may exist in one type of employment market and not in another.

Studies comparing job satisfaction with such criteria as termination and production may result in erroneous conclusions. It is necessary to be aware that many hypotheses can exist concerning the relation between job satisfaction and produc-

tion and/or termination or any other stated criterion. The hypothesis offered here is that job satisfaction can bear various relations to production depending upon a host of factors not usually even considered by some psychologists in their studies. For example, a correlation is computed as a result of taking a measure of production at a particular time. It is clear from studies investigating production that it varies from one period to another. In other words, production in the long run may differ from production in the short run—and which should be regarded as the "truer" measure? If they differ, then each bears a different relationship to job satisfaction.

It is most important to establish what is meant by job satisfaction or dissatisfaction and *then* to measure it. If what is meant by job dissatisfaction is the number of gripes that a person has about a job, it may not be correct to assume that the person who gripes will necessarily leave the job. It may happen that the griper is competent and feels free to gripe, whereas the person who is unsure and is afraid to complain may leave the job because he is really dissatisfied. This, again, may be cited as an example of "unseasonable sweetness." The fact that a psychologist establishes a correlation between two concepts does not mean that he established a cause and effect. Further, it does not mean that this correlation would continue if any variable in addition to the two being measured were to change.

Research findings often lead to meaningful hypotheses, but they must also be mixed with reality. Many times the researcher will be stumped by the meaning of his results. All he has to do is ask the secretary or the factory worker involved to explain the phenomenon, and the answer will be forthcoming immediately.

Kates (1950) found that there was a positive correlation between scores on the Strong interest blank and job satisfaction in clerical workers, but he did not find that the expression of job satisfaction bore any relation to Rorschach signs of adjustment or maladjustment in these same clerical workers. He also found that the higher the score on the clerical scale of the Strong blank, the more numerous were the signs of personality maladjustment as determined by Rorschach respondents. This would mean that routine clerks may be satisfied with their jobs despite Rorschach's signs of maladjustment. This finding does not allow the generalization that routine clerks who are satisfied on their jobs are emotionally maladjusted. It merely means that the possibility exists that individuals may have satisfaction in one area in relation to their personality structures, and if these people were working in different areas they might have dissatisfactions. Thus, it is unsafe to assume that all "good" things ought to go together, and if they do not, then there is something wrong.

Watson (1939) has conducted a few studies in job satisfaction, the results of which are summarized here to shed additional light on the role of salary as a factor in job satisfaction. He states:

> Without in any way denying the need of many workers for higher wages, it may be recognized that sometimes workers, vaguely dissatisfied with the way things are going, know no other way of meeting the situation. They ask for more pay or shorter hours, and perhaps they should have these, but if these demands were attained, the relief would be only temporary. The underlying cause of dissatisfaction may be not in the pay envelope or the time clock, but in the work itself.

In this study, and in a similar study by Seidman and Watson (1940), a sampling of men and women were asked to report on the job previously held which was most satisfactory to them and to give the reasons for their selection. The results, which

are shown in Table 12.6 present further evidence that recognition, friendly associations, work fitted to vocational level, and variety of duties are more important contributing factors in job satisfaction than salary.

TABLE 12.6 *Reasons Why Workers Preferred One Job Rather Than Another (in Percentages)*

Reasons	Watson Study		Seidman-Watson Study
	Men	Women	Young Men
Congenial work conditions and social contacts	21%	38%	24%
Responsibility, initiative, prestige	27	23	19
In line with vocational aspirations	15	13	29
Variety	15	12	12
Salary	13	6	4
Shorter hours	3	6	4

SOURCE: Adapted from J. M. Seidman and G. Watson. Satisfaction in work. *Journal of Consulting Psychology*, 1940, *4*, 117–120; and K. Lewin and G. Watson (eds.). *Civilian Morale*. Houghton Mifflin, Boston, 1942.

This same problem was approached from a slightly different angle by Watson and Seidman (1941) and Seidman (1943). In these studies, unemployed adults were asked, "If you feel that you weren't entirely successful on some of the jobs you have held, to what do you attribute that fact?" The answers, given in percentages in Table 12.7, show again that the nature of the work, congeniality of working conditions, and social contacts are more conducive to work satisfaction than wages and working hours. These four studies of Watson and Seidman corroborate other findings that concluded that wages are not the most important factor in job satisfaction.

A review of the material presented thus far in this chapter, and of that in the preceding chapter, should make it clear that job satisfaction can be measured in

TABLE 12.7 *Reasons Given for Work Dissatisfaction (in Percentages)*

Reasons	Seidman Study		Watson-Seidman Study
	Men	Women	Young Men
Deficiencies within self (personality)	32%	41%	24%
Nature of work; monotony	35	20	47
Uncongenial work conditions	11	9	9
Lack of education and experience	7	20	9
No promotion opportunity	5	1	9
Ill health	4	5	0
Social and economic conditions	4	3	0
Low wages	1	1	1
Long hours	1	0	1

SOURCE: Adapted from G. Watson and M. Seidman. Dissatisfactions in work. *Journal of Social Psychology*, 1941, *13*, 183–186; and J. M. Seidman. Dissatisfaction in work. *Journal of Social Psychology*, 1943, *17*, 93–97.

one of two ways. One method is to investigate the specific factors on the job and the resulting attitudes. The other, which is much more comprehensive, includes the overall factors that contribute to satisfaction in life. Neither method is necessarily right or wrong. However, progress can be made in this important field only after there is an understanding of the many factors which are involved and when research findings are reported in such a way as to allow meaningful comparison. For example, one who wanted to measure intelligence could not do so by making up six or eight or even eighty questions and assuming that they would measure intelligence, because the present development of intelligence testing would make such a process completely foolish. But pretty much the same thing is being done in measuring job satisfaction at the present time, and we are congratulating ourselves on our own good judgment in the firm belief that the other fellow does not really know what it is all about.

Job satisfaction, as pointed out before, is a generalized attitude resulting from many specific attitudes in three areas: specific job factors, individual adjustment, and group relationships. Since it is a highly complex attitude, the techniques for attitude measurement are the appropriate tools to use to study job satisfaction. (For a review of these, refer to Chapter 9.)

PREDICTING JOB SATISFACTION

Is it possible to select people who are likely to exhibit high job satisfaction? This notion assumes that a person is predisposed to job satisfaction or dissatisfaction prior to being employed. It requires that all applicants be screened, using a process which supposedly provides measures of the individual's intelligence, abilities, interests, and personality, at least to the extent of emotional stability.

Adjustment on the job, according to this philosophy, is dependent upon the individual's abilities, interests, and personality. When the specific job factors are related to his makeup, job satisfaction is possible. But when they clash there is job dissatisfaction, and the extent of the clash determines the depth of the dissatisfaction. After all, few people are likely to understand the problem in relation to their own limitations; it is much easier to avoid undermining the self and instead to ascribe the trouble to some outside factor such as the job.

The numerous studies on the relation of intelligence test scores and job performance suggest that for a particular occupation a score within a certain range is likely to be best. That is, a person may have either too much or too little intelligence to do a job successfully. Obviously, it is poor employment procedure always to hire the individual with the highest intelligence, regardless of the job; this usually results in considerable harm to both the individual and the job. As long ago as 1918, the Army Alpha testing program established that scores on intelligence tests varied according to previous occupation; Table 12.8 presents a portion of the findings reported by Douglas Fryer (1922). Although each of the selected occupations has a considerable range and there is overlapping from occupation to occupation, the hierarchy is plainly established.

Considerable differences are to be found within the same occupation. For example, Miner (1921) reports that salesmen for a technical product averaged 27 more points on an intelligence test than insurance salesmen. In turn, insurance

TABLE 12.8 *Occupational Intelligence Standards*

Occupation	*Average Score*	*Ranges of Scores of Middle 50% of Population (25th–75th Percentile)*
Engineer	161	110–183
Accountant	137	103–155
Chemist	119	94–139
Dentist	110	80–128
Clerk (office)	96	74–121
Conductor (railroad)	83	64–106
Electrician	81	57–109
Druggist	78	61–106
Toolmaker	67	50–92
Carpenter	60	40–84
Barber	55	34–78
Teamster	50	30–72
Mason	40	19–60
Shoemaker	35	19–57
Leather worker	30	16–41
Textile worker	26	18–60
Fisherman	20	15–51

SOURCE: D. Fryer. Occupational intelligence standards. *School and Society*, 1922, *16*, 273–277.

salesmen averaged 23 more points than wholesale salesmen, but the latter scored on the average 33 more points than retail salespeople. According to Snow (1927), duller individuals showed least dissatisfaction in highly repetitive work; but when the work was fairly complex, considerable dissatisfaction was manifest. The point to be stressed here is that an employee's intelligence is a factor in predetermining his job satisfaction. Too much intelligence—that is, more than the job requires—may well lead to dissatisfaction. Similarly, too little will make the job prove too much of a challenge and this may also lead to dissatisfaction.

This brief discussion of intelligence tests and job satisfaction should not lead to the assumption that there is necessarily a high correlation between such tests and job achievement. Nothing is further from the truth. The subject is introduced solely to suggest that maximum and minimum scores often furnish a lead regarding job satisfaction.

There are other abilities and aptitudes which can furnish similar clues. The previous job history is highly informative and should not be overlooked. Most employees who stay on a job from six months to one year can be assumed to be capable of doing the job. A person who does not have the required ability will find the job frustrating and leave it. When a job history is not available, as is often the case with young applicants, a battery of psychological tests to measure clerical ability, mechanical ability, and abilities in many other fields is often helpful. Tiffin and Greenly (1939) report a correlation of + 0.63 between scores on a hand precision test and foreman's ratings of a group of electrical fixture assemblers. Blum (1940)

finds a correlation of + 0.39 between a combined finger and tweezer dexterity score and earnings in a group of watch factory workers. Cook (1941) found that only 8 percent of the average group failed a coil winding test, whereas 72 percent in the below-average group failed it. Crissey (1944) points out that among test-selected employees, the turnover for personal reasons is 5 percent, as against a turnover of 12 percent for non–test-selected employees. He says further that those who score in the high third on a battery of tests maintain high production and contribute greatly to the improvement of employee morale on the job.

In addition to intelligence and other abilities, interest in the work contributes to job satisfaction. When a person's interests are in line with the job, he can be expected to be absorbed in the job. On gross analysis, interest can be divided into two categories: interest in people and interest in things. Individuals in the first group find the greatest outlet for their interest in jobs that essentially involve people —salesmen, lawyers, teachers, etc. People in the second group, on all levels from factory worker to the professional electrical engineer, find their maximum outlet in jobs which require designing or producing articles, tools, etc. More minute measurements of specific interests in relation to varieties of jobs are afforded by such test inventories as those constructed by Strong, Brainard, and Kuder.

A somewhat difficult question is the relation between interest and ability. In some respects, this is not unlike the famous "chicken and egg" problem, for it is probably true that in some cases the interest comes first but in other cases it grows out of an aptitude. Nevertheless, it is generally agreed that although the two are different, they tend to go together; the correlation between them is usually found to be + 0.50.

Last but not least of the contributors to job satisfaction is personality. One of the dimensions of personality is emotional stability—or "neurotic tendency." An individual's emotional stability is likely to manifest itself in satisfaction or dissatisfaction in a specific job. According to Fisher and Hanna (1931), "a large part of vocational maladjustment and industrial unrest are secondary to, and but a reflection of, emotional adjustment."

It is very likely that when everything is going smoothly the emotionally stable and the unstable show little difference on the job. However, when the pressure is on and difficult situations develop, petty annoyances take on major significance. It may well be that the individual reacts to these situations in proportion to his stability. The person who "flies off the handle" seems to go to pieces every time. The supervisor's choice is whether to ride out the situation or avoid it. Supervisors often say, "I can't tell him anything because he will get excited and rave," or "If I call the error to that girl's attention she will cry and then I'll be in a mess." What is really being said is that such people are emotionally unstable, even if the supervisors do not know the term or do not recognize the concept when it is called to their attention. A further characteristic of the emotionally unstable is the degree to which they allow one situation to affect a totally different situation. Thus a slight exchange of words with a coworker can have the individual "tied up in knots" not only on the job but even at home after the day's work is done. Similarly, such a person will bring a home situation to the job much more often than a stable individual will.

Security must also be considered as an important dimension of personality as it affects job satisfaction. Our earlier discussion of job security suggested that security was a relative rather than an absolute concept; that security is an attribute of the

individual. An insecure person will remain insecure even though his job is secure. Family background and many similar factors contribute to individual security.

One additional factor about the individual which must be taken into consideration if there is to be a complete understanding of job satisfaction is his adjustment to his life. Has his schooling led to satisfaction or dissatisfaction? This holds also for friends, hobbies, marital status, and all the other adjustments that the normal person makes adequately and with satisfaction. If the individual has a long list of "gripes" about being "done wrong," it is probable that sooner or later—probably sooner—he will find gripes and accompanying dissatisfaction in the job.

REFERENCES

Ash, P. The SRA employee inventory—a statistical analysis. *Personnel Psychology*, 1954, 7, 337–364.

Baehr, M. E. A factorial study of the SRA employee inventory. *Personnel Psychology*, 1954, 7, 319–336.

Bernberg, R. E. Socio-psychological factors in industrial morale: I The prediction of specific indicators. *Journal of Social Psychology*, 1952, 36, 73–82.

Blum, M. L. A contribution to manual aptitude measurement in industry. *Journal of Applied Psychology*, 1940, 24, 381–416.

Brayfield, A. H., and W. H. Crockett. Employee attitudes and employee performance. *Psychological Bulletin*, 1955, 52, 396–424.

Brody, M. The relationship between efficiency and job satisfaction. Unpublished master's thesis New York University, New York, 1945.

Cook, D. W. Psychological tests for unskilled jobs. *Personnel Series No. 50, American Management Association*, 1941.

Crissey, O. L. Aptitude testing improves shop morale. *American Machinist*, 1944, 88, 91–94.

Dabas, Z. S. The dimensions of morale An item factorization of the SRA employee inventory. *Personnel Psychology*, 1958, *11*, 217–234

Dunnette, M D. Factor structures of unusually satisfying and unusually dissatisfying job situations for six occupational groups. Paper presented at Midwestern Psychological Association Meeting, Chicago, 1965.

Ewen, R. B. Some determinants of job satisfaction: A study of the generality of Herzberg's theory. *Journal of Applied Psychology*, 1964, 48(3), 161–163.

Fisher, V. E., and J. V. Hanna *The dissatisfied worker*. Macmillan, New York, 1931.

Fleishman, E. A., E. F. Harris, and H. E. Burtt. *Leadership and supervision in industry*. Ohio State University, Bureau of Educational Research, Columbus, 1955.

Friedlander, F. Job characteristics as satisfiers and dissatisfiers. *Journal of Applied Psychology*, 1964, 48, 388–392.

Friedlander, F. Underlying sources of job satisfaction. *Journal of Applied Psychology*, 1963, 47(4), 246–250.

Friedlander, F., and E. Walton. Positive and negative motivations toward work. *Administrative Science Quarterly*, 1964, 9, 194–207.

Fryer, D. Occupational intelligence standards. *School and Sociology*, 1922, *16*, 273–277.

Gadel, M. S., and P. H. Kriedt. Relationship of aptitude, interest, performance, and job satisfaction of IBM operators. *Personnel Psychology*, 1952, 5, 207–212.

Giese, W. J., and H. W. Ruter. An objective analysis of morale. *Journal of Applied Psychology*, 1949, 33, 421–427.

Hamid ud-Din, Mohammed. The relationship between job performance and job satisfaction. *Dissertation Abstract*, 1953, *13*, 434–435 (Abstract).

Harrison, R. Cumulative communality cluster analysis of workers' job attitudes. *Journal of Applied Psychology*, 1961, 45, 123–125.

Heron, A. Satisfaction and satisfactoriness Complementary aspects of occupational adjustment. *Occupational Psychology*, 1954, 28, 140–153.

Herzberg, F., B. Mausner, R. O. Peterson, and D. F. Capwell. *Job attitudes: Review of research and opinion.* Psychological Services of Pittsburgh, Pittsburgh, 1957.

Herzberg, F., B. Mausner, and B. B. Snyderman. *The motivation to work.* Wiley, New York, 1959, 157.

Hoppock, R. *Job satisfaction.* Harper & Row, New York, 1935.

Hulin, C. L., and P. C. Smith. Sex differences in job satisfaction. *Journal of Applied Psychology*, April 1964, 48(2), 88–92.

Kahn, R. L. Productivity and job satisfaction. *Personnel Psychology*, 1960, 13, 275–287.

Kates, S. L. Rorschach responses related to vocational interests and job satisfaction. *Psychological Monographs*, 1950, 64(3).

Katzell, R. A., R. S. Barrett, and T. C. Parker. Job satisfaction, job performance, and situational characteristics. *Journal of Applied Psychology*, 1961, 45, 65–72.

Kendall, L. M., P. C. Smith, C. L. Hulin, and E. A. Locke. Cornell studies of job satisfaction· IV: The relative validity of the job description index and other methods of measurement of job satisfaction. Unpublished manuscript, Ithaca, N.Y., 1963

Knutson, A. L. Personal security as related to station in life. *Psychological Monographs*, 1952, 66(4).

Kornhauser, A. Analysis of class structure of contemporary American society. In G. W. Hartmann and T. M. Newcomb (eds). *Industrial conflict: A psychological interpretation.* Cordon Co., New York, 1940.

Kornhauser, A. W. *Mental health of the industrial worker.* Wiley, New York, 1965.

Lawshe, C. H., and B. F. Nagle Productivity and attitude toward supervision. *Journal of Applied Psychology*, 1953, 37, 159–162.

Locke, E. A., P. C. Smith, L. M. Kendall, C. L Hulin, and A. M. Miller. Convergent and discriminant validity for areas and methods of rating job satisfaction. *Journal of Applied Psychology*, 1964, 48, 313–319.

Lodahl, T. M. Patterns of job attitudes in two assembly technologies. *Administrative Science Quarterly*, 1964, 8, 482–519.

Lopez, F. M. A psychological analysis of the relationship of role consensus and personality consensus to job satisfaction and job performance. Unpublished doctoral dissertation. Columbia University, New York, 1962.

Mann, F. C., B. P. Indik, and V. H. Vroom. *The productivity of work groups.* University of Michigan, Institute for Social Research, Survey Research Center, Ann Arbor, 1963.

Miner, J. B. Standardizing tests for vocational guidance. *School and Sociology*, 1921, 13, 629–633.

Morse, N. C *Satisfaction in the white-collar job.* University of Michigan Press, Ann Arbor, 1953.

Mossin, A. C. *Selling performance and contentment in relation to school background.* Columbia University, Teachers' College, Bureau of Publications, New York, 1949.

Myers, M. S. Who are your motivated workers? *Harvard Business Review*, 1964, 42(1), 73–88.

Porter, L. W. Personnel management. *Annual Review of Psychology*, 1966, 17, 395–422.

Roach, D. E. Dimensions of employee morale. *Personnel Psychology*, 1958, 11, 419–431.

Robinson, H. A., and R. Hoppock. Job satisfaction résumé of 1951. *Occupations*, 1952, 30, 594–598.

Schaffer, R. H. Job satisfaction as related to need satisfaction in work. *Psychological Monographs*, 1953, 67(14).

Schwartz, M. M., E. Jenusaitis, and H. Stark. Motivational factors among supervisors in the utility industry. *Personnel Psychology*, 1963, 16, 45–53.

Seidman, J. M. Dissatisfaction in work. *Journal of Social Psychology*, 1943, 17, 93–97.

Seidman, J. M., and G. Watson. Satisfaction in work. *Journal of Consulting Psychology*, 1940, *4*, 117–120.

Severin, D. The predictability of various kinds of criteria. *Personnel Psychology*, 1952, *5*, 93–104.

Sirota, D. Job performance as related to attitudes, motivation and understanding. Unpublished research report (mimeographed). University of Michigan, Institute for Social Research, Ann Arbor, 1958.

Snow, A. J. Labor turnover and mental alertness test scores. *Journal of Applied Psychology*, 1927, *2*, 191–195.

Stogdill, R. M. *Individual behavior and group achievement*. Oxford, Fair Lawn, N.J., 1959.

Tiffin, J., and A. J. Greenly. Employee selection tests for electrical fixture assemblers. *Journal of Applied Psychology*, 1939, *23*, 240–263.

Twery, R., J. Schmid, Jr., and C. Wrigley. Some factors in job satisfaction· A comparison of three methods of analysis. *Educational Psychological Measurement*, 1958, *18*, 189–202.

Vroom, V. H. *Some personality determinants of the effects of participation*. Prentice-Hall, Englewood Cliffs, N.J., 1960.

Vroom, V. H. *Work and motivation*. Wiley New York, 1964.

Watson, G. Work satisfaction. In G. W. Hartmann, and T. Newcomb (eds.) *Industrial conflict*. Cordon Co., New York, 1939.

Watson, G., and M. Seidman. Dissatisfaction in work. *Journal of Social Psychology*, 1941, *13*, 183–186.

Weitz, J. A neglected concept in the study of job satisfaction. *Personnel Psychology*, 1952, *5*, 201–205.

Wernimont, P. F. *Intrinsic and extrinsic factors in job satisfaction*. Doctoral thesis. University of Minnesota, Minneapolis, 1964.

Wherry, R. J. An orthogonal re-rotation of the Baehr and Ash studies of the SRA employee inventory. *Personnel Psychology*, 1954, *7*, 365–380.

INDUSTRIAL MORALE 13

Morale is a term usually applied to civilian populations and armies during wartime; its conspicuous presence or absence is often considered vital to the outcome of the war. The term is also applied to athletic teams, and here also it can contribute to success. Sport fans know that high morale can overcome many obstacles and physical disadvantages. The term is similarly applied to industry. Of late, the concept of industrial morale has received considerable attention on the part of management.

DEFINING MORALE

Industrial morale may be defined as *the possession of a feeling, on the part of the employee, of being accepted and belonging to a group of employees through adherence to common goals and confidence in the desirability of these goals.* Let's look at this definition more closely. It consists of three different aspects: (1) feeling accepted by one's work group; (2) sharing common goals with one's work group, and (3) having confidence in the desirability of these goals.

The term "work group" as used in this definition can refer to groups of various sizes within the total organization. Thus a person may be part of a small work group (e.g., a section) which in turn is part of a larger work group such as a department which in turn is part of a larger group called a division which in turn is part of some larger group called a company. One can talk about sectional morale, departmental morale, divisional morale, and company morale. Each means something different, however, since the goals of the four reference groups may be quite different. For example, one might conceive of a department work group which had very low morale because the group members could not agree among themselves concerning goals. However, each individual member might feel that his own goals were quite compatible with those of men in other departments throughout the plant and also compatible with those of management. Thus their feeling of belonging might be minimal within the department but quite high if the reference group were shifted to all other plant employees.

Although morale is related to job satisfaction, it is not the same thing. There is no justification for using the two terms interchangeably. Job satisfaction, as we have seen, is the result of the various attitudes the *individual* holds toward his job, related factors, and life in general. Industrial morale is the composite expression

of the attitudes of the *various individuals* employed by a company. It is generated by the group and may best be considered as a by-product of the group.

Definitions of morale vary widely. Guion (1958) has collected a number of different definitions of morale from various sources.

1. Morale defined as the absence of conflict
2. Morale defined as a feeling of happiness
3. Morale defined as good personal adjustment
4. Morale defined as ego-involvement in one's job
5. Morale defined as group cohesiveness
6. Morale defined as a collection of job-related attitudes
7. Morale defined as an individual's acceptance of the goals of the group

Examining each of these in turn, Guion finds each by itself to be less than satisfactory and proposes what he feels to be a more general definition which includes the best aspects of all seven.[1] He defines morale as "the extent to which an individual's needs are satisfied and the extent to which the individual perceives that satisfaction as stemming from his total job situation." According to Guion, "High morale, under this concept, would be reflected by little aggressive or frustration-instigated conflict, by a reasonably euphoric work force, by fairly well-adjusted employees who can become quite ego involved in their work, by many favorable attitudes, and by the cohesiveness which comes from finding personal need satisfactions within a group."

Guion does another interesting bit of speculating. In the same article (Guion, 1958) he proposes five different attributes which he feels are essential for an adequate concept of morale.

GUION'S CRITERIA FOR A CONCEPT OF MORALE

1. It must recognize the dynamic complexity of morale.
2. It must treat morale as basically an attribute of the individual.
3. It must recognize that morale exists with reference to the job situation.
4. It must recognize the role of motivation in morale
5. It must be applicable to employees at any job level.

Stagner (1958) has also defined morale. His definition would seem to fall in Guion's category 7. Stagner has been sufficiently explicit in attempting to formally conceptualize the construct of morale that a few words of discussion about his "model" are worth including. To quote Stagner:

> Morale, I think, must always be defined in terms of an individual-group relationship, it is an index of the extent to which the individual perceives a probability of satisfying his own motives through cooperation with the group. Obviously, then, there is no such phenomenon as morale *in general*, the state of an individual's morale must be gauged relative to some specific group, such as his company, his informal work group, or his union.

and

> High morale exists when the individual perceives himself as a member of a group, and perceives a high probability of achieving both individual and group goals through a course of action. He consciously seeks to achieve the goals of the group because they are im-

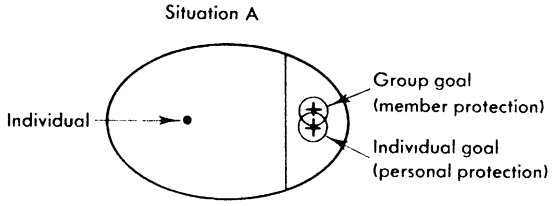
[1] Note the similarity between definitions 5 and 7 and the first two parts of the definition given by your authors.

portant to him—they have become individual goals—or because they provide a *pathway* to his own personal goals which are not related to the group.

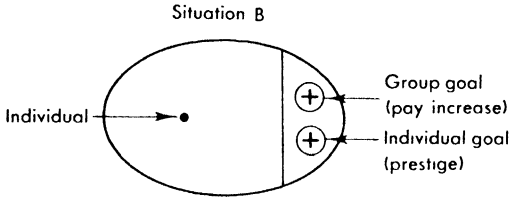
When we attempt to apply motivational analysis to the problem so defined, we find that there are three separate phases to be examined (a) the individual's private goals, (b) the group goals, and (c) the perceived relationship between these sets of goals.

Thus morale to Stagner depends upon (1) the person's perception of himself as a component of a group and (2) perception of his goals as being identical to or contiguous with group goals. He gives several examples of the latter aspect of this, as shown below:

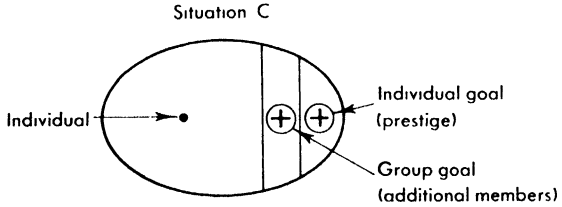
Situation A: Achievement of the group implies achievement of the individual goal—i.e., they are a single process. An example is the union leader who wins a contract that protects the union members and also protects the leader. According to Stagner, morale will be at a maximum when situation A exists.[2]



Situation B: Achievement of the group goal is closely associated with the individual goal. The example is a union leader getting a contract for his union which brings a *pay raise* to the members and more *prestige* to the leader.

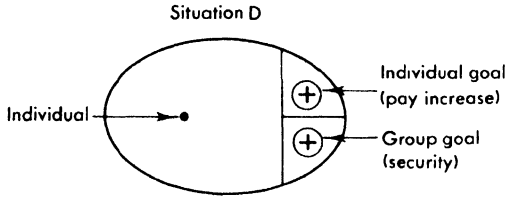


Situation C: Achievement of the group goal is a necessary precondition in order to achieve the individual goal. The example would be the union leader who tries to organize more workers hoping to increase his prestige.



[2] These diagrams are modified from Stagner (1958).

Situation D: Achievement of the group goal and the individual goal may be accomplished separately or independently. The example is a worker who produces at a very high rate to achieve the individual goal of increased pay while at the same time ignoring or violating the group goal of security through restricted work output. The situation leads to lowest morale.



Stagner concludes that the way to maximize morale is to make every attempt to assure that group goals and individual goals coincide as closely as possible.

WHAT MORALE IS NOT

We mentioned earlier that morale and job satisfaction are not to be confused. Industrial morale is definitely not an average of individual attitudes. For example, a company's morale may be low even though many employees enjoy considerable job satisfaction. Certain individuals who are dissatisfied with their jobs may lower the overall morale. Even one or two completely disgruntled workers, who are not recognized as such but rather are respected by their fellow employees, will tend to lower the morale of the group appreciably. Similarly, an erratic, unpopular, or ineffective foreman will lower group morale, even though most of the factors contributing to job satisfaction are present.

Conversely, a few people may have a noticeable effect in raising industrial morale even when little is offered to most of the employees in terms of the specific factors of job satisfaction. World War II showed that two otherwise similar squadrons can be very dissimilar as far as morale is concerned. Morale in similar departments in a factory often varies in ways that cannot be understood by studying the job satisfaction factors. An individual or a few members of a group can raise the morale of the entire group if his attitudes and behaviors can be adopted by the others.

The fact that morale is a by-product of the group and can often be generated by small segments of the group is important. It explains many of the paradoxes that are connected with morale. For example, the discharge of an inefficient and troublesome worker may have no effect on the morale of the group if it has not considered him a part of it; the group will regard this as a single and separate event. However, if he was liked by the others and they feel that he was not really troublesome and inefficient, then the group's morale will go down. On the other hand, if his fellow employees dislike him and hope that he will "get his," the group morale could conceivably be raised by his discharge. Another paradox in regard to morale is the statement that workers strike when their morale is low. This is contradicted by union organizers, who insist that morale must be high before a strike call can be promulgated. Now morale cannot be both high and low at the same time in the same group of workers. Obviously, in interpreting such statements

about the group's morale, the point of view must be considered. From management's point of view, the morale is low if workers side with the organizer; but from the union's point of view under these circumstances, the morale is high.

DETERMINERS OF MORALE

Morale can best be understood in terms of four determiners. The most outstanding determinant is a "feeling of togetherness" or group cooperation. The second is the need for a goal. Third, there must be observable progress toward the goal. Fourth, the individuals in the group must have specific meaningful tasks that are necessary to the achievement of the goal.

GROUP COHESIVENESS

Employees in a normal work situation rarely function as totally isolated individuals. Whether management is aware of it or not, the workers are likely to form a group or a number of subgroups. The ideal situation, and one which makes for the highest morale, is one in which there is a single group that includes all employees, the representatives of the employer, and the employer. The extent to which management attempts to achieve this is the extent to which high morale is likely to be present. Unfortunately, the factor of group formation is too often overlooked in industry. The introduction of favorable environmental factors such as changes in illumination, music, or air conditioning may or may not work, that is, increase production. In and of themselves such factors are unpredictable. The key to whether they will work lies in the group's reaction to the change. If the group believes that the change will benefit it, the reaction will be favorable. On the other hand, if the group is suspicious of the change, it will resist it. Changes instituted by management, with the most altruistic intentions, often boomerang because management has overlooked the importance of the "feeling of togetherness," or group cooperation. It cannot be assumed that increasing material efficiency necessarily increases people's capacity to work together. In fact, unless management takes specific steps to encourage the formation of a group that includes labor and management working together, the employees are likely to form their own group and deliberately exclude management. Management's best chance for inclusion is the adoption of a democratic group structure and the providing of an opportunity for the four determiners of morale to function.

It is impossible to deny the existence of a social structure in any company. To do so is to refuse to face reality. But management does this when it insists that people work solely for money. Men and women lead a social life on the job; they make friends and enemies; they exchange confidences, meet socially after hours, eat together, and do favors for one another. In short, consciously and unconsciously they form groups. These groups can be the basic core of morale formation if management will only recognize this and if it can funnel this energy into the proper channels of cooperation. Management's failure to recognize this leads to many of the mistakes it makes. Technical changes result in social changes; and while the technical changes may be logical from management's point of view, they are often unsuccessful because the social changes which accompany them have been overlooked or miscalculated.

GOALS

Promoting group cooperation is easier if the group has a goal to achieve. The goals in industry may be more obscure than such goals as winning a war or a football game, but with proper guidance they can be made clear. Doing the best one can on the job may constitute a goal, provided the employee receives evidence that this goal is understood. Advancement, security, increased earnings, and individual and group welfare can all be goals, provided management encourages them and the employee has evidence that they are real and attainable.

Some "scientific managers" have suggested that slogans are an excellent means of increasing morale, on the assumption that a good slogan may very well become the goal. This is best illustrated by the story about a big boss who attended one of these "scientific" sessions and was impressed by the suggestion. He set the slogan "Do it now" as the goal and had it posted in various conspicuous spots in the plant. From his point of view, this goal turned out to be anything but desirable, for the bookkeeper immediately absconded with the company's funds, the salesman married the boss's daughter, and the office boy put the wrong solution into the copying machine.

PROGRESS TOWARD GOALS

In addition to having a goal, it must be possible for employees to make observable progress toward it. A good example is seen in Marrow's (1942) work in his own plant. One group of power machine operators was told the level of production they

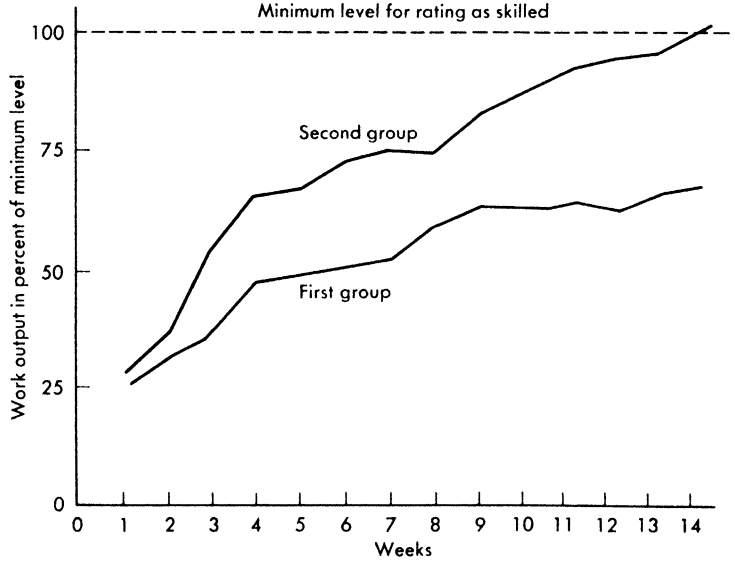


Figure 13.1. Effect of observable progress toward goal. (From A. J. Marrow. Cited in K. Lewin and G. Watson, eds. *Civilian morale.* Houghton Mifflin, Boston, 1942.)

would have to reach within 14 weeks. Another group was told the final goal but was also given weekly goals. Figure 13.1 shows the effect on performance of a too distant goal which the worker judges to be unattainable.

MEANINGFUL TASKS

The last of the four determiners of employee morale involves specific meaningful tasks for the individual in the group to perform, and a sense of participating in the group's work toward the goal. If the group's goal is 1000 units per day, each individual member must understand clearly how his specific job contributes to its attainment. If the output is merely to management's advantage and does not contribute in some definite way to the employee's advantage, the goal is likely to be regarded as management's and not the worker's. However, if the employees are consulted or properly informed about setting the goal and actually derive a fair share of the benefits that accrue, there is then a strong goal. Even a floor boy can see how he contributes by helping the assemblers, although he does no assembling. The first workman passes his completed share of the unit on to the second worker, and this continues, with every member of the group participating in a meaningful way in the achievement of the goal.

Yuzuk (1961) has reported the results of a study on employee morale which provides considerable information concerning the many facets and complexity of the term. Using factor analysis he found morale to consist of nine different dimensions:

1. Adequacy of communication
2. Hours of work
3. Overall competence of fellow employees
4. Interpersonal relations with fellow employees
5. Status and recognition
6. Conditions of work
7. Interpersonal relations with supervisor
8. Technical competence of supervisor
9. Job satisfaction

He also found that the workers having the highest morale tended to be the most experienced men who had the longest tenure and who held the highest labor grades.

Another factor analysis of moral dimensions has been done by Gordon (1955). Gordon defined morale as "the feeling of well-being that an individual experiences when his needs are filled to his satisfaction." He found four different dimensions:

1. General need satisfaction
2. Recognition and status
3. Need for self-respect
4. Need for self-expression

Although he did not attempt to relate morale to actual work performance, Gordon indicated that a better understanding of morale dimensions was essential to our knowledge of man and how he performs his work.

MEASUREMENT OF MORALE

SUBJECTIVE MEASURES

An interesting measure of morale is one that measures the group and subgroup formations. A technique has been devised for this purpose by Moreno (1943); he has called it *sociometry*. Jenkins (1947) has modified Moreno's technique and

called his method the *nominating technique*. He has used it with considerable success in studying morale in the Navy. Further reference to this work will be made in the section on leadership.

The technique is simple. Each person in the group is asked to name the person in the group whom he considers the best worker, or the one who would make the best supervisor, or the most fair-minded person. Each individual is represented on a diagram by a circle; an arrow is drawn from it to the circle representing the person he names. The leader of the group is the one who receives the greatest number of choices. If this individual is also the boss, then he is both the formal and the informal leader. However, if the formal leader does not receive many choices, it is immediately apparent that he is leader in name only and that the informal leader, if there is one, is the real force in the group. This person is known as the "star." Actually, there may be more than one star in a group. A person who is never chosen is called an "isolate"; he may be so unimportant that if he dropped out of the group its structure and morale would not change. Sometimes two workers choose each other. Such individuals are usually known as "mutual admiration societies" and may function as separates rather than in the group. A triangle occurs when X chooses Y who in turn chooses Z; this is the beginning of a subgroup and may be enlarged into a clique. Figures 13.2 through 13.4 are simple sociograms that illustrate some of the various types of group structure.

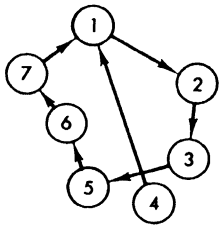


Figure 13.2. Sociogram of a cohesive group without a strong leader.

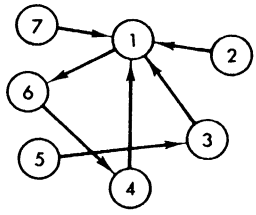


Figure 13.3. Sociogram of a group with a strong leader.

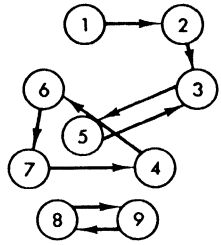


Figure 13.4. Sociogram of an unstructured group with cliques, isolates, and mutual admiration societies.

In actual practice, sociograms may become very complicated. Figure 13.5 shows a complex sociogram made by Jenkins in his work with a Navy squadron.

The study of sociograms is valuable. It can furnish leads as to clique formations and the extent to which they may act as disruptive influences in the organization. It can substantiate observations which formerly may have led merely to suspicions. It can pick a potential leader far in advance of any principle of seniority. A sociogram is much more than a popularity poll.

Sociometrically selected work teams can increase production. Van Zelst (1952) obtained a 5 percent saving in total production costs by allowing carpenters and bricklayers to choose their "buddies" as partners in work. Work teams were assigned based upon stated preferences. Twenty-two workers were assigned their first choices as partners, 28 had their second choice, and 16 obtained their third choices. Eight of the 74 were isolates, that is, not selected, but were of course assigned.

Van Zelst reports the following subjective report of one of the workers: "Seems as though everything flows a lot smoother. It makes you feel a lot more comfortable working. I don't waste any time bickering about who's going to do what and how.

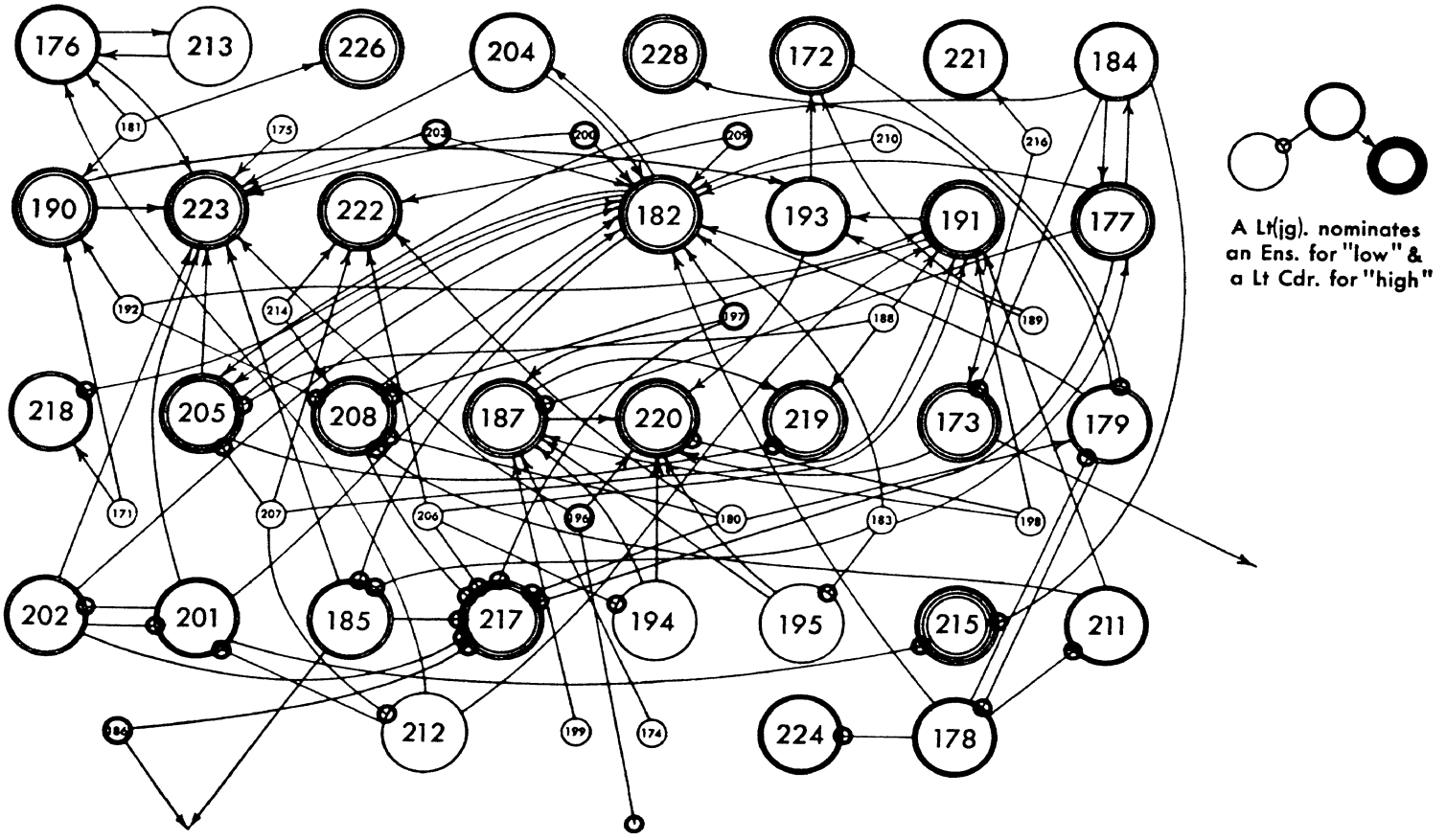


Figure 13.5. Sociogram of a Navy squadron. (Courtesy the Navy Department.)

We just seem to go ahead and do it. The work's a lot more interesting, too, when you've got your buddy working with you. You certainly like it a lot better anyway."

This principle of having workers assign themselves to pairs on work locations rather than have management superimpose rigid rules is not idle folly. It is of the essence in building morale. More of it should be applied to industry, since workers must often work in teams.

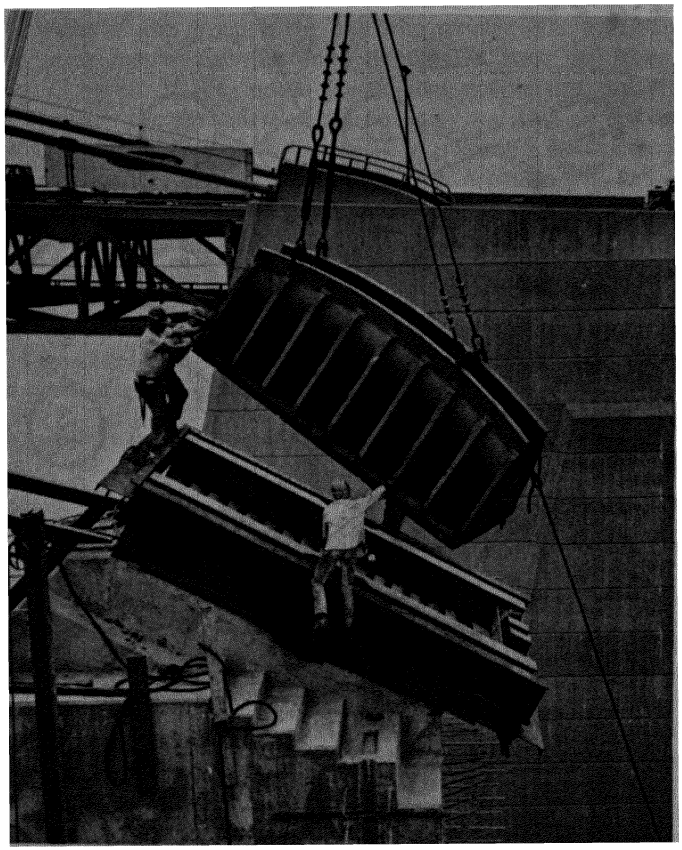


Figure 13.6. Without teamwork, there could be a critical incident. (Courtesy Bethlehem Steel.)

But in measuring the other three determinants of morale—goal, progress toward goal, and meaningful participation—the sociogram may be of little value. In order to obtain objective data in these fields, the attitude scale, the questionnaire, or the interview has to be used. The material already presented has indicated the approach and method that should be chosen.

OBJECTIVE MEASURES

The suggestion that the sociogram, together with questionnaires, attitude scales, and interviews, affords the best measure of morale has been made without overlooking the many indices that are usually considered measurements of this factor.

Among these indices are strikes, labor turnover, absenteeism, grievances, suggestions, and production figures.

Giese and Ruter (1949) have attempted to predict the morale of departments in a company from objective data. They obtained a multiple correlation of +0.71 between the six objective factors they studied and morale as measured by a questionnaire. Because of the high correlations, they proposed an objective morale index which could be obtained by measuring the following factors: productive efficiency, error efficiency not affecting customers, error efficiency affecting customers, turnover, lateness, and absenteeism. They found that when morale is low, department absenteeism and lateness tend to be high. They found only a slight relationship between morale and productive efficiency. The notion that morale of a department may be reflected through certain types of group performance in combination is an interesting one and is probably more meaningful than either attempting to correlate morale with a single criterion, such as production, or confusing morale (a group phenomenon) with job satisfaction (an individual phenomenon).

Bernberg (1952) reports a study based on 890 hourly paid employees in a large aircraft manufacturing plant. On the basis of a statistical analysis of the results, he found that no significant relationship exists between the tests of morale and the specific indicators for the prediction of individuals, but that tests of morale can be predicted for departments and factories. Bernberg's results underscore the fact that morale is a group phenomenon rather than an individual phenomenon.

He had five indicators: absences, tardiness, short-time absences, trips to a medical unit, and merit rating. In addition, he combined these into a total indicator. He used an indirect method of attitude measurement as a measure of morale and also a direct attitude measurement. His objective was to determine which of two morale theories was more predictive of performance indicators. One theory considered morale a group phenomenon and the other considered it as the degree of acceptance of formal organization. Actually, the difference between these two theories was about on the level of hair-splitting. Both were intended to measure a group phenomenon. Both avoided considering individual attitude and job satisfaction as equivalent to morale. His results indicated that these two measures correlate to the extent of +0.77. Clearly, both hypotheses predict group differences (i.e., departmental differences) but not individual differences on performance variables. This study obtained its data from 890 of the 1009 hourly workers in five departments of three factories in an aircraft company.

PROBLEMS WITH OBJECTIVE MORALE MEASURES

A word of caution is necessary with regard to the idea that objective indicators can be regarded as predictors of morale. It may be best to regard such concepts as external indicators flashing red lights demanding a stop-and-look reaction. While these indicators may indicate low morale, they may otherwise indicate ineffective leadership or inadequacies of training, selection, or work standards.

The difficulty with these indicators is that they may be due only in part to morale; there is no certainty as to how much other factors enter the picture. For example, a strike may indicate not low morale but a reflection of business conditions. Similarly, workers with low morale may not dare to strike because they feel much too defeated and beaten; in this instance the absence of a strike is due to low rather than high

morale. Nor can absenteeism and labor turnover always be looked upon as indications of low morale because conditions in the labor market, the weather, and other factors may induce such behavior on the part of the worker.

Workers whose morale is high may want to make suggestions but refrain from doing so because other employees might resent their ideas. Roethlisberger (1946) cites one company that turned its employee suggestions over to a judgment committee and made suitable awards. Management posted the notice of awards on its bulletin boards in order to stimulate further suggestions and give the winners recognition. This well-meaning program created increasing dissatisfaction. Foremen often discriminated against workers who won when the suggestion related to work within the foremen's area of responsibility, for they interpreted such suggestions as criticisms of their work. Likewise, workers whose jobs were simplified as a result of the suggestions, and who consequently received lower pay rates, put pressure on the person responsible for the suggestion. In addition to these troubles, when the winner owed money, claimants appeared and asked the company to settle on the basis of the employee's "windfall." Management finally had to modify its procedure and publish awards by code number, thereby eliminating the recognition factor which had been part of its original intention. To put it briefly, this suggestion system did not develop or promote cooperation, and consequently it served to lower rather than raise morale.

METHODS OF INCREASING INDUSTRIAL MORALE

The measurement of industrial morale serves as a barometer to management. If the findings indicate low morale, this must be regarded as a challenge and steps must then be taken to improve it. Four methods have been used at various times to increase industrial morale:

1. Expert approach
2. Industrial spy
3. Industrial counselor
4. Employee problem approach

These four methods differ, of course, in terms of the degree to which they have been used by industry. For example, the industrial spy method is now primarily of historical interest only, while the employee problem approach is continuing to gain in popularity. Unfortunately, a word of caution needs to be made about all four methods prior to examining them in detail. None of them has ever actually been demonstrated as a truly effective tool for increasing employee morale! There is almost a complete lack of objective data regarding their respective worth for this purpose. Still, they continue to be employed on the assumption that they do accomplish their intended purpose. Perhaps, eventually, the necessary empirical supporting evidence will be forthcoming to show their use is indeed justified. Until that occurs, they should be viewed and used with considerable caution.

EXPERT APPROACH

The expert approach, the most frequently used technique, has resulted in mixed success and failure. The basic feature of this approach is the use of an "expert" in making a spot survey. He tours the plant, talks to key people, and writes a report in which he gives advice. Sometimes he suggests any or all of the following items: posters, slogans, speeches, parties, welfare associations. These are likely to give morale a slight spurt for a short period. When things lapse back to normal, the company calls in another "expert." Since it is now somewhat more critical, it shows him the previous report and intimates that things did not go as predicted. Of course, the second expert knows this, for otherwise why would he have been called in? This being the case, he has to produce some new devices for boosting morale. He suggests bowling contests or better cafeteria meals, an effective house organ or a suggestion system. Again there is a slight lift in morale for a short time and then things take a turn for the worse.

The point is that any or all of the five suggestions of the first expert might work under certain conditions, but the conditions are never suggested or complied with. This is true also of the four suggestions of the second expert. These nine suggestions do not by any means exhaust the list; there are many, many more. Some companies and their executives seem to thrive on them. When any of the suggestions promote group cooperation, establish goals, allow for observable progress toward goals, and present an opportunity for employees to participate meaningfully and actively in their achievement, they can increase morale. Checking the above list of suggestions makes clear that few in themselves encourage cooperation or meet the requirements of the other determiners of morale. We are reminded of a dance given for employees of a large company which was supposedly promoted by their "benevolent association." It was a full-dress affair and the president of the company was present to show that he was just one of the boys and to award prizes. Early in the evening two of the employees put on an unscheduled dancing exhibition. In uninhibited fashion, they twisted, turned, and contorted, to the delight of the huge circle which formed around them. The president's reaction to what he considered a shocking and revolting performance was to order it stopped immediately and to warn the couple about conduct unbecoming ladies and gentlemen of the X Company. Needless to say, he effectively dampened everyone's spirits and unknowingly made himself very unpopular. Most of the employees felt that if he "couldn't take it" he should have stayed at home; besides, he was not their boss on Saturday night. The morale was lowered far out of proportion to the slight increase that such a party could have created.

INDUSTRIAL SPY TECHNIQUE

Although dangerous, this technique actually has been used as a means of increasing morale. An expert whose true identity is unknown to the employees, and usually to many of the executives, is given a job in a plant under the guise of an ordinary worker. It is his task to gain acceptance as a fellow employee and thus acquire intimate knowledge of the group and its complaints. He puts this information into a report. Sometimes such reports do not refer to specific employees but deal mainly with uncovering environmental conditions and technical difficulties which, in his opinion, lower morale.

The ethics involved in this type of report are not as distasteful as in the case of reports which take advantage of the confidences gained by the "spy" to locate the individuals who, according to management's definition, are "agitators" or "trouble-makers." Sometimes union organizers and members are put in this category. Eventually the "spy" disappears, and the workers wonder what happened to that "nice guy." His recommendations are acted upon through dismissals for "just cause."

Some readers may be shocked by such tactics, but to those who believe that all is fair in love and war this technique is very appealing. It should be emphasized that not all industrial "spies" are dastardly. Some do a very ethical job and their reports show that they are really concerned with employee welfare and increased morale. Some even go so far as to place the blame on ineffective management when they find that this is where the trouble lies.

INDUSTRIAL COUNSELOR METHOD

The industrial counselor has none of the unpleasant connotations that are attached to the industrial spy. As we saw earlier, the Hawthorne Studies culminated in the suggestion that the industrial counselor is useful in increasing morale by improving relations and acting as a communication channel between employee and employer. The employees are told that the industrial counselor is a representative of the personnel department, and they are paid their average earnings during the time they spend talking with him. Experiences at Western Electric indicate that employees are willing to bring their problems to and talk freely with a counselor.

Besides being useful in aiding supervisor-employee and employee-management relations, the industrial counseling program has also been valuable in improving the personal adjustment of individual employees. A case cited in *Fatigue of Workers*[3] is a good illustration of this point:

> One day an assistant foreman called the attention of the counselors to a problem which had been worrying him for some time. The problem was a man of forty-eight who had had about twenty years of service with the company His history was as follows. In his early years of service he was apparently an efficient worker He was made a group chief in 1923 and remained in that position until 1931. But at that time he was considered one of the least efficient supervisors in the group, and thereafter his fall was gradual but steady. He was demoted from job to job, and in each successive job he failed to measure up to the proper standards of work. When the foreman consulted the counselors this employee was at work on a job of the lowest grade, being paid considerably more than this grade warranted, and not earning his money. These were the essentials of the problem as they appeared to the heads of the department when the foreman called in the counseling organization. They felt that they had given the employee every possible opportunity to make good and that he had failed. They also were disturbed because he appeared to be drinking heavily, and, more recently, because he was not at all cooperative.
>
> The department proposed to shift the employee to a grade 1 (lowest grade) job in another organization and cut his hourly rate to the minimum of the grade. They proposed to explain to him that this would be his last chance and that if he failed he would be dismissed. In the meantime a counselor had been at work. He had had a preliminary interview with the employee which had indicated that the employee's failure was the result of a personal maladjustment. He asked that the transfer be delayed a few months or until a detailed study of the case could be made.

[3] National Research Council. *The fatigue of workers.* Reinhold Book Corporation, a subsidiary of Chapman-Reinhold, Inc., New York. Reprinted by permission.

The study consisted of interviews with the employee, a thorough physical examination, several intelligence and vocational tests, and a careful investigation of the employee's supervisors, his associates on the job, and the mechanics of all the jobs on which he had failed. Here the report may be cited again: "The interview showed an employee who was raised in one of the west-side slum districts The family consisted of two older sisters, an older brother, the employee, and a younger brother. The mother died when the employee was very young, and he does not remember her. The older sister managed the house. The father was a strict disciplinarian. The children were punished frequently and they were not allowed ordinary liberties This was so pronounced that the older brother left home as soon as he finished school. The father was employed as a laborer until his death, which was shortly after the employee graduated from grammar school The employee worked during vacation time while he was in seventh and eighth grades, and he found a steady job immediately after leaving school. He worked for several concerns until he enlisted in the army during the World War, and when he returned at the age of twenty-nine he was employed at Hawthorne

During this time he formed a close relationship with his older sister who never married. She kept house for him until a few years ago when he moved to a hotel across from the Plant. This move was made during the period when he was having difficulty on the job, but he continued to contribute toward her support and still has a very loyal attitude toward her.

Such was the past history of the employee, as it was revealed in his interviews with the personnel counselor. His attitudes in the immediate situation were described in the report as follows:

"1. Fear of the supervisors (foremen especially) Belief that they had it in for him. Feeling that they were watching him all the time. He could feel the foreman's eyes on him when the foreman was at the other end of the room.

"2. Belief that his nervousness was a 'shell shock' hangover from the war, in spite of the fact that he was in good health for years after the war and was not nervous until recently.

"3. Fear of sickness. He had been under the care of several doctors who tried to convince him that he was all right. He took 'nerve medicine' regularly four times a day. His closet shelf was full of pill boxes and cold cures. He wore a jacket on days when it was so hot that his shirt was wet with sweat He talked of his health, his sister's health, boyhood friends who had terrible diseases. He went into a tantrum if the hotel room was cold or the department was drafty.

"4. His only associate outside the Plant was a tubercular war veteran with a shady character, and he did not see him often. Most of his time was spent in his room with a cheap magazine. He also did not talk to anyone in the department unless they asked him a question."

The method used in dealing with the employee was typical of personnel counseling. He was interviewed daily until he had expressed much of the emotional disturbance in his thinking, and in expressing it, freed himself from it Thereafter, the interviews were limited to two a week, and the counselor concentrated on helping him to think sensibly about his work, his supervisor, and his associates outside the Plant As soon as an improvement was recognizable, the counselor went to the proper group and section chiefs and asked them to talk about the employee. They were asked to express their criticism in as much detail as possible. The counselor did not differ with any of the opinions expressed except to make the general comment that he thought the employee's attitude had improved. This process was carried so far that the counseling organization did not object when the division chief called with the information that the employee's hourly rate would have to be cut. This event occurred six months after the case had been opened. The organization asked only that the explanation be given to the employee in terms which

would carry conviction to him. Later the organization attempted to assist the section chief in preparing a statement for the employee, and in this conversation the section chief found his case very weak and decided to do nothing.

The report continues as follows: "This approach to the problem has resulted in the following changes in the situation:

"1. The supervisors are taking a genuine interest in the employee and they are helping him with encouragement instead of criticism. Both the group and the section chiefs are now taking personal credit for the improvement and they claim to have made arrangements to assign him to higher graded work in the near future.
"2. The employee has increased his efficiency from between 60% and 70% to about 100%. There appears to be little lost motion in his activities on the job, and he appears to be getting a real feeling of satisfaction in doing his job better than the rest.

"He is also very friendly with all of the employees, both men and women, who work near him, and he spends his rest periods talking with a group of employees who work in the other end of the room.

"He has developed several close friends in the hotel and he occasionally joins a group playing pinochle. He has been keeping steady company with a girl whom he expects to marry. They plan to move to a furnished apartment. Due to her influence he spends several evenings a week at the movies and dances, or with their friends.

"About a month ago he quit spending his money on doctors. He takes no medicine and says he feels better than he ever did. He has discarded the jacket that he was continually wearing and observes that he is no longer troubled with colds. This is interesting because he said he always had a cold about this time of the year. The only remnant of the old attitude toward sickness is his excuse for marrying. He says a man of his age needs some one to take care of him.

"From a social point of view, this employee is now in working equilibrium with his environment. His attention is on the job. The supervisors appreciate his efforts; the new employees respect his knowledge of the job, and the older employees are friends instead of sympathizers. The concentration of our efforts on a personal adjustment without taking into account the other areas—supervisor, associates, and the job itself—would have probably resulted in failure. The total adjustment depended upon work with the employee's attitude, the supervisor's attitude, his associates' attitudes, and with assistance in studying the job and developing an appreciation of coordination of activities."

The industrial counselor can be helpful in increasing morale because he is in a position to promote group cooperation and to help the person see the goal, his advancement toward it, and the meaning of his work in relation to achieving it. The only trouble with this method is that it does not go far enough; it may be merely a substitute for the employee's self-expression rather than the expression itself.

EMPLOYEE PROBLEM APPROACH

The fourth and most promising approach in increasing morale is the employee problem approach. It has been used more widely than the other three and has many advantages over them. However, it requires skillful handling and for that reason many people are unable to use it. Since it attempts to put democracy to work, this method deserves serious consideration.

In the problem approach, which is basically a form of role playing, the foreman, the boss, or best of all a trained psychologist (who does not play the role of expert)

presents a currently pressing problem to the employees. Then, with a minimum of interference but some guidance, he allows them to work it out for themselves. Obviously, group cooperation is a natural by-product of this method, and this in itself constitutes an important factor in increasing morale. In working out the problem the employees establish a goal: the solution. They see the progress they are making, and they actively and meaningfully participate in the effort to solve the problem. Hence all the determiners of morale are present, and increased morale is automatically achieved along with the solution of the specific problem.

In Chapter 8, the several different types of human relations training methods were discussed in considerable detail. Nearly all of these training methods, such as role playing, sensitivity training, and even management games are potentially useful ways of handling morale via the employee problem approach. The types of problems that can be used are almost limitless as long as they concern the specific group. Work hours, production increases, vacation schedules, and overtime work, as well as the many irritations which arise from day to day, are all suitable for this technique.

In the employee problem technique using a sensitivity training approach, a group meeting is called and the problem is briefly presented by the discussion leader, who may be an outsider, an employee, or the employer. Whoever he is, it is important that he be trained in the art of conducting a democratic meeting. He must encourage free discussion and he must be in a position to assure the group that the management will respect its final recommendation. In the first meeting the group is usually silent for a short time, but someone eventually speaks up and then there is bedlam. Within a few moments there will be shouts for order and one person will manage to be heard above the others and thus gain the group's attention. One view is discussed after another, and finally someone suggests that a vote be taken. It is important that no pressure be put on the group to reach a decision and that the leader handle the discussion in such a way that no one will get the impression that he is steering the group toward a certain decision. It is not necessary for these group meetings to be held regularly. One or two discussions on a specific problem are usually sufficient. When the group has reached its decision, it will stick to it and, what is more, feel proud of it and of its members.

Three brief examples of this technique will be given. Students often purchase a number of books which they leave in the library for their exclusive use. Soon some student complains that he cannot get his book because someone has removed it from the library. At this point, the instructor calls this to the attention of the class and asks what, if anything, should be done. There is the customary short silence; then many students speak at once and no one can be heard. Within a short while they recognize this and either calm down or ask the instructor to act as chairman. He calls upon one student to present his view. If his ideas are unpopular, others make suggestions. The interesting point is that semester after semester each group finally draws up a code of conduct. The usual decision is that all books remain in the library except over weekends, when all but two of them may be taken out from Friday afternoon until Monday morning. Do the students ever break these rules? Of course not. To do so would mean breaking rules they themselves formulated and relegating themselves to the position of social outcasts. Incidentally, the library has its own rules regarding fines on books kept out after the due date, and the students willingly pay the fines.

A high school principal who took a seminar which one writer conducted and in which the democratic principles governing group behavior were used was very sure

that this method would not work. After about six weeks she came to a session and announced, "I'm tired, but I found out that democracy works." Urged to explain, she went on (Blum and Selltiz, 1946, p. 426):

> I've just conducted a faculty meeting democratically. It was exhausting for me—but it worked. You see, we've always had trouble about advisers for afterschool clubs. The teachers always say they're too busy and they refuse to take on the job. This year they insisted there were too many clubs, and they wanted me as principal to decide which clubs should be eliminated. I turned it back to them, and asked them to discuss the values of each club and then decide which ones should be discontinued. As they talked, they decided every single club was important. They voted to keep every one, and each club got an adviser. I learned something; being democratic is not only a good theory—it brings results.

Although these two illustrations have been drawn from groups with above-average intelligence, the method works equally well with people of average intelligence. The success of this method is not a function of the group's intelligence. For example, the employees of a factory that manufactured curtains complained to the boss that the rate set on a particular item was too low, and they threatened to quit if it were not raised. Actually, the rate was fair and was equal to the previous average earnings, and the boss *believed* that he could not increase it. Using the employee problem method, he called the group together and presented these facts. He also told them that he was concerned about the problem and wanted to see it worked out fairly for everyone. First came the suggestion to grant the increase, whereupon he told them that in his opinion the time study had set the rate fairly. Then an employee suggested that they time the operation. The boss had nothing to lose by doing this, because the rate had been set honestly in the first place. The group then arranged to have some employees work on a few bundles of curtains and to have other employees do the timing. The necessary simple calculations were made and the facts presented to the group. They decided that even though they did not like the "number" they would do the job since their earnings would not be reduced. The employees felt better because they knew they were not being cheated, and the boss not only prevented employee turnover and a money loss but gained increased respect. He is very different from the authoritarian, "do it—or else" type of employer.

An extension of the employee problem approach naturally leads into the potent and meaningful concept of group dynamics in industry since a realistic study of human relations in industry demands study not only of the individual—either employer or employee—but of the group or groups that individuals form. Such awareness encourages the view that the solution to many industrial problems involving employees, supervisors, and employers lies in a better understanding of the group dynamics involved. Accordingly, the hypothesis is advanced that a solution to industrial conflict lies in the unraveling of the psychological dynamisms involved in the interrelationships of individuals on the job.

The organization of business, both big and small, demands that people work together in teams. In some few instances the team approach is recognized and even labeled. More often, men with different personalities represent different aspects of the same business and regard each other as obstacles. They do not perceive the other man's needs except from their own frame of reference. Accordingly, sales, engineering, research, accounting, and management regard each other as threats or interferers, are too often not even consciously aware of their hostile reactions to others, and seldom think of their work as part of a team effort.

The group dynamics approach attempts to solve this problem by creating group atmosphere and thereby allowing the various members to better understand one another's problems and needs. It results in changes in behavior and, when successful, creates the team approach. For the present, there is no one best way to achieve this, and many diversified approaches find equally enthusiastic support in varied quarters.

We shall discuss the group dynamics approach in considerable detail in the later chapter on leadership, since it has been within this context that the area has been most dynamic. It will become clearer at that time how closely tied are the areas of leadership and morale.

GROUPS AND GROUP DYNAMICS

In an industrial situation, as in any other, people form groups. As soon as one studies such groups it is obvious that:

1. The structures of these groups differ.
2. The groups are rarely distinct; rather, they overlap.
3. Certain individuals (not necessarily known to management) determine the group behavior.
4. The behavior of certain individuals is determined by group behavior.
5. Individuals in a group contribute to group structure and behavior in different degrees.
6. The group generates a climate that transcends individual feeling.

Further, it must be recognized that predictions of group attitudes and behavior are possible. As a result of encouraging group formations or working with those already formed, it has been found possible to:

1. Change the purpose of groups.
2. Change the structure of groups.
3. Change the personality of individuals in a group.
4. Promote an understanding of the motives of others in the group.
5. Improve morale.
6. Solve problems by reducing conflict

VERTICAL ROUND TABLE

One technique that serves the purpose of harnessing group dynamics may be called the vertical round table.[4] In a particular company about 30 sessions were held at a round table with twelve representatives of different levels of management. A round table was deliberately used to avoid seating according to status; and by having different levels participate, a vertical view of management was obtained. Meetings were held as luncheon sessions away from the plant, and no one was expected to report back to the plant that day. It was intended to have these sessions for frank talk and a good time: a bar and meals helped achieve the latter. Three topics were proposed for discussion: (1) human nature and what little we know about it; (2) the present American industrial scene and its problems; and (3) the problems at X Company, especially bridging the gap between management and the workers.

[4] The term *vertical* refers to the status of the executives participating in the round table.

The order of presentation of the topics was important, and developments have shown that, by not discussing company problems immediately, the difficulty of status consciousness and its interferences with frank expression was, to a large extent, reduced.

At the start, the six upper-level men grouped themselves and talked much, the six lower-level men (none of the twelve were on the same level) talked little. As the series progressed the groupings disappeared, talk was contributed more evenly, and the inhibitions of status disappeared.

Analysis of discussion content revealed the following changes:

Problems	*Period of Discussion*		
	Early	*Middle*	*Late*
Human nature	60%	30%	10%
Industrial problems	30	50	20
Company problems	10	20	70

Another shift was from fun (storytelling) to conflict and catharsis to discussion and listening to problem solving. Attitude scale measurements were used. Shifts took place and the evidence is that the liberals became slightly more conservative and the conservatives became a great deal more liberal.

It is now many years since the vertical round table was introduced in this company. At present, many vertical round tables are in existence in *various* divisions of the company. Specific changes and improvements have been effected, and in addition, as a result of the entire plan, a uniform statement of policy concerning employee relations has been evolved.

INVERTED PYRAMID

In another company, a program of testing and counseling executives evolved the inverted pyramid technique, thereby continuing to distort geometry. Working with individuals brought to light considerable rivalry, with many striving to outdo each other to gain top-management favor. The result was wasteful competition, personal rivalries, and little awareness of the need for cooperation. Each claimed he was willing to cooperate, but it was always the others who would not. The inverted pyramid technique resulted in a group conference, but only after interviews had been held with individuals, then pairs, then quartets, etc.

Briefly, this method operates as follows. An interview with Allen revealed that general conditions in the supervisory group are getting very bad; backbiting is becoming worse and a change is needed. He expressed loyalty to the company but concern about Baker, who is very erratic. He believes that Cox is close to Baker, and for that reason he feels he cannot talk to him because he carries tales. However, he likes and respects Cox. It turns out that Cox believes the company is running very smoothly. Each is willing to talk things over, however, and so a paired meeting is arranged. At this session Cox expresses admiration for Allen's ability to get things done but is concerned about the rough treatment Allen gives his supervisors. After citing a specific case, Allen responds by indicating he is afraid that if he does not do it that way he will be blamed for any shortcomings by Baker. He further

states his respect for Cox but fears that he tells everything to Baker. Getting down to specifics, Allen only reports suspicion and, in turn, Cox illustrates how he really stands up to Baker. They then become involved in a discussion of the confusion caused by lack of clear channels and the crossing of those that exist. They agree that Doe, the big boss, should be informed and other executives called in for such meetings, that is if they so desire. And so it goes, with Allen meeting Baker, etc.

This series of meetings of the consultant with individuals, pairs, and groups does seem to be more than just talk. Better understandings have been worked through, and strong evidence of constructive group performance is emerging.

CIRCLE WITH A POINT

To further belabor our perverse geometry, we may call the third example the circle with a point. It occurs in a much smaller company than those previously described. For example: The employer set up a human relations program which deliberately attempted to provide opportunities for people to express, be aware of, and understand feelings and attitudes that interfere with their effective functioning. Its main purpose was to improve interpersonal relations and improve the "we" spirit of the company. All key personnel were invited to the sessions. Actually, the conferences consisted of twelve individuals, plus the employer, and the group leader who was the consulting psychologist. The discussions were entirely free. There was never an agenda. Any topic that anyone cared to present was discussed. Actually, discussions centered around plant problems· problems which group members had in their work situation, interpersonal relations in the plant, and some personal problems.

The group leader actively led the group to analyze their participation in terms of the attitudes and feelings expressed. He never attempted to constrain members, even in moments of intense heat. His role was that of supporting some person when he believed it necessary, of drawing out members who were rather reluctant to participate, and of reflecting feelings in order to continue discussion leading to analysis and understanding.

As a result of weekly sessions for sixteen months, many things happened to the individuals as well as the group. Tex, a highly vocal, excitable person, a strong defender of the "rank and file," became a person who was now able to speak for *himself* as well as the "rank and file." He became more relaxed and secure and a team member. Cal, a suspicious and hostile person with a large "chip on his shoulder," became cooperative and full of suggestions. He was willing to be critical of management and willingly volunteered to take on unusually heavy assignments. Monte originally was deferential to authority figures but later was able to challenge them if he felt his stand was justified. Phil, a timid and almost nonparticipating member in the early sessions, became rather active and vocal in the later meetings. Needless to report, such personality changes did not occur in all the participants.

Fear and hostility toward the "boss" tended to diminish during the course of the sessions as the participants had the chance to express hostile feelings without punishment, to see that the "boss" was not the "sacred cow" and that he too had problems, and further, that he did not have magical solutions. During the sessions open expressions revealing anxiety, guilt, rejection, and insecurity, among others, came to the surface; and as these feelings were to some degree relieved, interpersonal relations improved. In turn, these led to the solution of operating and personnel problems, as well as the formulation of acceptable company policies.

A COMPARISON OF METHODS FOR INCREASING MORALE

One small but nonetheless serious problem which affects morale concerns the adjustment of windows. This is not only a time-consuming affair in almost any office or factory, but one that leads to almost continual friction and lowered morale because it breaks the employees up into at least two groups—the "window open" or "we need air" group and the "window closed" or "we'll get sick" group. If the supervisor refuses to settle the question arbitrarily, there is likely to be a continuous procession to and from the windows. Let us see how it can be solved by each of the four methods of increasing morale.

In the first method an *expert* is called in. He makes a study of the plant and comes to the conclusion that fans or air conditioning, or both, will solve the problem. The boss may reject this recommendation because of the outlay of capital required, or he may install an air-conditioning system at considerable expense only to find that some employees still want the windows opened because they need fresh air. The second method employs a *spy*. Unknown to the employees, he observes what goes on and he makes frequent trips to the lavatory so that he will have a full set of notes. He finally spots the person who is causing all the trouble and those who are on his side. The spy may suggest that someone be fired or he may present an "ideal" seating arrangement for the group. But even if the ringleader is moved away from the window, there is always someone else who will take the lead in the fight, so the problem remains unsettled. Furthermore, the "ideal" seating arrangement does not work out because it is not likely to take status and sociometric principles into consideration. With the third method, an *industrial counselor* interviews a number of employees. He finds that one employee was scared by an open window when he was two years old, whereupon he prescribes therapy that enables that individual to be a happier and better adjusted person.

The fourth method, the *employee problem approach*, affords the best means of solving the problem. The boss or supervisor calls the group together and says: "We have quite a problem on our hands with the continual opening and closing of windows. Now let's not act like children; let's try to work it out." There is some laughter and then the employees start making suggestions such as: "Let's have a schedule and keep the window open for 15 minutes every two hours." "Let those who don't mind open windows sit near them." When one of the group says, "We need air conditioning," the boss tells them that he has already investigated this possibility and that it will cost more than the budget allows. The group may then make other suggestions or decide to try one of the first two. If the discussion is not hurried and no pressure is brought to bear, the employees are likely to come to an amicable decision, such as trying the first suggestion for a month and then calling another meeting if there are still objections. The point is that the group makes the decision of its own free will, and its members will be reluctant to break the rules because they may be considered poor sports. Although some of them may change their seats, there will be a marked decline in raising and lowering the windows.

Another petty morale problem that often arises, especially among women in offices, is the "body-odor" problem. Someone accuses one girl of having "B.O.," and whether she has it or not, the other girls will refuse to work with her and insist that she be fired. Obviously, a group discussion of this problem as it applies to one girl is inadvisable and should never be used. Here is a tender but tense situation which is best

handled by the industrial counselor. If the accusation is unfounded—as it often is—he tells the girl about the unfair attack and tries to help her become accepted by the group. He may even have to speak to the leader of the persecuting group about the unfairness and the prejudice involved. On the other hand, the girl may have a distinct body odor. If this is a medical problem (the intense and constant cases sometimes are), he should refer her to a physician, who will treat the condition after determining its causes. Other cases of "B.O." involve the problem of body hygiene, and once a person is aware of her susceptibility she is usually willing to take the necessary precautions. The reason for mentioning this type of problem is to emphasize the fact that in handling individual problems of a personal nature the group technique should *not* be used, for it can result, psychologically, in more harm than good to the individual. The group approach to a problem related to the group not only solves the problem but also provides a promising means of increasing or bolstering morale.

REFERENCES

Bernberg, R. E. Socio-psychological factors in industrial morale I. The prediction of specific indicators *Journal of Social Psychology*, 1952, 36, 73–82

Blum, M. L., and C. Selltiz. The seminar as a method of in-service training. *Journal of Educational Sociology*, 1946, 19, 420–429.

Giese, W. J., and H. W. Ruter. An objective analysis of morale. *Journal of Applied Psychology*, 1949, 33, 421–427.

Gordon, O. J. A factor analysis of human needs and industrial morale. *Personnel Psychology*, 1955, 8, 1–18.

Guion, R. M. Industrial morale: The problem of terminology. *Personnel Psychology*, 1958, 11, 59–61.

Jenkins, J. G. The nominating technique: Its uses and limitations Paper presented at Eastern Psychological Association meeting, April 1947.

Lewin, K., and G. Watson (eds). *Civilian morale*. Houghton Mifflin, Boston, 1942.

Marrow, A. J. Cited in K. Lewin and G. Watson (eds.). *Civilian morale*. Houghton Mifflin, Boston, 1942.

Moreno, J. L. Foundations of sociometry. *Sociometry Monographs*, No. 4, Beacon Press, Boston, 1943.

National Research Council. *The Fatigue of workers*. Reinhold, New York, 1941.

Roethlisberger, F. J. *Management and morale*. Harvard University Press, Cambridge, Mass., 1946.

Stagner, R. Motivational aspects of industrial morale. *Personnel Psychology*, 1958, 11, 64–70.

Van Zelst, R. H. Sociometrically selected work teams increase production. *Personnel Psychology*, 1952, 5, 175–185.

Yuzuk, R. P. *The Assessment of employee morale*. Bureau of Business Research Monograph No. 99, Ohio State University, Columbus, 1961.

14 LEADERSHIP AND SUPERVISION

Another type of problem to whose solution the industrial psychologist can contribute is leadership. The role of leadership in industrial relations is gaining increasing recognition. Just as society looks for a leader to define its purpose and lead it forward, so both management and labor have been concerned in the selection and development of men who can successfully attack the many perplexing problems that confront them. The psychological aspects of effective leadership and of the training of leaders deserve serious thought.

We all know that leadership is necessary in any group. In the industrial setting either the employer or the employees will choose the leader. In the ideal situation both select the same individual. However, this does not often happen; the leader who is selected and the one who emerges from the group will have very different characteristics. If the leader is handpicked by management, his loyalties will be predominantly to management and his characteristics may be such as to encourage conflict rather than cooperation. By the same token, if the group chooses a leader, his loyalties may lie with the group, and the result may be conflict rather than cooperation. The militant labor leader is undoubtedly a product of the conflict between management and labor; when these two groups learn to cooperate, such a person will no longer be a successful leader.

Effective selection depends upon management's understanding of leadership and its ability to evaluate men. For management's purposes leadership may be classified at three levels: top management, middle management, and front-line management. Top management usually means the "big boss"; middle management is just called the "boss"; front-line management includes foremen and supervisors, who are usually referred to as "boss men" or "lead men." These three types of leaders operate on different levels and have different responsibilities and duties. Nevertheless, all three face basically the same problem—how to deal with and lead people.

A leader who loses contact with his group is no leader; hence it is important for any executive who serves as a leader to remember that he must maintain close contact with his group if he is to function as a leader. It is also important that he be able to delegate authority. Many leaders fail not because of limitations on their own general ability but because of their inability to delegate authority. Leadership is not a one-way affair; it involves interaction of the leader and the group.

Barnard (1938) says that executive work has three aspects: communication, cooperation, and purpose. The executive is responsible for devising a system of com-

munication which will allow the organization to function smoothly. Since the executive is always concerned with securing services from individuals, he must show a strong willingness to cooperate with people. The executives formulate the general purposes of the organization, and they must indoctrinate their subordinates with these general purposes so that the organization will function in a unified manner.

LEADERSHIP AND POLITICAL ATMOSPHERE

Possibly one of the great paradoxes of our times is represented by the leader who extols and defends the virtues of democracy by authoritarian means. Although this book is written in the context of industry, it recognizes that leadership patterns, whether they are demonstrated in the home, in government, or on the job, can run the gamut from the "little dictator" to democracy. Further, the label by which the leader identifies himself often merely confuses because his actions seem to be unrelated to his professed or stated goals. Many self-appointed defenders of freedom of the little man, the worker, etc., seem more to restrict freedom than to defend it. Although the reader may be thinking of political implications, the authors are more concerned with leadership patterns in industry. Yet they will not deny their bias for democratic leadership in the home, schools, government, and industry.

As we will see more clearly in later sections, it is easy to relate "leadership" to the kind of atmosphere that the supervisor establishes in his work group. For example, Bradford and Lippitt (1945) refer to four types of supervision: hard-boiled autocrat, benevolent autocrat, laissez faire, and democratic. For each of these types they postulate the characteristics, group reactions, and group personalities involved. Inspection of the differences due to kinds of leadership is fruitful.

The *hard-boiled autocrat* constantly checks on production, gives orders, and expects immediate acceptance. He is a rigid disciplinarian and believes that praise will spoil. He is status minded and does not trust the employees' initiative. The group reaction is submissive but resentful. Individual responsibility is at a minimum, but buck passing and backbiting are common. The group tends to be insecure, tense, aggressive, and egocentric.

The *benevolent autocrat* dominates all employees and is the source of all standards. Failure to meet these standards makes him feel hurt, angry, or surprised and is interpreted as a personal disloyalty. The benevolent autocrat rarely recognizes his autocracy. The group is submissive and lethargic and shows no initiative without checking with the supervisor. Only those who see through him dislike him intensely. In this group the employees are dependents who demonstrate a slow regression to more submission, dependency, and inability to accept responsibility.

Laissez-faire leaders often busy themselves in paper work and so stay away from employees. Such a leader sets no goals, makes no decisions, and believes he is the "good fellow." The group has scapegoating, instability, and a sloppy, low output. Frustration, failure, and insecurity are typical in this directionless group.

Democratic supervision shares group decisions, gives reasons for decisions, and devotes time to planning. Enthusiasm is high and basic needs tend to be satisfied. Confidence and security are typical.

According to Bradford and Lippitt, democratic leaders encourage employees to work in a democratic framework. They recognize that they play a role as a leader which fosters individuals. They recognize that:

1. Having problems is permitted.
2. Meetings are necessary for group thinking and action.
3. Group goals should be definite.
4. Performance standards must be accepted and determined.
5. Reasons for decisions must be understood.
6. Progressive growth results in independence and responsibility.

One might immediately conclude on the basis of the above descriptions that the goal should always be toward democratic leadership and away from autocratic leadership. However, it must be remembered that democratic leadership is a more difficult form, possibly because varieties of autocratic leadership tend to be more prevalent. Further, not all people understand democratic leadership. Some definitely believe it is a sign of weakness in the leader. Others simply cannot be anything else but dependent and submissive. They only want to be told to do things and assume that what they are told is easier and right.

LEADERSHIP CRITERIA

Probably the best place to begin our examination of the leadership concept is by examining the criterion problem as it applies here. Ask yourself "What is a successful leader? How can I tell that leader X is terrific while leader Z is a dud?" You would probably be able to make a list of things which are clues to whether or not a leader is being effective. For example, one might list the following:

1. His men respect him.
2. His men follow his orders without question.
3. His men like him.
4. His work group has high morale.
5. He looks out for his men.

Examination of these statements should serve to point out two rather striking difficulties inherent in leadership criterion development. First, there is a wide variety of different ways of defining success as a leader. Not all people will agree on which ways are appropriate, either in general or in any particular situation. Second, even if people can agree on which clues to use, there still remains the difficulty in measuring each leader on each characteristic. It is easy enough to say that one essential criterion for determining good leadership is whether or not his men follow his orders without question. But how does the psychologist go about measuring the degree to which that is true or not with a number of supervisors?

Objective measures of leadership are extremely difficult to find. One that is sometimes used is team or work group productivity. Whether this is legitimate or not is subject to question, since it is necessary to assume that productivity is a by-product or result of good leadership. Some of the research to be examined later would indicate this may *not* be a particularly good premise.

A much more frequently used measure of leadership effectiveness is the rating technique. Because of the difficulty in acquiring any objective measure, one is often forced into this method of leadership evaluation. Such a criterion is, of course, subject to the criticisms detailed in Chapter 6.

TWO APPROACHES TO THE STUDY OF LEADERSHIP

Historically, research on leadership can be classed into two approaches. These approaches differ in their philosophy concerning the best way of studying leadership. We shall call them the "trait" approach and the "behavior" approach.

TRAIT APPROACH

Most of the early leadership research has been of the trait approach type. The philosophy of the trait approach is quite simple and would seem, at least initially, quite logical. *Successful leaders are assumed to possess more (or less) of certain traits than are unsuccessful leaders.* This is the notion of the ideal leader, or the ideal leader profile which describes which traits a good leader should possess a lot of and which traits he should possess in only minimum amounts. The emphasis in this approach is on the personal characteristics of good and bad leaders. It says that the best way to investigate leadership is to look at good and bad leaders and see how they differ in terms of their traits. The usual method followed in such studies is to:

1. Identify a group of "good" leaders and a corresponding group of poor or "bad" leaders.
2. Measure these leaders on a variety of personality traits and individual characteristics.
3. Determine if the good leaders possess a significantly different amount of any trait or characteristic than do the poor leaders. If so, this trait is defined as a critical leadership trait.
4. Cross-validate results in steps 1–3 on a new sample of leaders (this step is often omitted in spite of the fact it is the most critical of all).

EXAMPLE OF TRAIT APPROACH A recent study by Randle (1956) is an excellent illustration of the use of the trait method. This study involved 1427 executives from 25 different companies. It involved four phases:

Phase I. Very extensive appraisals were carried out on each executive to provide a complete dossier on each man. This was achieved by several methods (Randle, 1956, p. 64)[1]:

> (1) Each executive's background and experience were analyzed in detail. This covered age; education; professional, social, and civic activities; work experience; health; and family relationships. Such an analysis indicated what areas of competency the executive possessed (both actual and potential), his past progress and recognition record, his leadership record, and his adjustments.
>
> (2) A thorough appraisal of each executive was made by five of his business associates who were best qualified for the task. This was done on an independent, noncollaborative basis, and covered job performance, human relations skills, mental attributes, and personal characteristics.

[1] The material on Phases I-IV is adapted with the permission of the *Harvard Business Review* from C. Wilson Randle's "How to identify promotable executives. © 1956 by the President and Fellows of Harvard College; all rights reserved.

(3) A battery of written tests was given each executive, covering mental ability, interests, and personality characteristics. Tests were kept to a supporting rather than a primary role. They were employed as a diagnostic tool to indicate what areas needed further exploration, as an aid in explaining other findings, and as further confirming evidence of executive characteristics.

(4) Each executive was given a thorough interview lasting from one and one-half to three hours. This interview was conducted only after a complete review of all other appraisal results. It was employed to clear up areas of question, to verify other appraisal findings, and to gain firsthand impressions of the executive such as were not available from the other techniques.

Phase II. Each executive's dossier was examined to determine his "degree of promotability." They were classed into three groups as shown below:

	Number	*Percent*
Promotable	498	35
Satisfactory	770	54
Inadequate	159	11
Total	1427	100

Phase III. All appraisals were content analyzed to determine what identifiable characteristics were possessed by the 1427 men in the sample. These analyses reveal over 100 different characteristics or traits. However only 30 occurred with sufficient frequency to be regarded as "common denominator" characteristics. These 30 are given in the following list:

1. *Position performance*: How well the executive carries out the duties of his present job
2. *Intellectual ability*: Ability to solve problems, to adapt to new situations, to analyze and make judgments
3. *Human relations skill*: Ability to motivate people and get them to work together
4. *Personal characteristics*: The total of temperament or personality characteristics bearing on executive functioning
5. *Technical knowledge*: The knowledge of functional skills needed to carry out position requirements
6. *Breadth of knowledge*: Range of interests, use of information and concepts from other related fields of knowledge
7. *Planning*: Looking ahead; developing programs and work schedules
8. *Administration*: Organizing own work and that of others, delegation, follow-up, control of position activities
9. *Accomplishment*: Effective use of time; amount of work produced
10. *Quality*. Accuracy and thoroughness; high standards
11. *Dependability*. Meets schedules and deadlines; adheres to instructions and policy
12. *Acuteness*: Mentally alert; understands instructions, explanations, unusual situations and circumstances quickly
13. *Capacity*: Mental depth and breadth; reservoir of mental ability
14. *Flexibility*: Adaptable, adjusts rapidly to changing conditions; copes with the unexpected
15. *Analysis and judgment*: Critical observer; breaks problem into components, weighs and relates; arrives at sound conclusions
16. *Creativeness*: Original ideas; inquiring mind; fresh approaches to problems
17. *Verbal facility*: articulate, communicative, generally understood by persons at all levels
18. *Socialness*: Makes friends easily; works "comfortably" with others; has sincere interest in people
19. *Acceptance*: Gains confidence of others; earns respect
20. *Sensitivity*: Has a "feel" for people, recognizes their problems; quick to pick up "the way the wind is blowing"; is considerate of others

21. *Leadership*: Receives loyalty and cooperation from others, manages and motivates others to full effectiveness
22. *Developing others* Develops competent successors and replacements
23. *Motivation*. Has well-planned goals; willingly assumes greater responsibilities; realistically ambitious
24. *Attitude*: Enthusiastic, constructive, optimistic, loyal; good orientation to company, position, and associates
25. *Vision*: Has foresight, sees new opportunities, appreciates, but not bound by, tradition or custom
26. *Self-control*. Calm and poised under pressure
27. *Initiative*. Self-starting; prompt to take hold of a problem, sees and acts on new opportunities
28. *Drive* Works with energy, not easily discouraged, basic urge to get things done
29. *Self-confidence*: Assured bearing, inner security, self-reliant; takes new developments in stride
30. *Objectivity* Has an open mind, keeps emotional or personal interests from influencing decisions

Phase IV. The final phase was to determine which of these 30 characteristics were related to promotability. To be considered a predictor of promotability, a characteristic had to be either significantly more or significantly less present in those judged promotable than in those judged inadequate.

The results indicated that the "composite executive" was a person who was distinguished by eight qualities shown in Table 14.1.

TABLE 14.1 *Qualities of the "Composite Executive"*

	Percentage Present	
Characteristics	*Promotable Executives*	*Inadequate Executives*
Position performance	50	5
Drive	47	14
Intellectual ability	44	8
Leadership	41	6
Administration	40	6
Initiative	38	7
Motivation	34	8
Creativeness	30	6

SOURCE: C. Wilson Randle. How to identify promotable executives. *Harvard Business Review*, 1956, *34* (3), 126.

In addition to these general traits for successful executives, Randle also found that there was a tendency for certain traits to gain in importance as one goes up in the managerial hierarchy. This was particularly true of the *motivation* characteristic, which was judged outstanding in 84 percent of top executives considered promotable and in only 14 percent of top executives considered inadequate.

CRITICISMS OF TRAIT APPROACH There are a number of important difficulties involved in the trait method of studying leadership. First, there is the problem of defining and agreeing upon traits. The number of descriptive adjectives which can be

used to "type" people is tremendously large. For example, those who were at one time boy scouts will recall the oath to be the following:

Helpful	Cheerful
Friendly	Thrifty
Courteous	Brave
Kind	Clean
Obedient	Reverent

Each of these adjectives is a trait. Thus there are nearly as many traits of people as there are adjectives. Are they all really different? Which ones are similar? Which ones should we measure since we obviously cannot study them all?

A second difficulty exists in trying to measure traits. We have many personality tests available today, each of which lists the traits it proports to measure. Unfortunately all too often two tests which claim to measure the same trait turn out on close examination to be quite different and two other tests which ostensibly are designed to measure traits quite different from each other may turn out to have very similar contents. A common solution is to have judges rate each person on every trait, rather than giving tests. This was the procedure used in the Randle study. Since the additional problems created by the use of ratings have already been discussed, we shall not elaborate on them again.

The last objection to the trait approach is probably the most important of all. As a method it does not provide the psychologist with much insight into the basic dynamics of the leadership process.

SOME ADDITIONAL TRAIT STUDIES The above criticisms, although rather severe, do not imply that the trait approach is without value. A good deal of useful and important information has been acquired using this approach, with the work of Porter and Ghiselli at the University of California being excellent examples of such research.

We previously looked at Porter's work in the context of motivational theory (Maslow) in industry. His research also dealt with the importance of different managerial traits. In his first study Porter (1961) found that some traits tended to be judged more important at high management levels ("aggressive," "dominant," "independent," and "original") while other traits tended to decrease in importance at higher levels ("conforming," "cooperative," "flexible," and "sociable"). In a second study Porter and Henry (1964a) examined the judged importance of five other-directed type traits and five inner-directed type traits. They found that the inner-directed traits were perceived as more important at each higher level of management and that the other-directed traits were seen as being less important the higher one went in the managerial structure. A follow-up study using 1800 managers found that staff managers placed more emphasis on other-directed traits than did line managers, with managers in combined line-staff jobs falling intermediate between the other two groups (Porter and Henry, 1964b).

Inner-Directed Traits	*Other-Directed Traits*
Forceful	Cooperative
Imaginative	Adaptable
Independent	Cautious
Self-Confident	Agreeable
Decisive	Tactful

Ghiselli and Lodahl (1958) reported no success using the traits of supervisory ability and decision-making approach to predict group productivity and carefulness. However, in another study Ghiselli (1959) found all of the following traits to differentiate between high-level managers, middle-level managers, and low-level managers:

Intelligence
Supervisory ability
Initiative
Self-assurance
Occupational level

The higher one went in the managerial structure, the more pronounced the trait. The only exception was intelligence, which did not differentiate middle and top managers.

BEHAVIORAL APPROACH

Research in the field of leadership characteristics has indicated a trend. More psychologists recognize that looking for specific traits even as measured in psychological tests does not enhance the understanding of leadership qualities. From widely different sources, work and enthusiasm seem to be centered around a more broad and meaningful concept of leadership. This view regards leadership as behavioral, situational, or related to the interaction of the leader and the group.

The layman still believes that such traits as superintelligence, initiative, aggressiveness, tact, etc. identify the leader. Various disciplines less sophisticated in experimental methodology still extol the virtues of the trait concept. The simplest way to dispel this view is to name leaders in politics, the community, or the company. Then assign the traits that most typically identify these people. The result generally is complete disagreement in trait assignment from leader to leader even though there is general agreement that one leader speaks well, another is aggressive, another ruthless, and still another kindly.

The behavioral approach is also quite simple in its philosophy. It states that the best way to study and to define leadership is in terms of *what leaders do* rather than in terms of what leaders are. Thus one is concerned with leader "behaviors" rather than leader traits. In outline form, the behavior approach:

1. Collects critical incidents of good and bad leadership behavior, usually by interview procedures.
2. Scales each incident to determine, in the eyes of experts, how "good" or how "bad" each behavior is judged to be.
3. Develops a check-list type questionnaire that can then be used to check off which of these behaviors have been exhibited by each leader or supervisor. A leadership "score" can be computed by using the median scale value of the behaviors which have been checked.

OHIO STATE LEADERSHIP STUDIES The behavioral method of leadership research is best exemplified by the work done in the 1950s at the Personnel Research Board at Ohio State University by Hemphill, Fleishman, Stogdill, Shartle, and Pepinsky. A detailed description of this research is impossible here so only the highlights will be reviewed.

The early OSU research, carried out primarily by Hemphill and his coworkers, involved the development of a Leader Behavior Description Questionnaire containing 150 items which described how people in leadership positions operate in

their leadership role. These 150 items were derived from an original item "pool" of over 1800 items which had originally been classified into nine *assumed* dimensions of leadership by a group of experts. These nine assumed dimensions were:

1. Integration behaviors
2. Communication behaviors
3. Production emphasis behaviors
4. Representation behaviors
5. Fraternization behaviors
6. Organization behaviors
7. Evaluation behaviors
8. Initiation behaviors
9. Domination behaviors

TABLE 14.2 *Items Selected for the Revised Form of the Supervisory Behavior Description*

	Orthogonal Factor Loading	
	"Consideration"	*"Initiating Structure"*
"Consideration"—Revised Key		
He refuses to give in when people disagree with him.	−0.68	0.06
He does personal favors for the foremen under him.	0.40	0.06
He expresses appreciation when one of us does a good job.	0.70	0.19
He is easy to understand.	0 70	0.13
He demands more than we can do.	−0.40	−0.08
He helps his foremen with their personal problems.	0.32	0.05
He criticizes his foremen in front of others.	−0.49	0.03
He stands up for his foremen even though it makes him unpopular.	0.54	0.08
He insists that everything be done his way.	−0.52	−0.01
He sees that a foreman is rewarded for a job well done.	0.70	0.05
He rejects suggestions for changes.	−0.62	−0.06
He changes the duties of people under him without first talking it over with them.	−0.65	0.09
He treats people under him without considering their feelings.	−0.72	0.41
He tries to keep the foremen under him in good standing with those in higher authority.	0.68	0.17
He resists changes in ways of doing things.	−0.57	0.19
He "rides" the foreman who makes a mistake.	−0.61	0.37
He refuses to explain his actions.	−0.72	0.23
He acts without consulting his foremen first.	−0.73	0.01
He stresses the importance of high morale among those under him.	0.73	−0.11
He backs up his foremen in their actions.	0.62	0.16
He is slow to accept new ideas.	−0.66	0.06
He treats all his foremen as his equal.	0.66	0.28
He criticizes a specific act rather than a particular individual.	0.63	0.14
He is willing to make changes.	0.78	0.09
He makes those under him feel at ease when talking with him.	0.86	0.17
He is friendly and can be easily approached.	0.82	−0.02
He puts suggestions that are made by foremen under him into operation.	0.87	0.11
He gets the approval of his foremen on important matters before going ahead.	0.65	−0.02

TABLE 14.2 (*Continued*)

	Orthogonal Factor Loading	
	"Consideration"	*"Initiating Structure"*
"Initiating Structure"—Revised Key		
He encourages overtime work.	0.20	0.40
He tries out his new ideas.	−0.10	0.42
He rules with an iron hand.	−0.20	0.58
He criticizes poor work.	−0.18	0.59
He talks about how much should be done.	−0 20	0.60
He encourages slow-working foremen to greater effort.	0.17	0.33
He waits for his foremen to push new ideas before he does.	−0.07	−0.28
He assigns people under him to particular tasks.	0.00	0.26
He asks for sacrifices from his foremen for the good of the entire department.	0.00	0.46
He insists that his foremen follow standard ways of doing things in every detail.	0.25	0.72
He sees to it that people under him are working up to their limits.	−0.17	0.87
He offers new approaches to problems	0.36	0.72
He insists that he be informed on decisions made by foremen under him	0 13	0.51
He lets others do their work the way they think best.	−0.17	−0.33
He stresses being ahead of competing work groups	0 05	0.34
He "needles" foremen under him for greater effort.	−0.17	0.50
He decides in detail what shall be done and how it shall be done.	0.37	0.63
He emphasizes meeting of deadlines.	0.10	0.68
He asks foremen who have slow groups to get more out of their groups.	−0 22	0.40
He emphasizes the quantity of work.	0 17	0.51

SOURCE: E. A. Fleishman. The description of supervisory behavior. *Journal of Applied Psychology,* 1953, *37,* 1–6.

As Fleishman has pointed out (Fleishman, 1953a), while the resulting scores on these nine subscales turned out to be very reliable, the subscales also turned out to be very highly correlated with each other—thus they were apparently not really *different* aspects of leadership behavior. The 150 items were therefore analyzed further using factor analysis methods. The results were rather interesting. Two major distinct groupings of supervisory behavior emerged through this process. These two categories were defined as "consideration" and "initiation of structure." The items on behavior which defined these two leadership dimensions can be seen in Table 14.2, taken from Fleishman's revised scale. The consideration dimension represents the degree to which a supervisor is considerate of the feelings of those under him. It is analogous in many respects to the traditional notion of human relations in the work setting. The initiation of structure dimension refers to the degree to which the supervisor facilitates or defines group interactions toward goal attainment. His research concludes that these two concepts are independent of each other; that is, any one person may behave in a manner to demonstrate either or both of these characteristics.

In applying these concepts at four plant levels, he found people in the higher plant hierarchy inclined to less consideration and more structure. The attitudes of the foreman group on each dimension fell somewhere between what the workers expected and what their own supervisors expected, but were much more like the attitudes of their supervisors.

Fleishman concluded, "It is possible that future research will indicate that combinations of measures of such things as group characteristics, needs and expectations, leadership attitudes, behaviors and perceptions, pressures from supervisors, etc. can yield more successful predictions where ordinary testing procedures have failed in the complex field of leadership and group effectiveness."

These two primary dimensions of leadership were subsequently found to be very stable. Later in the chapter we shall examine some other research which tried to determine if a person's leadership ability (as measured by these scales) could be influenced through training.

Empirical Evaluation of OSU Dimensions Fleishman went one step further. In a well-designed study he and Harris (1962) showed that both the amount of consideration and the amount of structure exhibited by a supervisor were related to the effectiveness of their subordinates. They defined effectiveness in terms of the amount of turnover and the number of grievances exhibited by the subordinates.

Figures 14.1 through 14.5 show the relationships obtained by Fleishman and Harris. The most interesting is the trend shown in Figure 14.3 relating both leadership dimensions to grievance rate. Note that for supervisors low on consideration the amount of structure in their behavior has no importance. All these supervisors had a high rate. Similarly, for supervisors high on considerations, structure also does not seem to be particularly important. All these foremen had low grievance rates. For supervisors of medium consideration, however, structure becomes very important. Low-structure foremen in this group had few grievances while high-structure foremen had many grievances. This would indicate that consideration, at its extreme values, is a more "dominant" leadership trait, and that only when one is dealing with supervisors of average consideration can structure have an effect.

MICHIGAN LEADERSHIP STUDIES The University of Michigan also had an effective and productive program in leadership behavior, carried out by people such as Likert, Katz, Maccoby, Kahn, and Seashore at The Survey Research Center. The initial study in their series, done by Katz, Maccoby, and Morse (1950), is now something of a classic and illustrates the early Michigan approach to leadership quite well.

The study was done in the home office of the Prudential Insurance Company, using clerical workers and their supervisors. Twelve pairs of sections were studied, with each pair of sections matched as to the kind of people, the number of people, and the type of work involved. One section in each pair was a *high productivity section* and the other a *low productivity section* as determined from prior work records. Heads of the sections were then compared on a number of variables to see if there were any significant differences between them.

There were no differences between the *high* and *low* supervisors in terms of such demographic variables as age, sex, marital status, etc. However, supervisors in charge of high-producing sections were found to be *employee-centered* in terms of their attitudes; supervisors in charge of low-producing sections were found to be *production-centered* in their orientation. (Note the similarity of these attitudes to the

Figure 14.1. Relation between consideration and grievance rate.

Figure 14.2. Relation between structure and grievance rate.

Figure 14.3. Combinations of consideration and structure related to grievance rate.

Figure 14.4. Relation between consideration and turnover rate.

Figure 14.5. Relation between structure and turnover rate.

(From E. A. Fleishman and E. F. Harris. Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*, 1962, *15*, 43–56.)

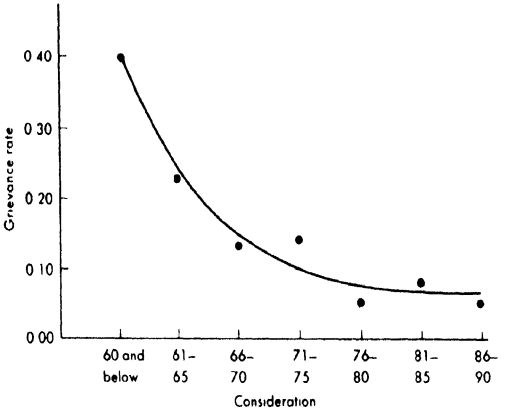


Figure 1

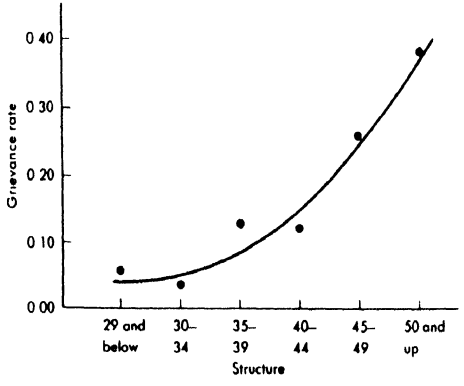


Figure 2

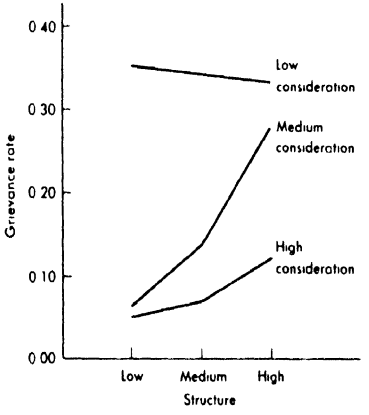


Figure 3

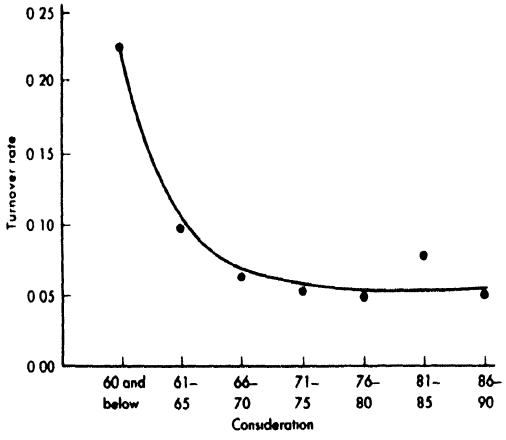


Figure 4

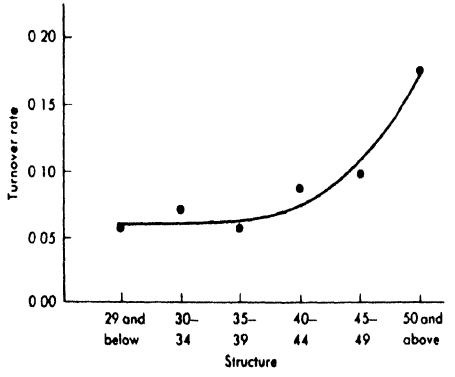


Figure 5

consideration and structure dimensions found in the Ohio State Studies.) This pattern is illustrated in Table 14.3. Clearly, the attitude of the supervisor appeared to be related to the productivity of the work group.

TABLE 14.3 *Relation of Section Heads' Attitudes Toward Employees to Section Productivity*

	Employee-centered	*Production-centered*	*Not Ascertained*	*Total*
Heads of high-producing sections	6	1	5	12
Heads of low-producing sections	3	7	2	12

SOURCE: M. S. Viteles. *Motivation and morale in industry.* W. W. Norton, New York, 1953, 152. After *Productivity, supervision and morale in an office situation.* Survey Research Center, University of Michigan, Ann Arbor, 1950.

While the University of Michigan group (Katz and Kahn, 1951) has tended to be concerned with determining characteristic differences of high- and low-producing supervisors, they have not been interested in trait characteristics as such. Instead, the emphasis has been on differences in the human relations and supervisory climate characteristics. They have found that more time spent in planning, greater degree of delegation of authority, employee oriented rather than production oriented, and the attitude of group pride are four factors related to effective supervision. This is made quite apparent by summarizing the characteristics which they found differentiated high productivity work groups from low productivity work groups. The high productivity groups:

1. Are under less close supervision from their own supervisors
2. Place less direct emphasis upon production as the goal
3. Encourage employee participation in the making of decisions
4. Are more employee centered
5. Spend more of their time in supervision and less in straight production work
6. Have a greater feeling of confidence in their supervisory roles
7. Feel that they know where they stand with the company

The common denominator of this list would appear to be behavioral, situational, and group-leader interaction rather than personal trait characteristics.

No attempt is made to add such concepts as responsibility level, level of authority, delegation of authority, goal and achievement index, consideration, initiating structure, perceptual flexibility, sensitivity, planning time, employee orientation, or group pride and reach a total executive pattern. What is intended is to emphasize that none of these concepts are "inherent" personality traits residing solely in an individual. Rather, leadership characteristics are best regarded as behavioral, situational, and the interaction of the leader and the group. This naturally leads to the view that training people is more appropriate than looking for "born" leaders.

THEORIES OF LEADERSHIP

Actually it is probably not appropriate to talk about "theories" of leadership behavior since there is little in the literature on leadership that is really of sufficient elegance to qualify for the title. However, it is convenient to stretch a point and to consider some of the different conceptual positions regarding leadership in our modern day society.

Historically, the concept of supervision has had two rather distinct "phases" which differ considerably in their philosophy. The first phase might be called the *scientific management* phase, and the second is probably best labeled the *human relations* phase. Scientific management in the early 1900s was really established and exemplified by Frederick W. Taylor. The leader or manager in this system was perceived exclusively as an individual whose sole purpose was to expedite the goals of the organization. Management was completely impersonal in that the interaction of the leader with his subordinates was ignored and the notion that subordinate attitudes or goals might have some relevance to the work situation was not considered. As Bennis (1966, p. 67) puts it, the philosophy of scientific management was that "The only road to efficiency and productivity was to surrender man's needs to the service of a bloodless machine."

The Hawthorne Studies marked the beginning of the end of scientific management. Indeed, the acceptance of the human relations notion of leadership has sometimes been carried to the extreme opposite of Taylorism. That is, if Taylorism implied organizations without considering the human element, then the extreme human relations notion has tended to consider people as if they existed free of organizational constraints. Both positions are quite unrealistic: any meaningful model of leadership must certainly deal not only with the human element but must also be willing to admit that the formal structure of the organization and its objectives are critically important variables.

It is far beyond the scope of this chapter to prevent any comprehensive treatment of the variety of leadership models which various individuals have proposed. However, we shall mention some of the more outstanding people with a brief description of the "essence" of their frame of reference in viewing the leadership problem. All of these positions can be included under the category of human relations theories or models, although they can in turn be categorized as being either fundamentally empirical or humanistic in nature.

EMPIRICAL THEORIES

The first group of leadership theorists are the empiricists. These men have tended to develop their theories of leadership upon a foundation of empirical research evidence. Generally they have (1) examined a wide variety of psychological research dealing with interpersonal and group behavior, (2) developed principles of leader behavior based upon these research findings, and (3) tested their theories or principles in subsequent research investigations of their own design. Their approach has been more deductive than inductive in its nature, although both components have certainly been involved. The men most noted for using this approach are Stogdill, Bass, Likert, Fiedler, and Blake.

STOGDILL In his outstanding book *Individual Behavior and Group Achievement* (1959) Stogdill emphasizes the critical importance of *roles* in the phenomenon of leadership. He lists three factors as being important in shaping the expectations that define an individual's role in a group. First is the status and function of the position occupied by the individual. Second are the demands made upon the individual by his group due to variations in the group's structural and operational requirements. Last are the perceptions of the group members concerning the kind of person the individual is. The leadership role can emerge out of any or all of these three contributing factors.

A second essential element in Stogdill's view is that the basic process of emerging leadership is the process of *role differentiation*. Thus the roles of the group member (including that of the leader) tend to become defined over time as a function of the three factors linked above. Differentiation of role typically proceeds along two dimensions. *Functional differentiation* is determined by such things as the nature of the task and differences in task ability. *Status differentiation* is related to task ability and personality variables. In Stogdill's model reinforcement and learning principles are central to effective role differentiation among team members. He formally defines leadership as "the initiation and maintenance of structure in expectations and interactions."

BASS One of the most prolific contributors of quality research on the problem of leadership during the past decade, Bass has outlined his concept of leadership in relation to organizational factors in his text *Leadership, Psychology, and Organizational Behavior* (1960). He defines leadership as "the observed effort of one member to change another members' behavior by altering the motivation of the other members or by changing their habits." Thus the effectiveness of a leader is determined by whether there is any change in the observed behavior of the followers. This observed change results from motivational changes or from the leader initiating structure into the work situation. Motivation is changed by classical reinforcement principles (reward and punishment). In a formal organization, leaders use such things as promise of reward, support, consideration, etc. to motivate. Bass calls this motivational aspect of leadership a variation along a dimension of consideration. He suggests the second way of accomplishing successful leadership is via the process of initiating structure—"making others more able to overcome the obstacles thwarting goal attainment."

While Bass's view of leadership is obviously influenced by the Ohio State Studies, his development of the constructs of consideration and structure is achieved by a careful examination of leadership research far more extensive than the factor-analytic studies carried out at that university. Also, it would be an extreme oversimplification of his model to reduce leadership to these two basic elements; however, they do seem to provide the central mechanism for his view of the process.

LIKERT The Michigan approach to leadership is best exemplified in the theory of leadership proposed by Likert (1959a, 1959b). Sometimes referred to as his "modified theory of management," his model is based upon the continued series of research studies carried out over the years at Michigan. The heart of this position is that successful leadership must involve the process of *employee participation* in the structuring of his work and work environment. Democratic supervision is suggested as facilitating the involvement of the worker in the decision-making processes.

The theory emphasizes the importance of open and full communication within the group as a necessary condition to effective leadership.

FIEDLER Recently Fiedler (1964) has proposed a model for leadership effectiveness which is unique in that it (1) proposes that the qualities necessary for effectiveness vary as a function of the characteristics of the situation and (2) outlines a three-dimensional system for classifying leadership situations. He suggests that the three critical dimensions are (in order of hypothesized importance):

1. The personal relations of a leader with his group members
2. The degree of structure in the task which the group must perform
3. The power and authority inherent in the leadership position

He then states (1964, p. 164) that "the type of leader attitude required for effective group performance depends upon the degree to which the group situation is favorable or unfavorable to the leader." Figure 14.6 is a graphic representation of his model.

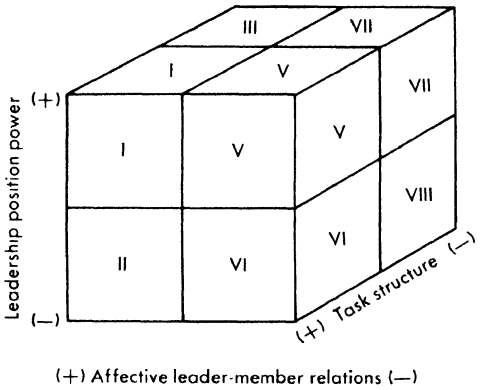


Figure 14 6. A model for the classification of group task situations. (From F. E. Fiedler. A contingency model of leadership effectiveness. In L. Berkowitz, ed. *Advances in experimental social psychology*. Academic Press, New York, 1964, 150–190.)

His research results to date appear to substantiate the notion not only that these three dimensions are critical, but that it is the particular combination or cell in block 1 which determines what kind of leader will be most effective.

BLAKE A theory of effective management known as the "Managerial Grid" has recently been outlined by Blake and Mouton (1964). Blake takes the position that the two critical dimensions of effective leadership are (1) *concern for people* and (2) *concern for production*. (Note the similarity of these dimensions to the dimensions found in the Ohio State leadership research.) According to Blake these two dimensions are independent, a manager can be high on both, low on both, high on one and low on the other, etc. Any combination is possible. His theory or model is illustrated in his 9 × 9 grid shown in Figure 14.7.

Optimal management in this model is what Blake refers to as the 9,9 manager. This type of leader is extremely concerned with both production *and* the people working with and for him. Blake proposes that managers be trained to behave in the 9,9 pattern through a two-stage training process (management development followed by organizational development) involving six phases.

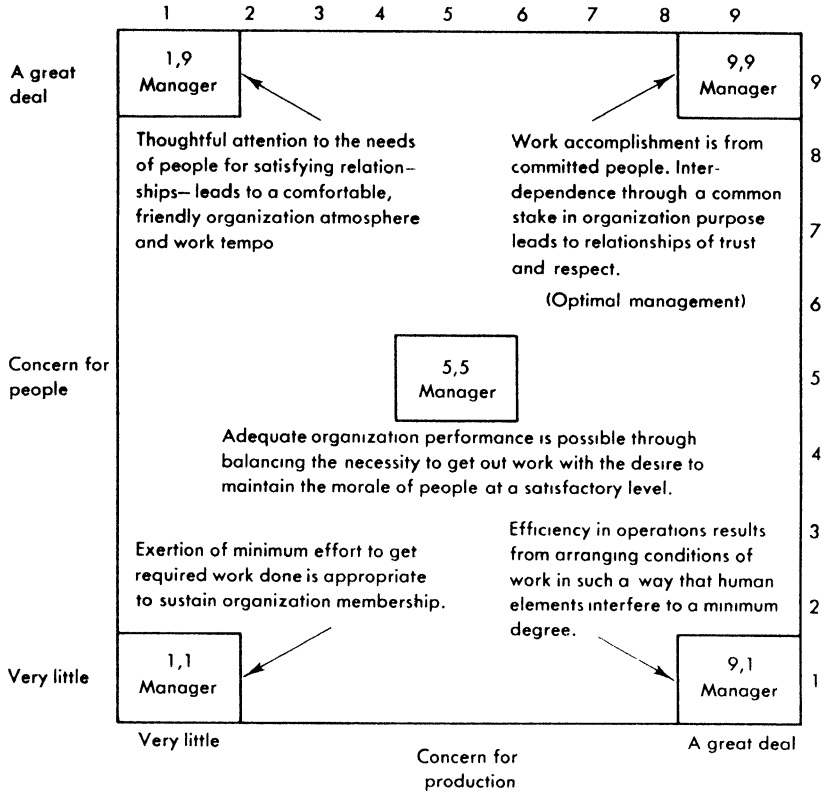


Figure 14.7. "Managerial Grid" suggested by Blake and Mouton. (Adapted from R. R. Blake *et al* Breakthrough in organization development. *Harvard Business Review*, 1964, 42 [November–December], 133–155.)

HUMANISTIC THEORIES[2]

The second grouping of leadership theorists consists of those men who have developed their models of leadership and group behavior in a much more inductive fashion. Because of this, these theories are as yet generally lacking in factual support in the form of research results. Thus at this point in time it is more difficult to evaluate the promise of these theories on any objective basis than it is to evaluate the more empirically based theories in the first group.

MCMURRAY McMurray (1950) has presented a model of leadership which moves away from the human relations notion so prevalent today and toward the old notions of subverting the individual to the goals of the organization. He argues for leaders who are "benevolent autocrats." His premise is that democratic procedures are just not practical or possible in most bureaucratic organizations. Also, he does not feel that one can ever really convince top management that democratic leadership is the way to run a business. Thus, one solves the human relations aspect of leader-

[2] For an excellent discussion of these in greater detail see W. G. Bennis (1966).

ship (the consideration dimension) by having the leader be benevolent. One solves the structure aspect by making him an autocrat.

ARGYRIS A completely different view of leadership is developed by Argyris in his book *Personality and Organization* (1957). He suggests that the needs of the individual and his goals are not compatible with the needs and goals of the organization. Thus, the organization will often make unreasonable demands on the individual. To Argyris, ideal leadership would be that which fosters the development and growth of the individual. He relies heavily upon Maslow's notion of a need structure and would have us believe that the best leadership is that which permits the worker to self-actualize. He suggests three steps to enhance the work environment:

1. Job enlargement
2. Employee-centered leadership
3. Reality-leadership

MCGREGOR A theorist with ideas very similar to those of Argyris, McGregor has presented his views in *The Human Side of Enterprise* (1960). He sets forth two basic theoretic positions which he calls "Theory X" and "Theory Y." In Theory X the emphasis in leadership is on the goals of the organization. In Theory Y the emphasis is directed toward the goals of the individual. An advocate of the Theory Y approach, he proposes that the key to the success of this approach is the concept of *integration*. This is the process of structuring the work situation in such a fashion that the goals of the organization are integrated with the goals of the individual so that members of the work group can achieve their own goals best if they follow the strategy of directing their efforts toward achieving the goals of the organization. Another important concept in effective leadership for McGregor is *collaboration*. He feels that collaboration between superior and subordinate is a necessary condition for achieving integration of goals.

TANNENBAUM The final theory of leadership which we shall mention is that proposed by Tannenbaum, Weschler, and Massarik (1961) in their book *Leadership and Organization*. Their view of effective leadership is probably the one most oriented toward individual goals and least concerned with organizational structures and goals of the enterprise. They suggest that a primary requirement for effective leadership is flexibility. Thus they define leadership as "*interpersonal influence, exercised in a situation and directed through the communication process, toward the attainment of a specified goal or goals.*" Leadership always involves attempts on the part of a *leader* (influencer) to affect (influence) the behavior of a follower (influencee) or followers in a *situation*. The construct of *perceptual flexibility* concerns a leader's sensitivity to the quality and quantity of stimuli in the environment, which then provides him with a basis for attempts at influence. A second construct, *action flexibility*, concerns the leader's repertoire of available communication behaviors, i.e., the pool of actions the leader can employ to exert influence. The model considers sensitivity as basic to successful leadership in that it will determine the leader's "psychological map" of a leadership situation. Since they attach so much importance to sensitivity in leadership, they are strong supporters of the process of sensitivity training (see Chapter 8) as an essential ingredient to effective supervision.

COMMENTS ON BLAKE'S MANAGERIAL GRID

It is unfortunate that space does not permit greater elaboration of the different theories of leadership. However, for several reasons your authors feel it necessary to comment in somewhat greater detail on one of these theories—the Managerial Grid of Blake. First, the managerial grid "theory" is the only leadership theory that is being offered today to industry as a packaged managerial development program designed to improve managerial effectiveness and to improve organizational efficiency. Second, it is the only model for which substantial claims have been made concerning its actual effectiveness in terms of dollars-and-cents benefits to the firms which have tried it out. Thus, it seems important to look at this theory a little more closely than the others to evaluate its merits since it apparently is being promoted rather successfully in industry by its proponents.

We mentioned earlier that the theory is predicated upon the notion that leadership effectiveness is based upon two dimensions—concern for people and concern for production. As a theory then, it is somewhat similar but far less comprehensive than the view of Bass. Optimal management consists of behaving in a way so as to maximize both leadership dimensions—that is, managers should be trained to the 9.9 pattern. Unfortunately, the research data examined earlier in this chapter would lead one to the conclusion that effective leadership is just not that simple. For example, the data from Fleishman and Harris (1962) found noticeable and important interactions between the very similar variables of consideration and structure. Also, Fiedler in his research seems to find that leadership is a very situational kind of thing and that the type of leader required depends upon the characteristics of the situation. Finally, there is no absolute certainty that these two dimensions of Blake's are *the* basic dimensions of managerial effectiveness. There would appear to be a considerable need for much more extensive research evidence that the 9,9 managerial style is indeed optimal across a wide range of situations before the grid model could be viewed with confidence as a reasonable approach to managerial development.

In implementing his managerial development ideas, Blake utilizes a six-phase training program which can be divided into two major parts:

A. Management Development
   - Phase 1. Laboratory-seminar training
   - Phase 2. Team development (intra-group training)

B. Organizational Development
   - Phase 3. Inter-group development
   - Phase 4. Organizational goal setting
   - Phase 5. Goal attainment
   - Phase 6. Stabilization

Again, however, there is as yet no adequate empirical evidence available to show that this training does indeed accomplish the desired objectives. Blake has provided certain anecdotal evidence in support of the effectiveness of the program, but nothing formal in the way of research evidence seems available at the present time. Even assuming that (1) the 9,9 goal is appropriate and (2) the training program does direct managers toward 9,9 behavior patterns, we *still* might well want to know

if this behavior continues on after training or do managers quickly revert to old patterns when back on the job (as Fleishman found in his research)? The answer to this question is essential before one can make any evaluation of the worth of any type of management development program.

Last, but equally important, is the data presented in support of Blake's model based upon changes in the company profit structure (see Blake, Mouton, Barnes, and Greiner, 1964). At best, the evidence presented seems interesting but highly equivocal. The data are longitudinal in nature in that the profit structure prior to the installation of the managerial development program is compared to the profit structure after installation. Unfortunately, such studies are virtually worthless as objective empirical evidence for any type of organization change since one has no "control" condition against which the experimental condition can be compared. Therefore, one is never certain just what might have caused the change in profit structure. True, perhaps the change *is* a function of the Blake system, but it might just as easily be due to (1) a changeover to automation, (2) a dramatic change in general market conditions within the industry, or (3) a general change or upswing in the entire economy.

At present it appears that the managerial grid model does not possess enough substantive data in its support to allow us to take the position that it has provided the long-sought answer to the leadership problem. Until such evidence is forthcoming, one must view its unquestionable acceptance by industry with more than a little concern.

COMMUNICATION

Without communication, effective leadership cannot be maintained. Any organization requires a system of communication in which orders and information can travel from higher to lower levels, and from lower to higher as well. Leadership is often limited by its inability to channel information correctly throughout the organization. Material that is transmitted down is usually referred to as an "order"; material transmitted from subordinate to superior is usually referred to as a "report." Order or report, both are important in the functioning of an organization, and the smooth flow of work often depends upon their accurate and rapid transmission. It is the direct responsibility of the superior to devise an efficient system of communication and see that it is kept in operation. Listening facilitates upward communication.

Communication in an organization is effective when it meets a number of requirements. First, the channels of communication must be known to all. Organizations function differently in this respect. At one extreme are the companies which require their employees to stay in line and communicate only with the employee immediately above or below them. At the other extreme are the companies which permit any employee to transmit communications to anyone else, regardless of position. Whatever the system, every employee must be familiar with it.

Second, the system of communication must reach all the employees with dispatch. If a communication involving many employees is not received by all of them at approximately the same time, rumors will flourish and employees will feel insecure; they may decide that favoritism is being shown somewhere along the line. Third, to avoid misunderstandings and make certain of correct transmission, the lines of

communication should be as short as possible. The outstanding example of direct communication is the situation where the president of the company speaks over a loudspeaker system to all the employees. There is nothing like the feeling one has when one gets information "straight from the feed box."

Fourth, a communication should be authoritative. According to Barnard (1938), a person regards a communication as authoritative when four conditions exist simultaneously: (1) when he understands the communication, (2) when he believes it to be consistent with the purposes of the organization, (3) when it is compatible with his interests as an employee, and (4) when he is mentally and physically able to comply with it. Pigors (1949) points out a fifth necessity for effective communication: It should not be limited to the minimum of factual information nor should it be given at the last moment.

The sixth and possibly the most overlooked aspect of effective communication is that it is a joint process. Management is quick to recognize the need for communications to flow downward. Equally important is the need for communication to flow upward. Communication, therefore, must be regarded as a two-way process in which employer and employee have equal opportunity to convey their meaning, their feelings, and their actions to each other. A stumbling block which impedes communication and prevents understanding is the suspicion that may exist in either of the parties. All efforts should be exerted to remove any basis for suspicion. Free-flowing communication both upward and downward goes a long way to promote understanding.

DO'S AND DON'TS

It is very difficult to devise a list of specific recommendations for the successful leader because such a list must be general in nature, whereas most situations which confront the leader are specific. A further difficulty is that, while a person may memorize the list, there can be no guarantee that he will be able to apply the rules to the specific situation confronting him. In spite of this, it is nevertheless advisable to summarize the points made previously by briefly listing what a leader should do and what he *should not* do.

A leader *should* base his actions on the following five principles:

1. *Fair evaluation of work.* A subordinate is constantly seeking an appraisal of his work; hence the superior should not hesitate to give praise or criticism periodically as well as whenever the situation calls for it. He should remember that people may be praised publicly, but that no one should be criticized publicly.

2. *Sufficient delegation of authority.* A "perfect executive"—if there were such a thing—would be a person who has no work to do himself because his subordinates do it for him. His function is to assign the work and see that it is done properly. But once the superior has delegated authority, he should back up the subordinate at all times. However, if the subordinate misunderstands or fails to carry out instructions, he should not be supported at the expense of the organization. Under such circumstances the delegation of authority should be changed. News of such changes should be communicated to the employees in advance.

3. *Fair treatment for all.* The leader must remember that one employee is as important as any other employee, even though one may have higher status in the

company. All people deserve and demand equal consideration. Regardless of status, every employee likes to feel that the company considers him important; he will resent being treated as if he were unimportant.

4. *Availability to all employees.* The leader should be readily available to all employees. Just as it is his prerogative to send for a subordinate, so it should be the subordinate's prerogative to make an appointment with the leader for any cause he deems justified. The "big boss" in a large organization would do well to have one of the telephones on his desk connected directly with the main switchboard for employee use. The feeling that all the employees have an equal right to talk to the "big boss" often makes it unnecessary for an employee to call him. A leader who is readily available to his subordinates impresses them with his interest in their problems and thereby promotes a more effective subordinate-superior relationship.

5. *Discussion of employee problems with employees.* Management should not make decisions concerning employees without taking the employees' wishes into consideration. Many times, taking subordinates into the leader's confidence has resulted in suggestions that have solved the problem more successfully. The feeling created in employees as a result of such a policy has outstanding rewards, not only for the employees but for management as well.

A leader should *avoid* such common practices as the following:

1. *Dependence upon superiority.* Some people take advantage of their superior position; they believe that their orders should be obeyed simply because "I am the boss." A successful leader does not have to depend upon his superior position. Rather, his superior position will be recognized, and his merit as a leader is based upon his actions.

2. *Simulation of knowledge.* Subordinates are quick to estimate the ability of a superior. Thus a wise leader will not try to make believe that he knows more than he really does. An effective leader is one who has the technical knowledge required by his job; subordinates will look up to such a leader. Sometimes they actually know more about the specific work they perform than the superior, for he is not expected to know all the intricacies of every job.

3. *Interference with work.* A leader delegates work to his subordinates. Once this has been done, within reasonable limits he should allow the subordinate to finish the job by himself. Interference on his part is often resented; furthermore, it causes distraction. A leader must not interfere with the smooth flow of the work done by his subordinates unless he wants to curtail production.

4. *Favoritism and discrimination.* The leader, as an individual, will probably like some of his subordinates more than others. However, in the work situation he cannot allow personal likes and dislikes to influence the distribution of work or to interfere with discipline.

5. *Public reprimands.* A subordinate must maintain his self-respect, especially in the eyes of his fellow workers. Just as children should never be punished in the presence of their friends, so employees who are subject to disciplinary action should be told about it in private. A public reprimand often serves only to lower the position of the subordinate. The superior should not even raise his voice when talking to a specific individual in a group.

6. *Pettiness.* Subordinates expect a leader to be magnanimous. Harping on trivialities, continually going over minor details, and demonstrating in other ways

that he does not respect the subordinate's ability often prevents a superior from maintaining harmonious relations.

7. *Conflicting orders.* A leader must remember the orders he has issued. A subordinate is often confused when the leader, failing to remember a previous order, directs him to do something just the opposite. He is usually too embarrassed to call these conflicting orders to the leader's attention, but he usually decides that the leader does not know what he is doing. In some organizations lines of authority are not clear-cut, with the result that two people with equal status knowingly or unknowingly issue directly opposing orders. Such a situation can be avoided by establishing clear lines of authority.

8. *Superfluous orders.* An effective leader is one who issues few orders rather than a great number. Unless orders are within the scope of his authority, can be clearly understood, and are possible to carry out, they will not be obeyed. Too many orders not only tend to undermine that leader's authority, but also create insecurity in the employee and prevent the exercise of a healthy independence on his part. An effective leader does not find it necessary to issue many orders.

SELECTION AND TRAINING OF LEADERS

SELECTION

Psychological tests have been successful in the selection of many different types of employees, but to date they have not been of any great help in the selection of potential executives. If this problem is solved, as it may well be, the type of test will be vastly different from that used in selecting other employees. The value of psychological tests in executive selection was discussed in Chapter 4, and we saw then that tests can be used for this purpose only when a job criterion is available. No criterion of success at the executive level has been unequivocally established. Probably the most promising technique at the moment for selecting leaders is the in-basket method. However, it still lacks sufficient validation.

One difficulty in predicting executive success is that the philosophy of management in a particular company often determines the characteristics and behavior expected in its executives. A high-pressure company may require pushing and driving. A company with a different rationale would find such persons to be misfits. Psychologists can effectively describe a person's characteristics such as intelligence level, ambition level, maturity level, and interests and abilities. Using these variables and their probable interrelations as a basis, psychologists can offer descriptive evidence that will enable management to decide whether it wants such a person for a specific executive position. It appears safest for psychologists to describe individuals and for management to make the decision as to whether to hire them. If the criterion is known, then psychologists can do the matching. However, for most executive positions the criterion is usually not clearly defined or it varies from company to company.

Although the literature reports many studies using tests to predict supervisory effectiveness, most of the findings are rather uninspiring. Possibly the major reason for this is the simplicity of the tests and the impossibility that a short intelligence test or a personality trait inventory can really measure enough of what is demanded of the executive on the behavioral level. Typical of such studies is the one reported

by Bruce (1953). He used 23 predictor variables and found 7 of them to be significantly (but not highly) correlated with the criteria. He found that the major contribution of + 0.290 to the mulitple of R of + 0.398 was made by the Otis Self-Administering Test, a 20-minute intelligence test. Possibly the greatest value of his research is part of his concluding sentence: "Further research in this area should seek to find instruments that measure factors that can successfully predict the criterion. . . ."

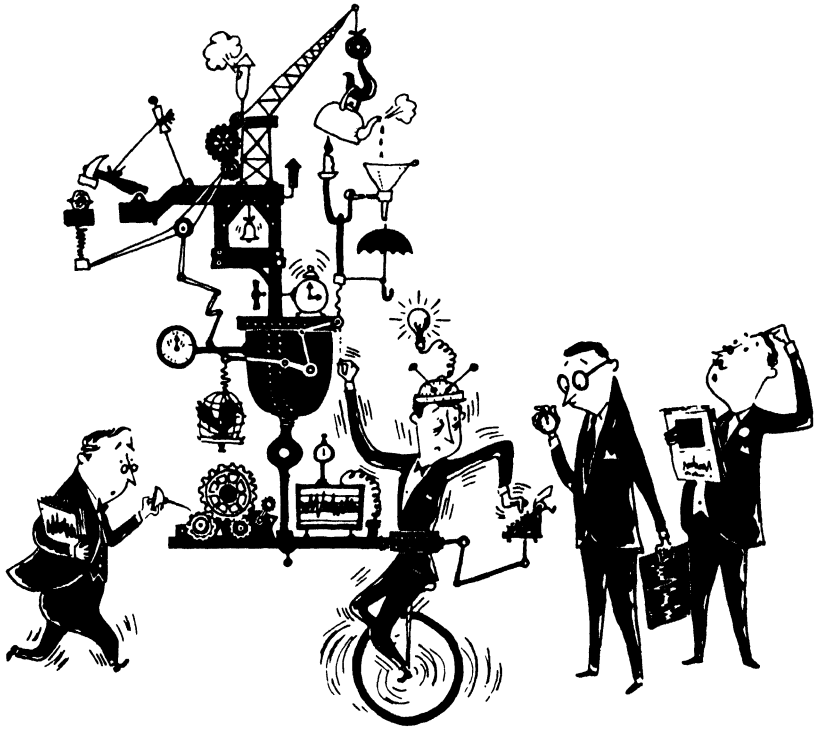


Figure 14.8. Training and development programs can get too complex.

A study published two years earlier in the same publication but not mentioned in Bruce's bibliography seems to be along such lines. Meyer's study (1951) presents possibilities and at least is refreshing in an approach that is different. Meyer did not try to rework the same hash that has been served before and expect to come up with a new recipe. His measures of leadership ability included, among others, the development of a multiple-choice projective test of social attitude. Examples of the type of item included are as follows:

29 Harry's supervisor is near the retirement age. Most of the men will be glad when he retires because he is so grouchy. How would you expect Harry to feel toward him?

_____ (a) He probably agrees with the rest of the men.
_____ (b) He probably feels that he might be grouchy too if he were as old as the boss.
_____ (c) He probably tries to avoid the boss.
_____ (d) He probably figures that something must be troubling the boss.

34. One of the men in the department, Joe Smock, borrowed 10 dollars from Harry and promised to return it in a week. When the week was up, Joe told Harry he could pay back the $10, but that he would like to keep the money for another week if it were all right. Although Harry would have liked to get the money then, he didn't need it. What would you expect Harry to have told him?
_____ (a) "That's all right."
_____ (b) "I'd rather have the money right now if you don't mind."
_____ (c) "Sure, Joe. Is there anything else I can do to help?"
_____ (d) "I'd like to but I'm short myself."

Meyer views this test as a measure of social perception, and he found that it was related to ratings of leadership ability used as the criterion. He infers that the way a person perceives a situation is determined by his attitudes. Implied is that change of performance as a supervisor should concentrate on changing attitudes rather than trying to teach skills.

Meyer presents evidence to indicate that the good supervisor regards others as individuals with motives, feelings, and goals of *their* own, whereas the poor supervisor is more likely to perceive others in relation to *his* own motives or goals.

Mahoney, Jerdee, and Nash (1960) present data indicating at least partial success in predicting managerial effectiveness with personality and demographic data. They validated 98 different measures based upon the following:

1. Intelligence test
2. Empathy test
3. Vocational interest test
4. Personality test
5. Personal history questionnaire

The criterion of managerial success was based upon rankings of supervisory effectiveness for each of 150 men. Eighteen of the 98 measures predicted the criterion significantly better than chance. The prediction equation based upon these 18 measures was then cross-validated upon a new sample of 156 managers and found to predict their rankings at better than a chance level.

Jenkins (1947) has proposed the nominating technique as an approach to the problem of selecting executives. This technique is very similar to one proposed by Moreno much earlier, but Moreno did not have in mind the specific problem of industrial leadership. Jenkins' original purpose was to obtain information on the "combat suitability" of naval aviators, but he soon realized that this study was producing interesting data on leadership.

Interviews were held with naval pilots, the conversation being centered about the following questions:

> Assume that you are to be shifted to a new air group tomorrow and that you may select your own combat-mates to go with you to this new air group. Of all the men known to you in Naval Aviation—living or dead—what two men would you like most to fly with in your new combat assignment? Why would you select these men?
> What two men would you least like to have flying with you? Why?

The pilots' selections showed diagrammatically the difference between the formal leadership imposed by chain-of-command and the informal, functional leadership set up by the men themselves. The squadron commander in one squadron was not

nominated once; the executive officer received multiple nominations—but they were all negative. There were many small cliques in this squadron and all the "not-wanted" nominations were men within the squadron. This squadron was recognized as having low morale, and the nominating technique showed the absence of leadership.

In another squadron known for its high morale and excellent combat record, the diagram resulting from the nominating technique was very different. The squadron commander was named most frequently as "wanted in combat"; the executive officer was the second most desired combat-mate. There was no evidence of subgroups, and more than 50 percent of the "not-wanted" nominations were men outside the squadron. In this squadron the man who ranked third was a lieutenant, junior grade. Jenkins points out that if this young lieutenant were to become a leader solely on the basis of seniority, it would take an unwarranted length of time for his leadership to be utilized.

Many people believe that the best way to detect potential leadership is to find a man who already has followers. The nominating technique can do just that.

Jenkins recognizes that this technique cannot be used indiscriminately. He believes that the following five conditions are a prerequisite for its success:

1. The tasks involved represent high levels of skill.
2. The motivation to do well is in proportion to the cost of failure.
3. The same basic skills are required of all members of the group.
4. Individual members have firsthand knowledge of the performance of the other members.
5. The group has been closely associated over a long period of time.

It appears certain that when a group selects its leader it does so not on the basis of popularity or personality but on the basis of a rather critical evaluation of all its members. Superimposing a leader on an already formed group may do much to change the general atmosphere. Such findings as these support the policy of promotion from within the ranks and even suggest that the group can aid in selecting the individuals to be promoted. Management should not forsake its accepted role of selecting the leader; but when, in selecting, it fails completely to consider possible candidates among the employees, its choice may not be very successful. It is possible, of course, that a group will be unable to reach a decision as to which member will make the most efficient leader, and under these circumstances management will have to bring in someone from the outside.

Bavelas (1947) points out that optimal face-to-face relationship is obscured by viewing this as a matter of "personality" or knowing "human nature." Personal relations depend upon social skills which are not learned out of books any more than tennis can be learned from books. Although lectures, conferences, or group discussions are successful in presenting a viewpoint, they have the limitation of "talking about" rather than "doing." Bavelas favors role playing as a management training technique.

TRAINING

One means of training leaders is by means of the technique variously called "reality practice," "role playing," or "the problem approach." A realistic problem is presented and the subjects are required to act out a solution to it (see Chapter 8).

For example, a group of foremen who are being trained in effective leadership may be presented with the problem of an employee's infraction of a company rule. Rather than allow discussion on an unrealistic basis, one of the foremen is asked to play the role of the employee and another to take the role of the foreman. The two men set the stage by planning their handling of the problem, and then they act out the situation before the group. Afterward the group discusses the problem, paying special attention to what it considers mistakes made by the two foremen. The performance is then repeated; two other men play the two roles to see if they can avoid the mistakes made in the first presentation. Again there is group discussion, the members trying to determine which solution was the better and also to analyze the behavior which led to this solution.

In a variation of this technique, one man takes the role of the employee who has broken the rule and he is confronted by three different foremen, none of whom knows the approach used by the other two. The group then decides which was the best way to handle the situation or what particular aspects of each technique would contribute most to the solution. The three foremen are asked to repeat their performances, modifying them according to the suggestions of the group.

This type of training encourages learning at a very meaningful level. The person in charge of the training has a passive but nonetheless important part. It is up to the group to choose the problems which are of greatest concern to them; the leader must encourage active participation. At best, he can steer the discussion but he can never really direct it; therefore the spontaneous aspects of the training are allowed to emerge.

French (1946) reports that in training foremen to handle interpersonal relations, role playing has a number of distinct advantages over the more conventional methods of lectures, conferences, reading, and discussions. Role playing provides an excellent bridge between discussion of interpersonal relations and their actual handling. It presents test situations to determine how well a trainee can handle various problems and makes the leader sensitive to different methods and styles of leadership.

An illustration of role practice as a means of training foremen is presented below. Here the problem was to increase an inefficient factory operator's production; the quotations (French, 1946)[3] show how three people attempted to handle it.

FOREMAN *(getting up reluctantly and going over to girl at the desk)*: Well, Dottie, you only made 30 units yesterday. Did you have any special trouble? After I brought you all your work, too.

GIRL: I didn't feel good.

FOREMAN: Did you have any machine trouble or anything?

GIRL: Yes, I did.

FOREMAN: Well, why didn't you put your little red light on?

GIRL: I don't know. I guess I forgot. And the thread breaks all the time.

FOREMAN: Well, you should tell me about those things so I can help you. You'll do that after this, won't you? And you'll try to do better too, won't you?

GIRL: Yes, but I don't know if I can. It's hard to do.

The discussion continues with suggestions for using check studies and giving additional training. When the group is discussing "making excuses," the personnel manager, Mr. Jones, thinks that the supervisor should find out whether the girl has

[3] Abridged and adapted from John R. P. French, Jr. Role playing as a method of training foremen. In Schuyler Dean Hoslett (ed.). *Human factors in management*. Copyright 1946 by Harper & Row, Publishers, Incorporated. Used by permission of the publishers.

any personal problems. Because the trainer knows that this foreman has created resentment by prying into the personal affairs of her girls and because he wants to criticize the argumentative technique without criticizing her personally, he sets up a special situation.

TRAINER: Let's see how Mr. Jones would tackle this problem. I'll be the girl, the same girl as the last time. (Sits at the "*machine.*")

MR. JONES: Well, it just seems to me this way. There's lots of things that might be holding the girl back. Possibly she got a letter from her boy friend that had bad news in it, or something like that.

TRAINER: Well, let's try it out and see how it works.

MR. JONES (*getting up from his chair*): What did you say your name was?

TRAINER: Dottie Sholley.

MR. JONES (*now acting the role of supervisor*): I have some bad news for you here, Dottie. It seems you have fallen down a little in your units. What seems to be the trouble?

TRAINER: Well, I didn't feel so good.

MR. JONES: But when you asked me to be your supply girl you seemed to be feeling well enough.

TRAINER: Well, I got some machine trouble and that slows you up. And these old threads break all the time. You can't do much when that happens.

MR. JONES (*pauses for a moment*): Did you go to the show last night?

TRAINER: No.

MR. JONES: Anyone in your family sick?

TRAINER: No.

MR. JONES: Did you have a date last night?

TRAINER: No, I'm married.

MR. JONES: And you say you haven't been feeling well?

TRAINER: No, I wasn't feeling well, but that was yesterday.

MR. JONES: You don't feel sick most of the time?

TRAINER: No, that was just a little stomach trouble. There's nothing wrong with me.

MR. JONES (*laughing*): You're sure bucking me. I give up!

BILL: If you let them get into an argument with you, you'll never get out. They answer and answer and answer.

TRAINER: Thanks. I certainly was being a tough one. Now let me give you my reactions. When you came up saying you had bad news I felt nervous. I didn't like that, so I was sort of on the defensive. I was thinking, now what am I going to say? I thought of something, and then while I thought of that I thought of something else to have ready for the next question. Then he asked me about my family and that scared me. I thought maybe something was wrong. Then he asked me if I had had a date, and I was married. That made me so mad that I nearly slapped him. And when he asked me again about my health, I tried to assure him that I was in good health because I was afraid maybe he would fire me if he thought I was sick all the time. Now, this is the toughest kind of case you will get. A girl who doesn't do what she can do, and you just can't find out why. Now would you like me to try the skunk oil method?

BILL: Yeah, I think so.

TRAINER: O.K. You be the girl, Bill.

BILL: Sure, I'll answer your questions.

TRAINER: You try to be the same girl that Anne was and that I was. Be as tough as you want to.

BILL: O.K. (*Sits down at the table.*)

TRAINER (*approaching Bill with the sheet in his hands*): Hello, Dottie. Here's the unit sheet for today. Let's see, where is your name? (*Turning the sheets with Bill's help.*) I guess it's over on another page. What have you been doing?

BILL: Well, I made 30.

TRAINER: How does that compare with what you've been doing? Is that good for you or not so good?

BILL: Well, I have done better.

TRAINER: How long have you been on the job?

BILL: About six weeks.

TRAINER: Well, it usually takes a girl three or four months to make 60. You say you have done better?

BILL: Yes.

TRAINER: Have any trouble yesterday?

BILL: Yes, the thread breaks all the time. And I had such little bundles. I had to get more all the time.

TRAINER: Oh, I'm sorry. I told you yesterday I was going to bring you a lot.

BILL: But they're too little. You run through them in no time.

TRAINER: What you want to do is not worry about your progress one day or another day. How much do you suppose you will make a week from today? Maybe you'll get some small bundles and maybe your machine will give you trouble, but counting that in, what do you suppose you will make in a week?

BILL: I don't know. I might make 40 or 45.

TRAINER: You think you could make 40 or 45! Why I've known girls who have taken three or four weeks to get up there from 30! What's the best you have made?

BILL: I think it's 48.

TRAINER: Well, maybe you could then. How'd you like to try and make 40 by next Friday?

BILL: You mean just do 40 by next Friday?

TRAINER: Yes, that gives you a good chance in spite of machine trouble and those things that you can't help that come up. Do you think you could do it?

BILL: I believe so.

TRAINER: Now I don't think you can do it if you have troubles that aren't your fault. Now on the matter of thread breaks, sometimes that's the way you hold your cloth and sometimes the trouble is with the machine. When you get trouble like that, we can call the mechanic in or we can get the trainer over to see what's wrong. You want to have perfect working conditions. I'll come over Friday to see if you have made it, and I'll come around every other day, too, to see if I can help in some way. *(End of role playing.)*

MR. JONES: Fine!

TRAINER: I don't think I did that very well, but I was trying to use a different technique. Now what's the difference?

MR. JONES: Well, you weren't on the defensive all the time.

TRAINER: You mean Bill didn't put me on the defensive?

BILL: What he means is when you were the operator you answered him back. And I could answer all your questions to me this time, but there never was any blame on me.

A detailed discussion continued for fifteen minutes on how to avoid arguments, putting a person on the defensive, the use of production goals, why the trainer used the *lower* of the girl's two estimates as the goal, etc.

Tannenbaum, Kallejian, and Weschler (1954) cite three limitations of conventional training programs: (1) the trainee is removed from the social setting in which he customarily performs; (2) imparting human relations information may have little or no effect in inducing changes in behavior; and (3) the carry-over from the training to the work situation may be negligible. To minimize these limitations they propose that a trainee group be vertically structured, that is, include different levels of management as in the work situation, and that the training emphasize the development of the trainee's sensitivity to himself, to others, and to the ways in which people interact. They regard the functions of the trainer as creating situations

conducive to learning by developing interpersonal skills in the group members. In more recent writings, Tannenbaum and his coworkers (Tannenbaum, Weschler, and Massarik, 1961) strongly support the use of the "T" group strategy for training executives.

The outstanding characteristic of companies that have training programs is the combination of many techniques. Planty (1952) describes the program at Johnson & Johnson as including 15 aspects:

1. Review of economic conditions
2. Multiple management
3. Role playing
4. Case studies
5. Specialized conferences
6. Harvard advanced management course
7. Wharton School of Finance
8. Reading
9. Conference leadership
10. Public speaking
11. In-plant conferences
12. Psychiatric group study
13. Out-of-plant conferences
14. Short courses, seminars, school and college programs
15. Counseling

Van Ark (1951) describes the General Foods Corporation program as including job rotation, understudying, outside courses, internal informal training, individual counseling, and staff meeting participation.

Worthy (1953), referring to the Sears, Roebuck & Company program, recognizes that four things are basic to a successful development program: (1) system and organization; (2) two kinds of measurements, one focused on the man and the other on performance; (3) the program must be based on certain principles of morality and fair dealing; and (4) whole-hearted understanding and support of top management. The Sears plan is comprehensive and includes the following five elements:

1. Analysis of organization needs
2. Selection of persons who appear to have potentialities for advancement
3. Preparation of training plans to aid in realization of such potentialities
4. Periodic appraisal to evaluate individual progress
5. An executive inventory control to help maintain the necessary balance between numbers and kinds of persons coming up the line

Habbe (1950), in cooperation with the National Industrial Conference Board, made a study of nine companies offering training programs for executive development. He summarizes the twenty techniques found in use as follows:

Almost always used:

1. Merit or performance reviews
2. Visits to other companies
3. Attendance at technical meetings and management conferences
4. Distribution of reading lists and management bulletins
5. Group meetings

Often used:

6. Job rotation
7. Advanced management courses at Harvard and elsewhere
8. Consultants
9. Committee assignments
10. Executive inventories
11. Job descriptions

Sometimes used:

12. Special trainee positions
13. Multiple management plans
14. Management cabinets
15. Understudy or assistant-to positions
16. Community leadership
17. Health and fitness programs
18. Individual counseling
19. Assignment to training or sales department
20. Sponsor plan

The additional value of this National Industrial Conference Board report is the inclusion of case studies of the following company programs:

1. Standard Oil Company (New Jersey)
2. Lockheed Aircraft Corporation
3. Hardware Mutual Casualty Company
4. McCormick & Company
5. Sears, Roebuck & Company
6. Bank of America
7. Swank, Inc.
8. Eastman Kodak Company
9. Bigelow-Sanford Carpet Company

The Standard Oil Company (New Jersey) initiated an executive development program in 1944. It has grown in importance with the years. Its basis is reducible to four rather simple principles: (1) Know what the management structure of your company is, (2) select candidates who have the qualifications required by the jobs, (3) establish a simple method of appraising the candidates selected, and (4) provide the training necessary to complete the candidate's experience (Suman, 1954). The best way to determine the managements structure and estimate manpower needs is to study a chart such as the one in Figure 14.9.

A word of caution must be added to the listing of the "big" companies. "Of course these companies can afford such programs" is the rationalization offered by small companies who sometimes are not so small. The fact is that small companies, that is, 50 employees or less, may need training and development programs even more urgently. In a large company a single executive may be "covered up," "kicked upstairs," or otherwise "swallowed." His mistakes may be neutralized by the bigness of the company with its counterchecks. The smaller company executive occupies a relatively more important role in his company. Smaller companies often are strong but have no personnel reserves. For example, consider a crude analogy from football: Two teams may have equally good first stringers, but their reserves often determine which team wins. Likewise, the smaller company must be concerned with its reserves, especially when they are nonexistent. No company, big or small, can expect to continue without potential replacements.

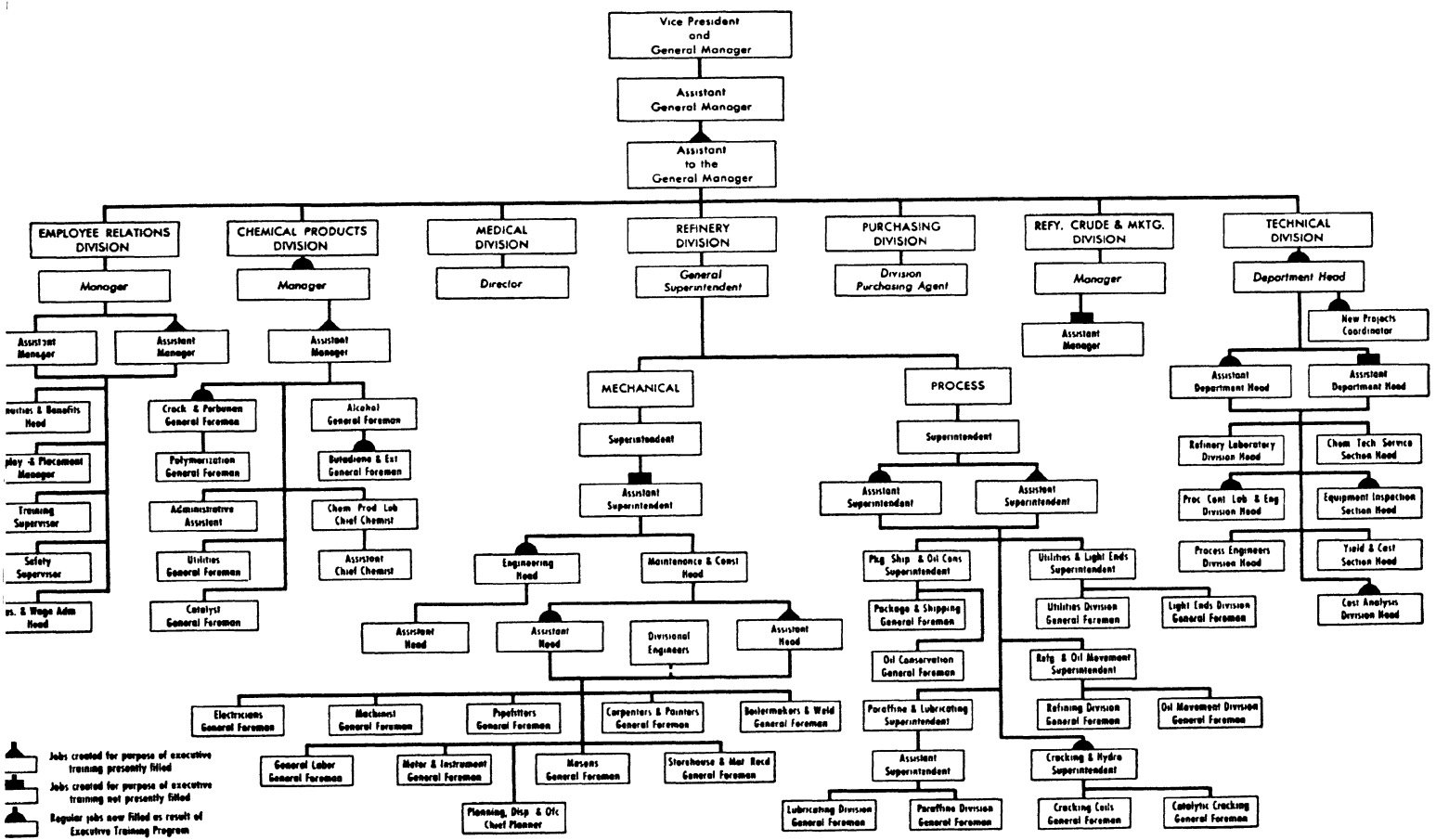


Figure 14.9. How jobs may be created for training purposes.

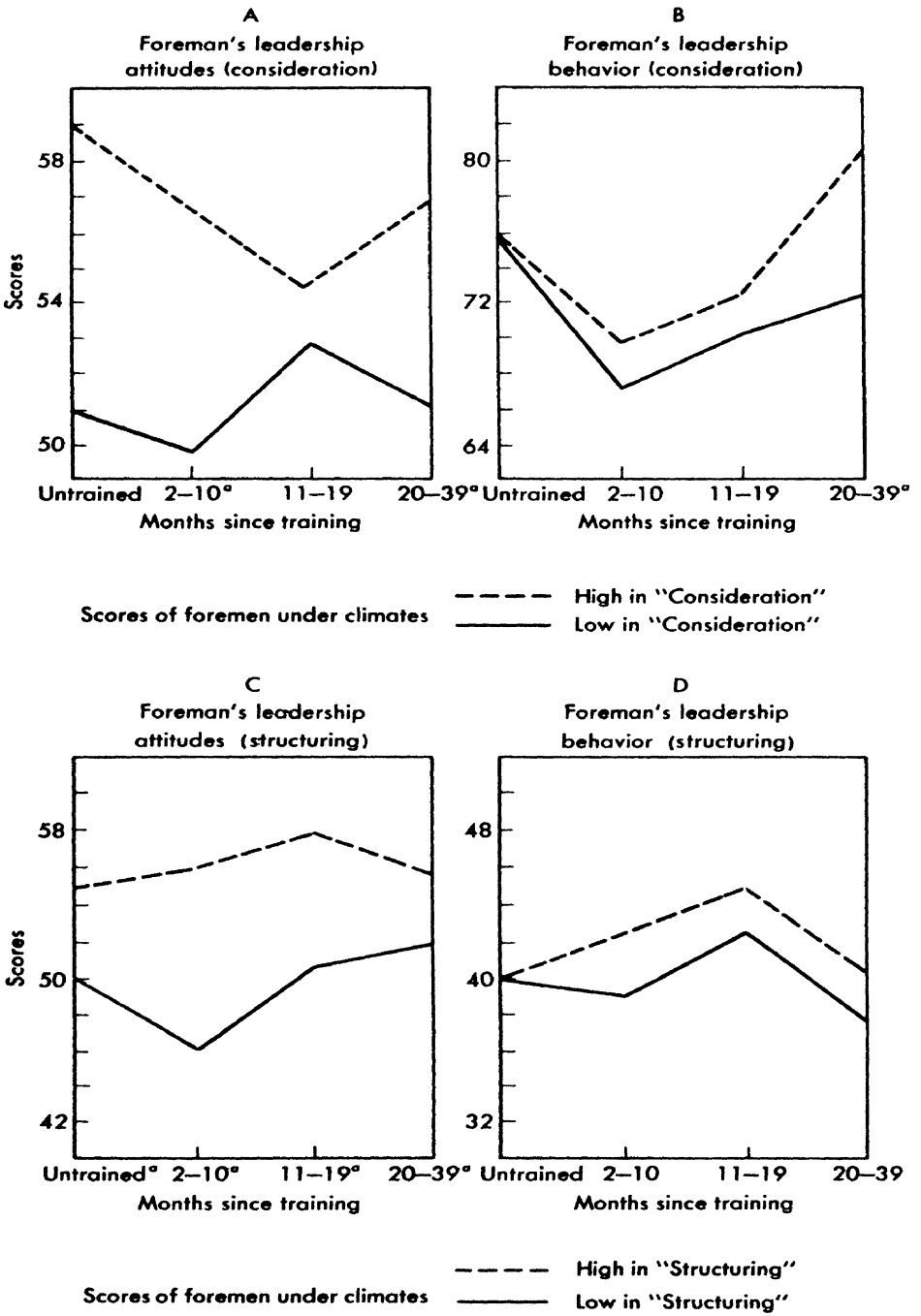


Figure 14.10. Comparison of the leadership attitudes and behavior of foremen operating under different "leadership climates" in the plant. (From E. A. Fleishman. Leadership climate, human relations training and supervisory behavior. *Personnel Psychology*, 1953, *6*, 205–222.)

TRAINING EFFECTIVENESS RESEARCH

No area in industrial psychology has had less good research than the area of evaluating the effectiveness of executive training programs. The rare definitive research efforts which have been carried out often indicate that attitudes do change (at least ostensibly) over the course of the training period, but one wonders what happens when the executive gets back on the job. Fleishman (1953b) has provided a rather dramatic answer to that very question—it casts grave doubts as to whether *any* type of human relations training is really worth the time and effort put into it.

His later study examined what happens to the attitudes and the behavior of a supervisor who has received human relations training and who is then thrust back into his original work environment. The results are shown very clearly in Figure 14.10.

The curves for "behavior" are more dramatic than those for "attitude." The findings show several things:

1. The effect of training may or may not result in a clear initial change of attitude or behavior.
2. Even if training does have an effect, it may be only temporary.
3. The behavior of a supervisor seems to be primarily related to the behavior of his own immediate supervisor. If his supervisor is considerate, he will be also, etc.

The "back-home" leadership climate is an important variable related to the behavior and attitudes of foremen in the work situation. This would mean that leadership training cannot be considered in isolation from the social environment in which the foremen must function. Changes in foremen, as a result of this training, may be difficult if the work situation remains constant. In other words, certain aspects of the foreman's environment may have to be recognized if the training is to be effective in modifying his behavior.

Training may be effective if the work environment tolerates the changes; if not, the person may be unable to use what he has learned. He may even be frustrated by not being able to do anything and as a result be unhappy with his training and the plant conditions because he *was* trained.

The findings of Fleishman would seem to indicate that if we want to be effective in our human relations training we had better start at the top levels first and work down, rather than trying to effect changes at the lowermost levels of supervision.

ROLE OF EXECUTIVE WIVES

An often ignored aspect of leadership and executive behavior is the importance of the wife. Not too much is really known about the role of the wife and her "effect" upon the effectiveness of her husband as an executive. Helfrich (1965) has provided some interesting data that indicate the role of a wife varies as a function of the level of management at which her husband is located (see Table 14.4).

TABLE 14.4 *Roles of Executives' Wives, Classified by Husband's Level of Management*

Role	Level of Management		
	Top Level	*Other*	*Total*
Family-centered	3	20	23
Community-centered	11	0	11
Creative	4	2	6
Consultant	1	5	6
Career	1	2	3
Student	1	1	2
Total	21	30	51

SOURCE: Margaret L. Helfrich. *The Social role of the executive's wife*. Bureau of Business Research, Ohio State University, Columbus, 1965, 35–36.

One cannot tell from Helfrich's data why there is a strong tendency for wives of top executives to be community-centered. The same problem of interpretation applies to the tendency of lower management wives to be family-centered (for example, perhaps they are simply younger and have more children at home). However, Helfrich's study merits attention in that it points out an important and neglected research area.

REFERENCES

Argyris, C. *Personality and organization: The conflict between system and the individual.* Harper & Row, New York, 1957.

Barnard, C. I. *The functions of the executive.* Harvard University Press, Cambridge, Mass., 1938.

Bass, B. M. *Leadership, psychology, and organizational behavior.* Harper & Row, New York, 1960.

Bavelas, A. Role playing and management training. *Sociatry*, 1947, *1*, 183–191.

Bennis, W. G. *Changing organizations.* McGraw-Hill, New York, 1966.

Blake, R. R., and Jane S. Mouton. *The managerial grid.* Gulf, Houston, Texas, 1964.

Blake, R. R., Jane S. Mouton, L. B. Barnes, and L. E. Greiner. Breakthrough in organization development. *Harvard Business Review*, 1964, *42*, 133–155.

Bradford, L. P., and R. Lippitt. Building a democratic work group. *Personnel*, 1945, 22, 1–12.

Bruce, M. M. The prediction of effectiveness as a factory foreman. *Psychological Monographs*, 1953, *67* (12).

Fiedler, F. E. A contingency model of leadership effectiveness. In L. Berkowitz (ed.). *Advances in experimental social psychology.* Academic Press, New York, 1964, 150–190.

Fleishman, E. A. The description of supervisory behavior. *Journal of Applied Psychology*, 1953(a), 37, 1–6.

Fleishman, E. A. Leadership climate, human relations training and supervisory behavior. *Personnel Psychology*, 1953(b), *6*, 205–222.

Fleishman, E. A., and E. F. Harris. Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*, 1962, *15*, 43–56.

French, J. R. P. Role playing as a method of training foremen. In S. D. Hoslett (ed.). *Human factors in management*. Harper & Row, New York, 1946.

Ghiselli, E. E. Traits differentiating management personnel. *Personnel Psychology*, 1959, *12*, 535–544.

Ghiselli, E. E., and T. M. Lodahl. Problems of managerial traits and group effectiveness. *Journal of Abnormal Social Psychology*, 1958, *57*, 61–66.

Habbe, S. Company programs of executive development. In *Studies in personnel policy #107*. National Industrial Conference Board, 1950.

Helfrich, M. L. *Social role of the executive's wife* Bureau of Business Research, Ohio State University, Columbus, 1965, 35–36.

Jenkins, J. G. The nominating technique; its uses and limitations. Paper presented at Eastern Psychological Association, 1947.

Katz, D., and R. L. Kahn. *Human organization and worker motivation in industrial production*. Industrial Relations Research Association Publication #7, 1951.

Katz, D., N. Maccoby, and N. C. Morse. *Productivity, supervision and morale in an office situation*. Survey Research Center, University of Michigan, Ann Arbor, 1950.

Likert, R. Motivational approach to management development. *Harvard Business Review*, 1959(a), *37*, 75–82.

Likert, R. A. A motivation approach to a modified theory of organization and management. In M. Haire (ed.). *Modern organizational theory*. Wiley, New York, 1959(b).

McGregor, D. *The human side of enterprise*. McGraw-Hill, New York, 1960.

McMurray, R. N. The case for benevolent autocracy. *Harvard Business Review*, January–February, 1950, 82.

Mahoney, T. A., T. H. Jerdee, and A. N. Nash. Predicting managerial effectiveness. *Personnel Psychology*, 1960, *13*, 147–163.

Meyer, H. H. Factors related to success in the human relations aspect of work group leadership. *Psychological Monographs*, 1951, *65* (3).

Pigors, P. *Effective communication in industry*. Lt. Rush Toland Memorial Study #1, National Association of Manufacturers, New York, 1949.

Planty, E. G., and C. A. Efferson. Developing leadership for tomorrow's tasks. *Dun's Review*, January, February, 1952.

Porter, L. W. Perceived trait requirements in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, *45*, 232–236.

Porter, L. W., and M. M. Henry. Job attitudes in management: V. Perceptions of importance of certain personality traits as a function of job level. *Journal of Applied Psychology*, 1964(a), *48*, 31–36.

Porter, L. W., and M. M. Henry. Job attitudes in management: VI. Perceptions of the importance of certain personality traits as a function of line versus staff type of job. *Journal of Applied Psychology*, 1964(b), *48*, 305–309.

Randle, C. W. How to identify promotable executives. *Harvard Business Review*, 1956, *34* (3), 122–134.

Stogdill, R. M. *Individual behavior and group achievement*. Oxford, Fair Lawn, N.J., 1959.

Suman, J. R. Growing a good executive crop. Speech, Standard Oil of New Jersey, 1954.

Tannenbaum, R., V. Kallejian, and I. R. Weschler. Training management for leadership. *Personnel*, 1954, *30*, 254–260.

Tannenbaum, R., I. R. Weschler, and F. Massarik. *Leadership and organization*. McGraw-Hill, New York, 1961.

Van Ark, G. Development of executives. Talk presented to National Office Management Association, Chicago, 1951.

Viteles, M. S. *Motivation and morale in industry*. Norton, New York, 1953.

Worthy, J. C. Executive personnel development. *Advanced Management*, 1953, *18*, 5–8.

15 DECISION MAKING

In Chapter 8, "Training and Learning," when we were discussing leadership, or executive, training, we pointed out that the executive "role" really consists of two major components—a decision-making component and a human relations component. In the previous chapter on leadership the human relations aspect of the executive's role was the major theme. Most research on leadership and most attempts at developing theories of leadership have tended to emphasize the human relations or interpersonal dimension. More recently, however, greater attention has been devoted to an examination of the process of decision making or "choice behavior" per se. This is, in our opinion, a long overdue occurrence. As we pointed out earlier, it is really the decision-making function which *defines* the leadership role. It is very difficult to conceptualize a leadership situation which does not involve making a decision of some type.

In this chapter we will look at the decision-making process itself, being as analytical as possible. We will look at some conceptual and quantitative schema which have been used to describe decision making, both in individuals and in organizations.

THE DECISION PROCESS

Each of us makes so many decisions in the course of any normal day that we may tend to view the decision process as a simple, relatively uncomplicated act. It is only when we find ourselves faced with one of those very "tough" decisions in which we just cannot seem to make up our minds that we begin to appreciate all the complexities of the process that a person goes through in trying to decide upon an appropriate course of action.

DECISION PROBLEMS

Larson (1962) believes that the key to understanding the decision process lies in understanding the problem which has created the need for a decision. He believes that failure to define and to understand the decision problem adequately is what causes the greatest difficulty in decision making. According to Larson there are four different types of decision problems.

1. The objectives have been defined and alternative courses of action have been identified which may lead to these objectives, but the problem is in deciding which course of action is best.
2. The objectives have been defined, but we do not know which courses of action will potentially achieve these objectives. The problem is in determining the possible alternatives and then selecting the most desirable alternative.
3. The objectives themselves have not yet been formulated, although a need for action is evident. The problem involves making a comprehensive choice concerning the alternatives.
4. A situation has been identified as being either currently undesirable or potentially undesirable at some future point in time. The problem involves devising actions which will eliminate or alleviate these undesirable conditions.

The major difference in (2) and (4) above would appear to be that in (2) the initial task involves "identifying" courses of action while in (4) the initial task involves "devising" courses of action.

The making of decisions usually implies responsibility. That is, whenever a decision is made, the person making it must be prepared to assume the responsibility for the eventual consequences of his actions. Thus the student must be prepared to explain poor grades to his parents if he decides not to study so that he may go on a date. The junior advertising executive must be prepared to take the consequences of his decision to select a particular advertisement from among those available to him. Even the company president has to be ready to defend his actions to the board of directors and, at times, even to his employees. The making of decisions is all too often not an enjoyable or pleasant task, particularly when one is forced to select from among several undesirable but necessary courses of action (the avoidance-avoidance conflict). Someone once said about decision making "Don't worry about it; if it's important enough, *someone* will do something about it!" While this is one way to avoid ulcers, it does *not* seem like a strategy of life that would lead a person to achieve much success as a maker of decisions.

STEPS IN THE DECISION PROCESS

The process of making a decision may be separated into a sequence of discrete and somewhat separate steps or phases for purposes of being analytical. For example, Calkins (1959) has suggested that appropriate decision making on the part of an administrator involves the following five steps:

1. *Identify the problem and understand it.* Unless the problem is clearly defined it is impossible to proceed in a rational fashion any further in the decision sequence.

2. *Define and clarify goals.* Given an understanding of the problem, the next step is to determine what eventual outcomes are most highly desired.

3. *Pose alternatives for the attainment of the goals.* Once there is agreement on the desired goals, various alternative courses of action must be identified or devised which will have some probability of leading to these goals.

4. *Analyze the anticipated consequences of each alternative.* After listing potential courses of action, each alternative must then be critically evaluated in terms of its end results and their desirability.

5. *Select a course of action.* The final step—that of actually selecting the most desirable alternative—is the phase we typically think of when discussing decision making. However, for Step 5 to be truly effective it should be preceded by the previous four analytical stages.

Of course, a rigid adherence to the above sequence is *not* a guarantee that the decision will be correct. It is, however, a way of assuring that the decision is as apt to be as correct as possible under the circumstances.

Schmidt (1958) has presented a generalized schema for viewing the decision situation faced by executives and managers. According to him, every decision involves a "hypothesis" about the future on the part of the decision maker. For example, an investment broker makes a hypothesis about a certain stock being apt to rise in value during the next six months. This hypothesis will subsequently turn out to be either true or false, depending upon the actual market performance of the stock in question.

Not only does the executive make a hypothesis about future events, but he also makes a decision to follow a particular course of action. In the case of the stock, he either *does* purchase some shares (a positive action) or he does *not* purchase some shares (a nonpositive action). Out of this, Schmidt develops a small 2 × 2 table to represent the four different executive decision situations (see Figure 15.1).

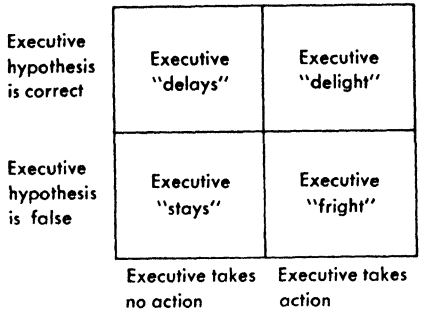


Figure 15.1. Various kinds of executive decision situations and their consequences. (From R. N. Schmidt. Executive decision making. *Academy of Management Journal*, 1958, *1*, 36–44.)

The four different decision situations which emerge from the figure are:

Executive "delight." The executive has correctly guessed the future state of things and has decided to act. Since his actions turn out to be justified he experiences delight over the outcome of his decision.

Executive "fright." The executive has incorrectly guessed the future state of things and has decided to act according to his incorrect prediction. Since his action will thus turn out to be inappropriate as he guessed wrong, this is termed executive fright.

Executive "delays." The executive has correctly guessed the future state of affairs but he decides not to take action. This simply causes a delay in his situation. Appropriate action, if eventually undertaken, will lag behind. The speculator who is sure land values will increase but who lacks the courage to take action on his conviction and buy land is the perfect example.

Executive "stays." The executive has incorrectly guessed the future state of events but has *also* failed to act on his prediction and has simply stayed in his present situation. He is "lucky" in that if he *had* acted on his hypothesis he would have been wrong.

DECISION "TREES"

A second schematical model for the decision process is suggested by Mager (1964). He proposes the process be viewed as a treelike structure, as shown in Figure 15.2. The base or "trunk" of the tree is the actual decision required of the executive.

The options or decision alternatives form the major branches leading away from the base. Sometimes there are many alternative courses of action; sometimes only two possible choices are open (the case shown in Figure 15.2). In the example, the decision maker must make a decision whether to expand his marketing into foreign markets. He has two choices—either "yes" or "no"—associated with his own decision in the problem of whether his competitors will decide on a similar strategy or not, that is, will they expand or keep the *status quo*? Since our decision maker has no control over what his competitors will do, he must view *their* decision as a "chance event." Thus, he must consider not only his own actions, but for each

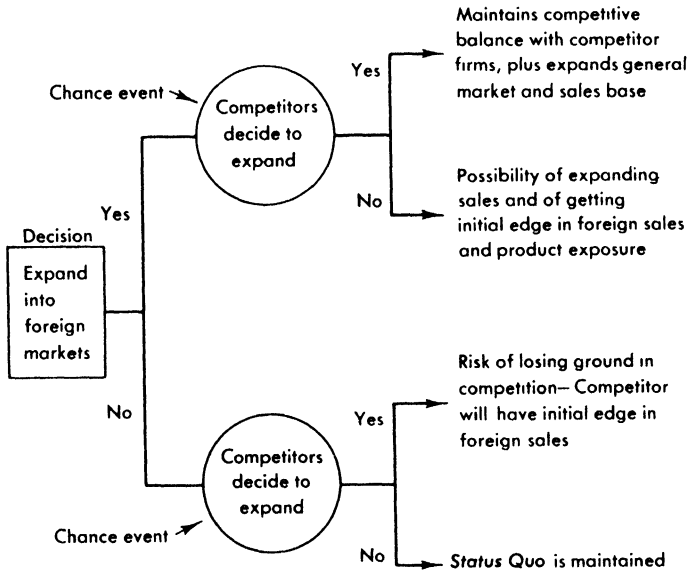


Figure 15.2. Example of a "decision-tree" diagram. (From John F. Magee. Decision trees for decision making. *Harvard Business Review*, 1964, 42, [July-August], 119–138.)

alternative open to him he must also consider what would result depending upon the actions of his competitors. The result is a series of branches upon branches ending up in all possible outcomes being listed at the right of Figure 15.2. Mager feels that this type of conceptual analysis, accompanied by formal mathematical analysis (he advocates the Bayes theorem approach discussed later in the chapter) is essential for making sound decisions.

TYPES OF DECISION MAKERS

All individuals are not going to be equally adept at making decisions. People differ in this ability just as they differ in all other characteristics. And not only do they differ in ability, but they also differ in terms of their basic strategy when it comes to dealing with decisions that involve risk, either personal or otherwise. Some individuals are exceedingly adverse to risk taking, while others seem to enjoy it immensely. Atkinson (1957) has suggested that there are two kinds of people—those whose motivation to achieve success is greater than their motivation to avoid failure, and those whose motivation to avoid failure is greater than their motivation to suc-

ceed. Those in the former category could be expected to have a tendency to "gamble" in a decision-making situation, whereas those in the latter category would probably exhibit a great deal of conservatism.

Traditionally, the economist has viewed the decision maker as a very rational type of person who responds exclusively in terms of economic value systems. Shubik (1958) has outlined the characteristics of *Homo oeconomicus*—"economic man." For economic man:

1. A decision is a conscious choice from among a set of well-defined alternatives.
2. The individual decision maker is able to attach a value to the outcomes arising from any set of alternatives.
3. The individual decision maker is motivated to behave so that the expected value of the chosen outcome is as high as possible.

Thus economic man is a person who behaves in an optimal manner in terms of his decision choices—he tries to maximize his return.

Simon (1957), however, has suggested a somewhat different type of model for man as a decision maker. He calls his "satisficing man." Satisficing man is a person who is at home in a bureaucratic environment. His decision strategy is such that he is more concerned with whether a decision alternative is feasible *and at least satisfactory*, rather than being concerned with choosing an optimal alternative. For satisficing man, an alternative is satisfactory if it meets certain minimum standards—it does not have to maximize gain.

A third type of hypothetical man has been suggested by Boulding (1958). He calls his model "heroic man." Heroic man has a decision ethic which calls for action irrespective of cost. It is the "damn the torpedoes, full speed ahead!" type of decision strategy. Boulding believes that in real life this type of decision strategy all too often transcends the philosophy of pure economic man.

The actual research which has examined the different types of strategies employed by decision makers seems to come closer to supporting the satisficing man notion of Simon than it does to supporting the other two models (Simon, 1959; Edwards, 1961). The data seem to indicate that man does *not* tend to behave the way in which an optimal decision strategy would dictate that he should behave. For example, in simple tasks where a person is asked to predict which of two lights will come on, the subject will tend to spread his predictions such that his probability of predicting light number one matches the actual probability of that light coming on, rather than adopting a strategy of always choosing the most likely light. In other words, even though past experience tells him that light 2 may be twice as likely to come on as light one, he will not always pick light 2; instead he will only predict light 2 about two-thirds of the time. This is called a probability matching strategy, whereas the optimal strategy (the one which would maximize his number of correct predictions) is to simply choose the most probable light on every trial.

RATIONAL VERSUS NONRATIONAL DECISION MAKING

Simon (1959) separated approaches to decision making into two broad categories which he calls (1) rational decision-making models and (2) nonrational decision-making models.

The rational decision-making group he refers to as the management science approach to decision making. The notion behind this approach is that it is possible to

quantify the decision-making process in terms of the utility value of various outcomes, the probability of the occurrence of various outcomes, etc., and then to develop formal mathematical systems which in turn will provide optimal decision choices. These normative models can then be used as idealized standards against which man's performance can be compared.

The nonrational decision-making models are referred to by Simon as the organismic approach. These models consider decision making in organizations to be a joint or group process and, therefore, related to and influenced by such things as organizational structure, group size, etc. The emphasis in these models is on motivational variables, needs, rewards and penalties, group processes, roles, etc. Atkinson (1957) represents the approach very well.

NORMATIVE MODELS OF INDIVIDUAL DECISION MAKING

There exist several normative models for individual decision making which differ in terms of their emphasis and complexity. The model which we will present in detail is one which has been used with a great deal of success in studying the basic characteristics of decision making. It also provides a nice conceptual framework for viewing and appreciating the decision process.

BRUNSWIK LENS MODEL

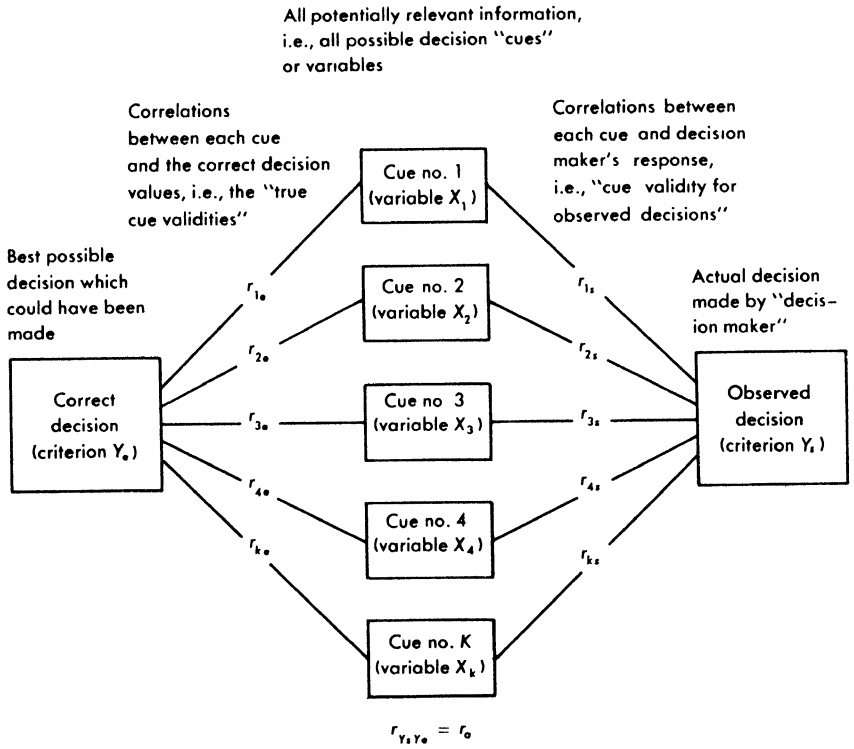
One way of viewing the decisions people make and how they go about making them is through the Lens Model of Brunswik (1956).[1] A diagram of the Lens Model is shown in Figure 15.3. The model assumes the decision process to be composed of three essential elements: (1) the basic information in the decision situation, (2) the actual decision made by the decision maker, and (3) the optimal or correct decision which should have been made in that particular situation. Each of these is shown in Figure 15.3.

BASIC INFORMATION Anytime a person makes a decision he has at his disposal a number of cues or indicators which he may or may not use as aids in the process. For example, take an executive faced each month with the problem of trying to decide how many units of product X to produce. There are obviously a wide variety of decision variables which he could *potentially* use to help him make his decision a good one, such as present inventory, current orders, general market indicators, advice from his immediate subordinates, etc. These are the potential cue variables shown in Figure 15.3.

OBSERVED DECISION Of course, any decision process must end in a response of some sort—even if the response is simply the decision *not* to make a response, it is probably safe to say that a response of sorts has been made. The *making* of a

[1] Much of the development outlined here can be found in the work of Hammond and his coworkers at Colorado (Hursch, Hammond, and Hursch, 1964; Hammond, Hursch, and Todd, 1964) along with Tucker (1964).

decision always involves a *choice* of action. Thus "decision behavior" and "choice behavior" are really quite indistinguishable phenomena. The box on the right-hand side of Figure 15.3 represents the course of action to which the decision maker finally commits himself.



Where r_a = correlation between the actual decisions made by a decision maker and the "correct" decision values. This is a measure of the *achievement* of the decision maker as he performs his decision-making function

Figure 15.3. Lens model paradigm for the decision-making process.

CORRECT DECISION Just as there is an observed course of action on the part of the decision maker, so is there an *optimal* response or choice associated with any decision. This optimal decision represents the best possible choice of action which could possibly have been selected by the decision maker in that particular situation. In a very real sense it represents the ultimate criterion against which the actual decision should be evaluated. In many decision situations it is hard to ever *really* determine or to know what this optimal decision is or was at a particular time. However, at least in theory, an optimal response on the part of the decision maker *always* exists. In Figure 15.3 this value is shown in the box on the left as the "correct" decision.

DYNAMICS OF THE MODEL Having defined the essential ingredients of the model it now becomes possible to examine the interrelationships between these elements. These interrelationships provide us with an indication of the complexity and the dynamic characteristic of the decision-making process.

True Cue Validity The true value of any single cue available to the decision maker is represented by the diagnostic or predictive "power" of that cue. In other words, how helpful is it to have that cue available during the decision process. The correlation between the cue and the correct decision, i.e., true cue validity, is the index which represents this predictive power.

TABLE 15.1 *Some Hypothetical Data Illustrating the Lens Model*

Month	*(a) Present Inventory*	*(b) Number of Units of X the Executive Decided To Produce*	*(c) Number of Units of X the Executive Should Have Decided To Produce*[a]
January	20	30	80
February	80	50	30
March	30	60	70
April	40	40	70
May	90	40	40
June	80	60	20
July	90	20	20
August	30	50	50
September	20	60	70
October	70	50	50
November	50	50	50
December	20	70	60

[a] Assuming that it is possible to look back and make this kind of assessment.

For example, take again the case of our executive who is continually faced with the problem of making a decision as to how many units of product X he should produce each month. One cue he probably would be using is the size of his present inventory. Suppose also that in looking back over the past year's records it is possible to specify, during each month, the number of X units which *should have been produced.* Table 15.1 provides a hypothetical example which shows, for each month in 1966, (a) the size of current inventory, (b) the number of X units our executive decided to produce, and (c) the number of X units which should have been produced that month.

If we plot the correlation between columns (a) and (c) as shown in Figure 15.4 we find that the trend is for low inventory values to correspond to a high number of units which should be produced. Indeed, the correlation between (a) and (c) is a *minus* 0.869! This tells us that size of present inventory is highly, but negatively, related to the number of units needed. In other words, this is an excellent cue —one that the decision maker should attend to very carefully.

Observed Cue Validity The next question we might ask about the decision process is "How well, or to what extent, did the decision maker use a given cue?" That is, given a cue which is available to him, does he tend to use it? This can be determined by examining the correlation between the cue values and what the decision maker *actually did* over a number of decisions, that is, columns (a) and (b) in Table 15.1. This correlation is also plotted in Figure 15.4, where we can see it

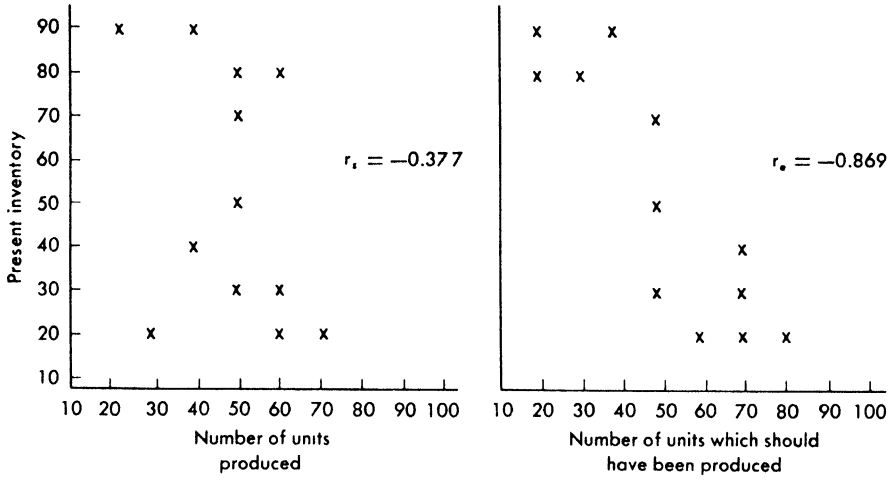


Figure 15.4. Scatterplots of decision-maker consistency (r_s) and true cue validity (r_e) based upon data given in Table 15.1.

has a value of − 0.377. Thus, our executive apparently used the cue, but not to the degree it should have been used (at least he had the direction of the true relationship estimated correctly).

Decision-Maker Achievement The third and perhaps most relevant question we should ask is the question of *how well* the decision maker performed his task. Did he have a high level of achievement insofar as the decisions he actually made were close to the decisions which, in retrospect, should have been made? This can be determined by looking at the degree of correlation between columns (b) and (c) in Table 15.1. The correlation between the number of units the executive decided to produce (column b) and the number he *should* have decided to produce (column c) turns out in our illustration to be 0.165—not very good achievement by any standard. Our decision maker is obviously not doing so well as he could with a cue which could be very helpful to him in these particular circumstances.

RESEARCH RESULTS The Lens Model is basically a descriptive conceptualization of the human decision process which provides a number of mathematical indices by which we can study the decision process in man. Most of the research based upon the model has been rather abstract laboratory research—it has not been applied in many realistic task settings. However, the research findings have indicated several rather interesting things about the ability of people to use cues in a decision-making situation, so a brief summary of these findings will be given.

First a number of studies (Schenck and Naylor, 1965, 1966; Dudycha and Naylor, 1966; Summers, 1962; and Peterson, Hammond, and Summers, 1966) have all shown that decision makers can learn to use cues appropriately. That is, they tend to learn which cues are *good* and which are *bad* and to give the good cues more attention than they do the poor cues. However, the Dudycha-Naylor study showed the very interesting finding that if a decision maker has a very good cue and then you give him a second cue which is poorer but still has *some* additional predictive value, his performance will decrease—poorer achievement results than if he just

had the single cue! Apparently poor cues add more static or "noise" to the decision-making process than they add predictive value. On the other hand, if the initial cue is only average in its predictive power and you give the decision maker a second, very good cue, his performance improves markedly.

Another interesting finding was recently reported by Clark (1966). He showed that cues with negative validity are not as useful to a decision maker as are cues having a direct or positive relationship. For some reason humans seem to have a more difficult time learning to use as aids information sources which give negative validity. The reader will remember that for *predictive* purposes the *sign* of a relationship is not important, that is, a cue with a validity of -0.80 is just as useful, potentially, as a cue having a validity of $+0.80$.

Other information which has been obtained about human decision makers using the lens model is (1) humans are better at learning to use cues which have *linear* relationships to the correct decision than they are at using cues which have a non-linear relationship (Dickinson and Naylor, 1966; Hammond and Summers, 1965) and (2) humans tend to use cues systematically even when the cues may possess no real predictive power whatsoever (Dudycha and Naylor, 1966). This latter finding simply means that if a decision maker is placed in a situation where none of the cues available to him is of any value, he will *still* tend to pick out and use some of them as if they did have value.

BAYES MODEL OF DECISION MAKING

Another mathematical model currently coming into increased use in the study of human decision making is known as *Bayes Theorem*. This is as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

where
$P(A|B)$ = probability of A given that B has occurred
$P(B|A)$ = probability of B given that A has occurred
$P(A)$ = probability of A
$P(\bar{A})$ = probability of not A, i.e., $1 - A$
$P(B|\bar{A})$ = probability of B given not A

Since expressions like Bayes Theorem often tend to be confusing, let us consider an example of a practical decision task and see how Bayes Model might apply.

One kind of typical decision-making task faced by all firms is that of deciding whom to select and whom to reject from a pool of job applicants. Consider the situation where a firm has decided to try out a new selection test. Consider further that experience has shown that only 60 percent of the employees applying actually turn out to be satisfactory. Also suppose that the practice of the company in the past was to hire everyone and give them a chance at working out.

Of the men who *do* turn out to be satisfactory, 80 percent have been found to be above a cutoff score on the new selection test, while only 40 percent of those who turn out to be unsatisfactory score above the cutoff. Now, if we use this test for selection, and if we only hire those men above the cutoff score, *what is the probability that a person above cutoff will turn out to be satisfactory*?

If we now define our symbols again, we have:

$P(A)$ = probability of being successful = 0.60
$P(B)$ = probability of passing test
$P(B|A)$ = probability of passing test given the employee is successful = 0.80
$P(B|\bar{A})$ = probability of passing test given the employee is unsuccessful = 0.40
$P(\bar{B}|A)$ = probability of not passing test given the employee is successful = 0.20
$P(\bar{B}|\bar{A})$ = probability of not passing test given the employee is unsuccessful = 0.60

We want to know $P(A|B)$, that is, the probability that a person will be successful *given* that he has passed the test. Bayes Theorem shows:

$$P(A|B) = \frac{(0.80)\,(0.60)}{(0.80)\,(0.60) + (0.40)\,(0.40)}$$
$$= \frac{0.48}{0.48 + 0.16} = 0.75$$

In other words, if we select only those who pass our screening test, we will end up with 75 percent successes in hiring, as compared to the 60 percent figure without the test.

The application of Bayes Theorem to decision making in industry is becoming more frequent. It is a very powerful tool and its use should increase greatly in future years.

SOME BASIC DECISION VARIABLES

There are several concepts or terms common to the area of decision behavior which are quite critical and important to a better understanding of the basic process. In particular, the terms *probability*, *utility*, *accuracy*, and *validity* are central to understanding the basic decision process. Only a brief exposition of each of these will be presented here—enough, it is hoped, to provide an idea of the meaningfulness and usefulness of each term as it applies to how people make decisions and how these decisions can be studied and evaluated.

PROBABILITY

In order to discuss probability as it applies to decision making, we need to consider a decision as "the process of making a *choice* among a set of alternatives." Each alternative may or may not turn out to be the *correct* choice on any given decision. For example, consider the simple act of tossing a coin and asking a friend to make a decision about whether it will fall heads or tails. The decision maker has two alternative choices, and on any given decision (toss) either may or may not be correct. Let

P_1 = probability of head = 0.5
P_2 = probability of tail = 0.5

that is, assume we have an honest coin and an honest coin tosser. P_1 and P_2 are the true or real probabilities associated with the various possible alternatives being correct on any single decision. Such probabilities are usually referred to as *objective*

probabilities. Objective probability differs from *subjective* probability, which is the probability that the decision maker *himself* associates with each outcome. The two probabilities may, in certain cases, be quite different. Consider the example of asking your friend to tell you what the probability of a head is on the next toss of a coin after he has seen heads come up five times in succession. He would probably still say $P = 0.5$. But then ask him to predict which will occur on the next coin toss and the chances are considerably greater than 0.5 that he will say tails! In other words, in spite of the fact that he knows objectively that a head is just as likely to occur on trial six as before, he still feels *subjectively* that after five heads a tail is long overdue. This kind of behavior is known as the "*gambler's fallacy*."

UTILITY OR VALUE

Given a decision situation which has a specifiable number of possible outcomes, each outcome also has a "payoff" associated with it. In the case of a coin-tossing game, the two possible outcomes associated with any decision or guess are "correct" or "incorrect." If the game is being played for money, the individual might win five cents every time he is correct and lose five cents every time he is incorrect. Thus, the value or *utility* of a correct decision is + 5 cents, while the value or *utility* of an incorrect decision is − 5 cents. However, it is important to point out that utility as measured in objective units such as money does not necessarily correspond to utility on a subjective or personal basis. Very often the subjective utility of an outcome can be noticeably different from the objective utility.

AN EXAMPLE

Perhaps an example may serve to clarify matters. The following illustration is taken with some modification from *Introduction to Statistics for Business Decisions* by Robert Schlaifer (1961, p. 3):

> *An Inventory Problem* A retailer is about to place an order for a number of units of a perishable commodity which spoils if it is not sold by the end of the day on which it is stocked Each unit costs the retailer $1; the retail price is $5. The retailer does not know what the demand for the item will be, but he must, nevertheless, decide on a definite number of units to stock.

This is a typical business decision problem. It has two essential characteristics:

1. The decision maker must choose among several alternative courses of action, that is, he must select one of several possible alternatives.
2. The chosen alternative will ultimately result in some definite payoff. This payoff can be either positive or negative in value.

From the above information it is possible to construct what is known as a "payoff table" which illustrates the monetary outcome which occurs for various combinations of chosen alternatives and actual outcomes.

What is the best "strategy" for the decision maker to follow? Is one choice a "better" choice than any of the others? One way of deciding which alternative to select is known in decision making as the *Minimax* principle. The minimax rule says that one should select the alternative which "minimizes the maximum possible loss." This is a very conservative type of decision rule which serves to protect the decision maker against any large adverse outcome. However, in many cases it also prevents large favorable outcomes from occurring. Note from Table 15.2 that if

TABLE 15.2 *Payoff Table for Inventory Problem*

Outcome	*Decision Alternative (Number of Units Actually Stocked)*					
(Number Demanded)	*0*	*1*	*2*	*3*	*4*	*5*
0	$0	−$1	−$2	−$ 3	−$ 4	−$ 5
1	0	+ 4	+ 3	+ 2	+ 1	0
2	0	+ 4	+ 8	+ 7	+ 6	+ 5
3	0	+ 4	+ 8	+ 12	+ 11	+ 10
4	0	+ 4	+ 8	+ 12	+ 16	+ 15
5	0	+ 4	+ 8	+ 12	+ 16	+ 20

SOURCE: R. Schlaifer. *Introduction to statistics for business decisions*. McGraw-Hill, New York, 1961, 5. © The President and Fellows of Harvard College.

we follow a minimax strategy we should select alternative 1, that is, stock *no* units whatsoever! If we do this we can be sure that we will never *lose* money. But neither will we ever make money—a rather silly alternative to select.

WEIGHTING THE OUTCOME In a very real sense, the minimax principle assumes that the least favorable outcome has a very high probability of occurring. Thus we ought to protect ourselves against this eventuality. In our inventory problem, the most unfavorable outcome would be to have *no* units purchased.

A more realistic decision strategy would be to weight each outcome by the estimated probability that the particular outcome will indeed occur. By doing this it becomes possible to make an appraisal of how good each decision alternative is, *given* that any of the possible outcomes is likely to occur with some specified probability. These probabilities may either be subjective or objective (based upon prior experience and knowledge).

For example, suppose our retailer assumes that each of the six possible outcomes is equally likely. That is, on any given day he is just as apt to have four units demanded as he is no units, etc. In tabular form, we could write his expectations out as follows:

Units Demanded	*Expected Probability*
0	1/6=0.1666
1	1/6=0.1666
2	1/6=0.1666
3	1/6=0.1666
4	1/6=0.1666
5	1/6=0.1666

Once the expected *probabilities* have been determined for each outcome, and if the *value* of each outcome has also been specified under each decision alternative, it is now possible to determine the *optimal* strategy or decision alternative. The formal reasoning process for doing so goes as follows (Schlaifer, 1961, p. 6):

1. Attach a definite numerical *value* to the consequence of every possible act given every possible event.
2. Attach a definite numerical *weight* to every possible event.

3. For each act separately, use these weights to compute a *weighted average* of all the values attached to that act.
4. Select the act whose weighted-average value is highest.

This weighted average across all outcomes for any given alternative is what is known as the *expected value* of an alternative. To illustrate, we shall compute the expected value for each of the six different decision alternatives available to our retailer.

Alternative No. 1 (no units are stocked):

Outcome	*Value of Outcome (V)*	*Probability of Outcome (P)*	(P)(V)
0	\$0	1/6	\$0
1	0	1/6	0
2	0	1/6	0
3	0	1/6	0
4	0	1/6	0
5	0	1/6	0
			Sum = \$0
			Average = \$0 = Expected value of alternative no. 1

Alternative No. 2 (one unit is stocked):

Outcome	*Value of Outcome (V)*	*Probability of Outcome (P)*	(P)(V)
0	−\$1	1/6	−\$.17
1	+ 4	1/6	.67
2	+ 4	1/6	.67
3	+ 4	1/6	.67
4	+ 4	1/6	67
5	+ 4	1/6	67
			Sum = \$3.18
			Average = \$ .53 = Expected value of alternative no. 2

Alternative No. 3 (two units are stocked):

Outcome	*Value of Outcome (V)*	*Probability of Outcome (P)*	(P)(V)
0	−\$2	1/6	−\$ .33
1	+ 3	1/6	.50
2	+ 8	1/6	1.33
3	+ 8	1/6	1.33
4	+ 8	1/6	1.33
5	+ 8	1/6	1.33
			Sum = \$5.49
			Average = \$ .91 = Expected value of alternative no. 3

Alternative No. 4 (three units are stocked):

Outcome	*Value of Outcome (V)*	*Probability of Outcome (P)*	(P)(V)
0	−$ 3	1/6	−$ .50
1	+ 2	1/6	.33
2	+ 7	1/6	1.17
3	+ 12	1/6	2.00
4	+ 12	1/6	2.00
5	+ 12	1/6	2.00
			Sum = $7.00
			Average = $1 17 = Expected value of alternative no. 4

Alternative No. 5 (four units are stocked):

Outcome	*Value of Outcome (V)*	*Probability of Outcome (P)*	(P)(V)
0	−$ 4	1/6	−$ .67
1	+ 1	1/6	.17
2	+ 6	1/6	1.00
3	+ 11	1/6	1.83
4	+ 16	1/6	2.67
5	+ 16	1/6	2.67
			Sum = $7.67
			Average = $1.28 = Expected value of alternative no. 5

Alternative No. 6 (five units are stocked):

Outcome	*Value of Outcome (V)*	*Probability of Outcome (P)*	(P)(V)
0	−$ 5	1/6	−$ .83
1	0	1/6	.00
2	+ 5	1/6	.83
3	+ 10	1/6	1.67
4	+ 15	1/6	2.50
5	+ 20	1/6	3.33
			Sum = $7.50
			Average = $1.25 = Expected value of alternative no. 6

Notice that alternative number 5, which calls for stocking four units, has the highest expected value of any of the choices available to the decision maker. This tells us that his best strategy is to choose this alternative *if indeed each of the outcomes is equally likely to occur on any given day!* The reader should keep in mind that if the probabilities were different, e.g., if the outcome of five units demanded had a probability of ¼ rather than 1/6, then the optimal strategy will in all likelihood change. We suggest that the reader try using a different set of probability values to demonstrate this fact to himself.

A SUMMARY OF RULES FOR COMPETITIVE DECISION MAKING

Friedman (1960, pp. 91–92) has summarized the steps or rules to follow in a competitive decision-making situation.

1. Determine the underlying rules of the game and the extent of the restrictions upon all the competitors.
2. Describe in as much detail as possible the particular situation in which competitive decisions are to be made by the competitors.
3. List all the alternative courses of action which are available to the competitors.
4. Calculate or estimate a final utility to each of the competitors as a result of the selection of the various alternatives.
5. Study individual competitors and their past behavior to determine possible strategies they might use in arriving at their decisions in the particular situation being studied.
6. Estimate a priori probabilities that the competitor used any given strategy, and a posteriori probabilities that the observed events would have occurred with the given strategies.
7. Calculate the final relative probabilities for the various hypotheses concerning competitive strategies.
8. Estimate the probabilities of a competitor using his alternative courses of action in the situation being studied.
9. Given the estimated probabilities of the competitor's use of his alternative courses of action, choose the alternative course of action which maximizes expected utility

ACCURACY VERSUS VALIDITY OF DECISIONS

Earlier in the chapter the basic elements of the Lens Model were presented. The model was shown to be a useful analytical device for analyzing the decision process itself. The Lens Model is a correlational model in that the correlation coefficient is used as an index of how well a person is performing. Thus, the correlation between a person's response and the correct response is the *validity* of a decision maker over time—it is an index of his achievement. One weakness in the correlation index as a measure of decision-maker achievement needs to have brief mention, however.

The problem which arises is that it is possible for a decision maker to have extremely high *validity* but at the same time be highly inaccurate in his predictions. For example, suppose we had the following situation:

Predictions of Decision Maker	*Correct Decision*	*Error*
8	6	2
10	8	2
7	5	2
9	7	2
11	9	2
5	3	2
14	12	2
13	11	2
19	17	2
12	10	2

Our hypothetical decision maker appears to be consistently *overestimating*—he has a kind of constant error to his decisions. However, the correlation between his decisions and the correct decisions is + 1.00—perfect validity. Thus, the correlational index really only tells us how well the predictions of the decision maker *co-vary* with the correct decisions and do not directly tell us how accurate he is.

A COGNITIVE MODEL

The models of decision making just described are what we might call "mathematical" models. They provide a way of formalizing the elements of the decision process, a way of stating or expressing what optimal decision performance should be, and with quantitative indices for the evaluation of decision behavior. Another kind of decision-theory approach is the "cognitive" model, which views the process of decision making in terms of the psychological processes felt to be involved. One such model is the Cognitive Dissonance Model proposed by Festinger (1957).

Actually it is inappropriate to call Festinger's Theory of Cognitive Dissonance a model for decision making. It is a much more general theory of human behavior. However, it has, as we will see, great relevance to the decision process and in particular to the *post*-decision behavior of individuals.

BASIC OUTLINE OF DISSONANCE THEORY

Following is a basic outline of the theory of Cognitive Dissonance as it relates to the decision process. The brief outline given here is merely a small "chunk" of the total theory outlined by Festinger. Those interested in the complete theory should examine the 1957 book.

Festinger (1957, pp. 260–261) summarized the background for his theory in the following paragraph:

> The basic background of the theory consists of the notion that the human organism tries to establish internal harmony, consistency, or congruity among his opinions, attitudes, knowledge, and values. That is, there is a drive toward consonance among cognitions. In order to deal with this notion in a somewhat more precise manner, I have imagined cognition to be decomposable into elements or, at least, clusters of elements. The following theoretical statements have been made about the relations among these cognitive elements:
>
> 1. Pairs of elements can exist in irrelevant, consonant, or dissonant relations.
> 2. Two cognitive elements are in an irrelevant relation if they have nothing to do with one another.
> 3. Two cognitive elements are in a dissonant relation if, considering these two alone, the obverse of one element follows from the other.
> 4. Two cognitive elements are in a consonant relation if, considering these two alone, one element follows from the other.

The relevance of dissonance to the decision-making act is then pointed out by Festinger (1957, p. 261) as follows:

Dissonance almost always exists after a decision has been made between two or more alternatives. The cognitive elements corresponding to positive characteristics of the rejected alternatives, and those corresponding to negative characteristics of the chosen alternative, are *dissonant* with the knowledge of the action that has been taken. Those cognitive elements corresponding to positive characteristics of the chosen alternative and negative characteristics of the rejected alternative are *consonant* with the cognitive elements corresponding to the action which has been taken.

Concerning the magnitude of dissonance which exists in any given situation, he suggests (1957, p. 262):

1. The magnitude of the dissonance or consonance which exists between two cognitive elements will be a direct function of the importance of these two elements.
2. The total magnitude of dissonance which exists between two clusters of cognitive elements is a function of the weighted proportion of all the relevant relations between the two clusters which are dissonant, each dissonant or consonant relation being weighted according to the importance of the elements involved in that relation.

Starting with these hypotheses about the magnitude of dissonance, a number of operational implications seem clear.

1. The magnitude of postdecision dissonance is an increasing function of the general importance of the decision and of the relative attractiveness of the unchosen alternatives.
2. The magnitude of postdecision dissonance *decreases* as the number of cognitive elements corresponding identically to characteristics of chosen and unchosen alternatives increases.

Concerning the *reduction* of postdecision dissonance, Festinger (1957, p. 264) suggests that:

1. Postdecision dissonance may be reduced by increasing the attractiveness of the chosen alternative, decreasing the attractiveness of the unchosen alternatives, or both
2. Postdecision dissonance may be reduced by perceiving some characteristics of the chosen and unchosen alternatives as identical.
3. Postdecision dissonance may be reduced by decreasing the importance of various aspects of the decision.

In summary, the dissonance theory states that the making of a decision is not the final step in the decision process. Instead, making a decision establishes some degree of disequilibrium with which the decision maker must cope. This need to cope with the created dissonance leads to predictable changes in behavior (and attitudes) on the part of the decision maker.

ORGANIZATIONAL DECISION MAKING

One aspect of the growing interest in the study of complex organizations on the part of psychologists has been their attention to the theme of *organizational* decision making. Indeed, decision making has, to some extent, become one of the central variables in modern organizational theory. Unfortunately, as Leavitt and Bass (1964) point out, there has not been complete agreement among organizational theorists as to the definition of decision making. In economics, however, the firm itself has classically been viewed as an individual decision maker. Decision rules have then

been developed for the firm to try and assure successful (profitable) operation in a specified economy and with specified competition. However, in general the decision situations to which these classical models can be applied are rather limited due to the difficulty in accurately specifying the characteristics of economy and the competition.

Recently there seems to be a great amount of interest in taking some of these classical decision models for situations like a bilateral monopoly and examining how they can be affected by such psychological variables as level of aspiration, amount and kind of information available to the bargainers (decision makers), and the value systems of the bargainers. These modified models of organizational decision making appear to hold much more promise than did the earlier classical models.

Leavitt and Bass (1964) cite three major issues which seem to lie at the center of the organizational decision-making problem. They call these (1) the *feasibility* issue, (2) the *normative-descriptive* issue, and (3) the *algorythmic-heuristic* issue.

Feasibility Issue: Basically, the argument here is that formal decision-making models lack feasibility because of the tendency of humans to be reluctant to give up the prerogative of having the final say in any decision.

Normative-Descriptive Issue: Should "optimal" decision-making strategies be developed to (1) implement directly or (2) serve as a standard toward which one attempts to get to aspire (the normative approach)?

Algorythmic-Heuristic Issue: The algorythmic approach is to use a standardized, guaranteed decision-making routine. The heuristic approach is to use simply gross strategies and simplifying approximate rules. What types of decision problems are best attached with these two alternatives? Evidence so far indicates that the more complex the decision problem, the more difficult it becomes to successfully apply the algorythmic method.

In summary, the problem of individual and organizational decision making appears to be one of the most promising and viable areas of research concerning man and his interaction with his environment.

REFERENCES

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review*, 1957, 64, 359–372.

Boulding, K. E. *Skills of the economist*. Howard Allen, Cleveland, 1958.

Brunswik, E. *Perception and the representative design of psychological experiments*. University of California Press, Berkeley, 1956.

Calkins, R. D. The decision process in administration. *Business Horizons*, 1959, 2, 19–25.

Clark, R. D. Decision task performance utilizing positively and negatively related information. Unpublished M.A. thesis, Ohio State University, Columbus, 1966.

Dickinson, T. L., and J. C. Naylor. Effects of task complexity and organization on team behavior using the multiple regression model. Paper presented at Midwestern Psychological Association, May 1966.

Dudycha, Linda, and J. C. Naylor. Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1966, 1, 110–128.

Edwards, W. Behavioral decision theory *Annual Review of Psychology*, 1961, 12, 473–498
Festinger, L A. *Theory of cognitive dissonance* Stanford University Press, Stanford, Calif, 1957.
Friedman, L. Decision making in competitive situations *Management Technology*, 1960, 1, 85–93.
Hammond, K R., C. J Hursch, and F J Todd Analyzing the components of clinical inference. *Psychological Review*, 1964, 71, 438–456
Hammond, K. R, and D A. Summers Cognitive dependence upon linear and nonlinear cues. *Psychological Review*, 1965, 72, 215–224.
Hursch, C. J, K R Hammond, and J L. Hursch. Some methodological considerations in multiple cue probability studies. *Psychological Review*, 1964, 71, 42–60
Larson, R. L How to define administrative problems *Harvard Business Review*, 1962, 40, 68–80
Leavitt, H J, and B M Bass. Organizational Psychology *Annual Review of Psychology*, 1964, 15, 371–398
Mager, J. F. Decision trees for decision making *Harvard Business Review*, 1964, 42, 119–398
Peterson, C. R., K R. Hammond, and D A Summers Optimal responding in multiple cue probability learning. *Journal of Experimental Psychology*, 1965, 70, 270–276
Schenck, A, and J C. Naylor. Some data concerning performance indices based upon the multiple regression model when applied in a standard multiple cue situation Paper presented at Midwestern Psychological Association, April 1965
Schenck, E A, and J C. Naylor. The effect of cue intercorrelation upon performance in a multiple cue situation Paper presented at Midwestern Psychological Association, May 1966
Schlaifer, R *Introduction to statistics for business decisions* McGraw-Hill, New York, 1961
Schmidt, R N Executive decision making *Academy of Management Journal*, 1958, 1, 36–44.
Shubik, M Studies and theories of decision-making *Administrative Science Quarterly*, 1958, 3, 289–306
Simon, H A *Models of man Social and rational* Wiley, New York, 1957
Simon, H. A Theories of decision-making in economics and behavioral science. *American Economic Review*, 1959, 49, 253–283
Summers, S A. The learning of responses to multiple weighted cues *Journal of Experimental Psychology*, 1962, 64, 29–34
Tucker, L R. A suggested formulation in the development by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd *Psychological Review*, 1964, 71, 528–530.

16 ORGANIZATIONS

No trend has been more dynamic in industrial psychology during the past decade than the growth of interest in organizations. Indeed, in 1961 Leavitt suggested that it was time to consider changing the name of industrial psychology to "organizational psychology." To quote Leavitt (1961, p. 7):

> Organizational Psychology occupies itself with the study of organizations and organizational processes. It is as much descriptive as normative; as much or more basic as applied; as much interested in developing theories of organizational behavior as ways of improving organizational practice. It is close on one of its boundaries to human relations psychology, especially the social-experimental part of that field. It is also touched with experimental psychology and the rapidly growing field of empirical sociology. Its other boundaries are quite different; predominately economics and mathematics—the game theorists, the operations researchers, the computer people.

Tannenbaum and Seashore (1965, p. 1) have also pointed out this change of emphasis in industrial psychology which seems to have taken place during the last fifteen or so years.

> Significant changes have occurred in several broad areas of our work. First, and perhaps most basic, has been the emergence of a concern for the larger organizational context within which the psychologist may study people in organizations. We now see important psychological processes within organizations—perceptions, attitudes, motivations and interpersonal relations among members—as being very much affected and conditioned by the larger organizational structure.

WHAT IS AN ORGANIZATION?

Several years ago one of your authors was invited to participate in a ten-week continuing symposium dealing with the study of organizations. The group of participants, who met twice each week for a three-hour period, consisted of scientists from such diverse areas as sociology, psychology, political science, anthropology,

economics, and mathematics. Each felt he had something unique to contribute and that he could benefit greatly from finding out the points of view held by his brethren in other disciplines. Being good and objective scholars, we decided at the first meeting that the initial order of business should be the definition of what was meant by the term "organization." Unless we had a common conception of the basic term, it would be difficult to exchange ideas, so we decided to hammer out a definition that would be acceptable to all concerned. Needless to say, ten weeks later the group was still hopelessly bogged down in the process of trying to agree on what were the essential characteristics of an organization. Everyone knew what *he* meant by the term organization, but no two people had similar conceptions.

One can, however, cite two essential characteristics of organizations. Every organization must have a *structure* and a *purpose*. Organizations may be classified in many ways, for example, there are social organizations, political organizations, military organizations, commercial organizations, religious organizations, etc. Bakke (1959, p. 37) has referred to what he calls the *Organizational Charter*. He states "It is essential that the organization as a whole means something definite, that the name of the organization calls to mind unique, identifying features. This image and its content we label the *Organizational Charter*." He then lists the basic identifying features of the organization which comprise the charter.

1. The name of the organization.
2. The function of the organization in relation to its environment and its participants.
3. The major goal or goals toward the realization of which the organization, through its system of activities, is expected by participants to employ its resources (including themselves).
4. The major policies related to the fulfilling of this function and the achievement of these major goals to which agents of the organization are committed.
5. The major characteristics of the reciprocal rights and obligations of the organization and its participants with respect to each other.
6. The major characteristics of the reciprocal rights and obligations with respect to each other of the organization, and people and organizations in the environment.
7. The significance of the organization for the self-realization of people and organizations inside and outside the organization in question.
8. The value premises legitimizing the function, goals, policies, rights and obligations, and significance for people inside and outside the organization.
9. The symbols used to clarify, focus attention upon, and reinforce the above, and to gain acceptance from people inside and outside the organization. These symbols are actually particular items of the several basic resources which serve as cues to bring to mind the content of the Organizational Charter and reinforce its hold upon the minds of both participants and outsiders.

ORGANIZATIONAL STRUCTURE

Everyone is familiar with the term "organizational chart." This refers to a schematic representation of the formal properties and interrelationships of organizational components. It is a representation of the way in which the organization is

indeed "organized." Questions such as "who reports to whom?" and "who has responsibility for what?" are made explicit by the organizational chart.

Nearly all organizations have a structure which when charted looks like a pyramid. This organizational pyramid-type structure is shown in Figure 16.1.

Each box in the organizational chart represents an organizational position or *unit*. These units are connected by organizational *linkages*. The linkage pattern between the units is what provides the structure to the entire organizational system.

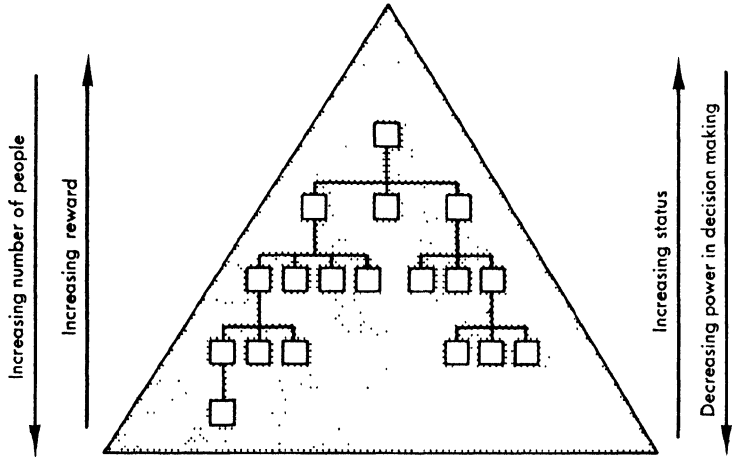


Figure 16.1. The organizational pyramid. (From Leonard R. Sayles and George Strauss. *Human behavior in organizations.* Prentice-Hall, Englewood Cliffs, N.J., © 1966. Reprinted by permission of the publisher.)

In its pure form, the links between units are supposed to represent the relationships that exist between units. If no direct link is shown between units, then no direct relationship is assumed to exist between them. Thus, the pyramid-type linkage assumes the following kinds of relationships between units (Sayles and Strauss, 1966, p. 349):

1. Nearly all contacts take the form of orders going *down* and reports of results going *up* the pyramid.
2. Each subordinate must receive instructions and orders from only one boss.
3. Important decisions are only made at the top of the pyramid
4. Each superior has only a limited "span of control," that is, he supervises only a limited number of individuals.
5. An individual at any level (but the top and bottom) has contact only with his boss above him and his subordinates below him.

This, of course, is the formal schematic of an organization which has traditionally held the implicit assumption that the only significant contacts or linkages in an organization are those which occur between superiors and subordinates—a view which is both dangerous and parochial. Recent research has begun to show that lateral, i.e., horizontal, linkages may be of extreme importance in the successful operation of any organization, especially as the concept of teamwork emerges.

LINKAGE SYSTEMS IN ORGANIZATIONS

Dubin (1959) has pointed out that the linkage pattern among organizational units may be classed as falling into three major types. Of course, more complex linkage patterns can, in turn, be created by combining these three basic linkage systems. The three basic patterns are (1) serial linkage, (2) radial linkage, and (3) circular linkage.

SERIAL LINKAGE

A serial linkage pattern is one in which each unit is linked only with the adjacent units on each side. It is the simplest type of linkage system. In a "pure" serial linkage system the number of links will always be given by

$$L = U - 1$$

where L = Number of links
U = Number of organizational units

Figure 16.2 shows a serial linkage system. Serial linkages can be either horizontal or vertical in nature. If they are vertical, each unit represents a different level of authority, that is, we have represented the typical "chain-of-command" notion. If the serial linkage is horizontal, then all units are of equal authority but of different function, such as an assembly line system in which the workers are at the same "level" but each has a different function to perform.

RADIAL LINKAGE

A pure radial system is one in which a number of peripheral units are each linked to a central unit but *not* to each other. This central unit is called the "hub" of the system or the "radial focus" unit. In such a system the number of links is given by

$$L = R(U - R)$$

where L = Number of links
R = Number of radial focus units
U = Total number of units in the system

As in the case of serial systems, radial systems may be either vertical or horizontal in nature (see Figure 16.2). A vertical radial system is one in which the peripheral units are subordinate to the radial focus unit, as in the case of a project coordinator. In a horizontal radial system, the hub unit may be a service unit which can interact with any of a number of line units (for example, the payroll department).

CIRCULAR LINKAGE

The third type of pure system is the circular linkage pattern. It is probably the least common of the three as far as its frequency in real organizational contexts. In

a circular linkage system all units are linked to all other units. The number of links in a circular system is given by

$$L = \frac{U(U-1)}{2}$$

where L = Number of links
U = Total number of units in the system

Figure 16.2 shows a diagram of a pure circular linkage system. Real-life examples of such systems might be an advisory board or a governing board of a large firm.

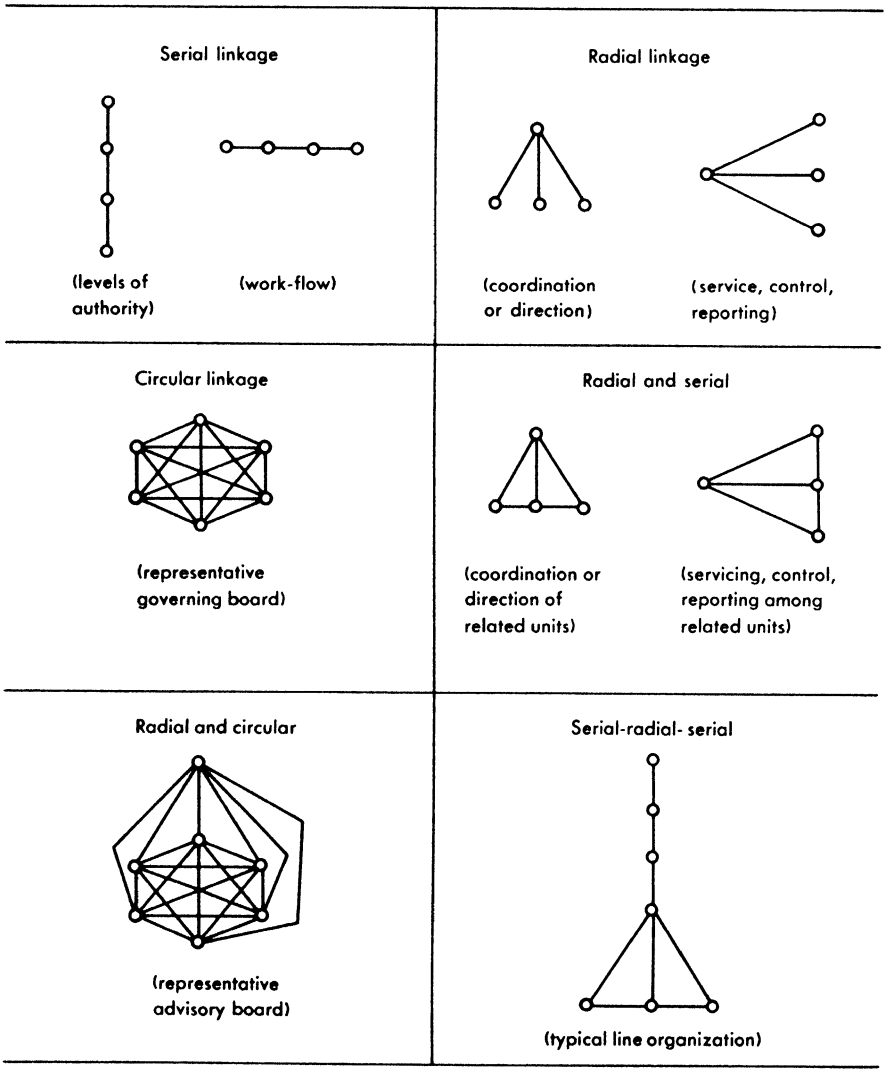


Figure 16.2. Unit linkage systems in organizations. (From R. Dubin. Stability of human organizations. In M. Haire, ed. *Modern organizational theory*. Wiley, New York, 1959.)

COMPOUND LINKAGE PATTERNS

As mentioned earlier, most organizational patterns in real organizations are probably *not* pure patterns, but are composed of some combination of these basic systems. These more complex systems are referred to as compound linkage systems, and some of these are of sufficient importance to mention briefly.

RADIAL-SERIAL Figure 16.2 shows a combination of a radial and a serial system. The number of links in this system having one serial and n radial-focus units is

$$L = R(U - R) + (S - 1)$$

where L = Number of links
R = Number of radial focus units
S = Number of serial units
U = Total number of units

but since $U = R + S$
then $U - R = S$
and $L = RS + (S - 1)$
$= S(R + 1) - 1$

As Dubin points out (1959), an excellent example of a vertical radial-serial system is the situation where one unit serves as coordinator for a number of other serial units, such as the case of a cost-accounting department linked to a number of serially tied production departments.

RADIAL-CIRCULAR A second compound pattern is the combination of the pure radial and the pure circular as shown in Figure 16.2. The number of links in this type system is

$$L = R(U - R) + \frac{C(C - 1)}{2}$$

where L = Number of links
R = Number of radial focus units
C = Number of units in circular system
U = Total number of units

again, since $U = R + C$
then $U - R = C$

and thus
$$L = RC + \frac{C(C - 1)}{2}$$
$$= \frac{2RC + C(C - 1)}{2}$$
$$= \frac{C(2R + C - 1)}{2}$$

This latter formula, then, gives the number of linkages among n radial focus units and *one* circular system. An example of such a compound system would be advisory board members acting under an advisory board chairman.

SERIAL-RADIAL-SERIAL A very typical structure is the linkage of a horizontal *and* a vertical serial through a radial as shown in Figure 16.2. The number of links in this type of system is

$$L = S_1 + S_2 (R + 1) - 1$$

where L = Number of links
S_1 = Number of units in first serial
S_2 = Number of units in second serial
R = Number of serial focus units

This type of structure is typical of the usual type of line organization.

RELEVANCE OF LINKAGE ANALYSES

Dubin (1959) has suggested that in designing organizations for maximum stability that the structure (linkage pattern) is of extreme importance. That is, given a fixed set of units necessary in an organization, the number of links will vary tremendously as a function of the type of linkage system employed. Dubin further suggests *the fewer the number of links among the units, the more stable (and thus the better) the system will be.*

What about the assumption that minimizing the number of linkages among organization units contributes to system stability? We could argue from a principle of parsimony and accept the principle as sufficient proof of our law relating stability to number of linkages. There is, however, some theoretical and empirical support for the law. For example, Dubin (1959, p. 227) writes:

> Given the problem of simultaneous stimuli being received by an organization unit or person (e.g., orders, plans, requests for information, services, the exercise of controls), how are the stimuli sorted by the receiver in order to provide the bases for response? Is there the potentiality of "stimuli-saturation," so that with a sufficiently large number of simultaneous stimuli it becomes difficult or impossible to respond to all (either because the stimulating cues become confused, or because the time available for response is insufficient to make all the required responses)? In short, can there be too many linkages between one organization unit and others so that simultaneous activation of these channels of linkage may make it impossible to evoke efficient, legitimate, or timely responses? The answers to these questions illuminate the "linkage problem" for organizations.

LINKAGE STRUCTURE AND ORGANIZATIONAL EFFECTIVENESS

Is there any real evidence that Dubin's notion of minimizing linkages is an accurate one? Do those organizations which have fewer links actually perform more effectively? A rather respectable amount of research evidence is now accumulating which seems to indicate that this may indeed be the case. Although most of this research has used individuals as organizational units and thus might more appropriately be called "small group" rather than "organizational structure" research, it nevertheless has an important bearing on the linkage problem as presented by Dubin.

The original research on linkage structure was done by Bavelas and his fellow workers (1950) at Massachusetts Institute of Technology in the early 1950s. How-

ever, the first systematic study was reported by Leavitt (1951). Figure 16.3 shows the different organizational structures used in Leavitt's study. His *circle* pattern is a closed series structure; his *chain* pattern is an open series; his "Y" pattern is a radial-series compound pattern; and his *wheel* structure is a pure radial system. When groups of individuals were organized in these different ways and then given problems to solve, Leavitt found noticeable performance and morale differences due to structure. The pure radial (wheel) and the radial-series ("Y") organizations per-

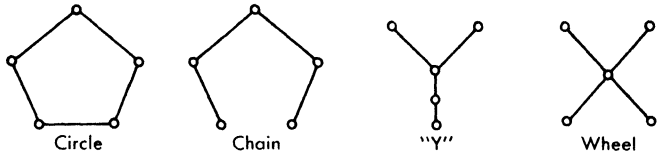


Figure 16.3. Organizational networks used in Leavitt's study. (From H. J. Leavitt. Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 1951, 46, 42.)

formed much more efficiently than the chain and the circle, which proved quite ineffective—particularly the circle. On the other hand, the individuals working in the circle networks reported much greater satisfaction with the work task. Thus, they liked the work better but did more poorly. Leavitt developed indices of *centrality* and *peripherality* for each group and showed that group or organizational stability was directly related to these measures. That is, groups which had one person in a very central position and others in peripheral positions were the most effective—leaders emerged sooner and the group got itself organized faster. However, the groups in which all members had about the same degree of centrality were the groups in which there was the greatest amount of morale.

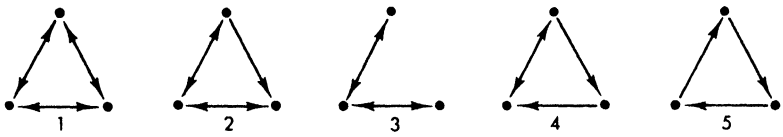


Figure 16.4. Organizational networks used in Heise and Miller study. Arrows indicate direction of information flow. (From G. A. Heise and G. Miller. Problem solving by small groups using various communication nets. *Journal of Abnormal and Social Psychology*, 1951, 46, 328.)

Heise and Miller (1951) did a somewhat similar study in which they examined the networks shown in Figure 16.4. All their networks used only three positions as contrasted with the five-position networks used by Leavitt. Heise and Miller also employed several different types of problem-solving tasks which differed in terms of the amount of communication they required among team members. They found that for a task requiring little communication among organizational positions no differences in effectiveness were observed among the different networks. For a task requiring a moderate amount of communication, network five was found to be worst, network three was intermediate in its effectiveness, and network one was the most efficient. When they used a task requiring a great deal of communication, net-

work five was still poorest, but now network three became most efficient. Note that network three is the only radial net used in the study—thus the findings seem to support what Leavitt has found in his study. Heise and Miller also found that introducing stress into the work situation tended to magnify differences between network effectiveness.

One of the most interesting studies on organizational networks—particularly in reference to Dubin's notion of minimal linkage—was that done by Guetzkow and Dill (1957). They used the same task as Leavitt had used in his study but employed the three networks shown in Figure 16.5. Instead of focusing their interest

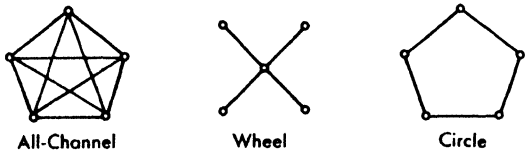


Figure 16.5. Organizational networks used in Guetzkow and Dill study. (From H. Guetzkow and W. R. Dill. Factors in the organizational development of task-oriented groups. *Sociometry*. American Sociological Association, 1957, 20, 175–204.)

exclusively on the performance of these different structures, Guetzkow and Dill also were vitally interested in seeing how the members of each system *restructured* themselves in order to perform their task. By giving a two-minute pause between each problem trial, they permitted the team members to hold conferences during which they could organize their work system within the limits of the communication restrictions already placed on them. The striking finding was that the tendency was overwhelming to set up work organizations having *fewer* links than the original system. These new systems were of two types—a pure vertical radial having four links (called a two-level hierarchy) and a vertical radial connected to two vertical serial chains (called a three-level hierarchy). Both structures possess only *four* links (see Figure 16.6)!

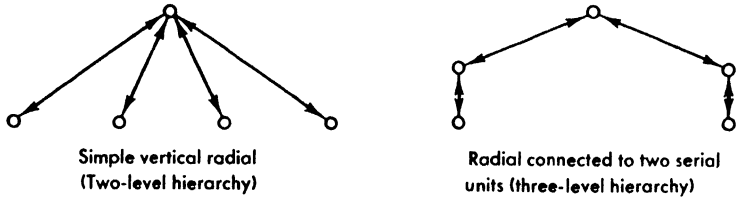


Figure 16.6. Resulting organizational structures found by Guetzkow and Dill. (From H. Guetzkow and W. R. Dill. Factors in the organizational development of task-oriented groups. *Sociometry*, American Sociological Association, 1957, 20, 175–204.)

Seventeen out of the twenty groups originally organized in a circular linkage system (called an all-channel by the authors)—a system with ten links—modified their task into one which required only the four links. Of the 15 groups which worked under the "wheel" condition (a pure radial), *all* retained a four-link system —either the two-level hierarchy or the three-level hierarchy. In no case did they reorganize into a more complex system. Of the 21 groups which worked under the "circle" system (a closed serial), ten reorganized into systems requiring only four links instead of the five they started with. *All told, then, 42 out of 56 groups reorganized into a minimum linkage system.*

At this point the objective evidence would certainly seem to support Dubin's hypothesis that minimum linkage is a desirable quality for an organization, at least if one is interested in promoting organizational efficiency. If group morale is the objective, then perhaps just the opposite may be true.

THEORIES OF ORGANIZATIONS

One can very legitimately ask whether there are such things as theories of organizational behavior. In recent years a considerable body of writings has appeared which seems to deal with so-called "organizational theory"—a fact which would seem to be *ipso facto* evidence that such theories do indeed exist. However, as far as your authors can tell, there seems to be no difference between these "organizational" theories and theories of management and/or leadership. Thus, as we outline the major organizational theories in this section, the reader may notice a tremendous overlap between the following points of view and the major theories of leadership which were discussed in Chapter 14.

Bennis (1966) has grouped current organizational theories into five categories:[1]

A. Exchange Theories
   1. Barnard-Simon
   2. Levinson
B. Group Theories
   1. Mayo
   2. Likert
C. Value Theories
   1. Argyris
   2. Blake and Mouton
D. Structural Theories
   1. Shepard
E. Situational Theories
   1. McGregor
   2. Leavitt
   3. Thompson and Tuden

EXCHANGE THEORIES

According to Bennis, the essential element of exchange theories is that they view the operation of an organization as the establishment of a balance or an equilibrium between the individual workers and the formal organization. Both parties give something to the other and, in turn, receive something in return for what they have given. The Barnard-Simon theory emphasizes the economic aspects of the exchange. Thus, they view the organization as offering wages, services, and income to the individual as inducements to his participation, while the individual offers his services as a worker as his part of the bargain. Organizational equilibrium is attained when the utility of the inducements is equal to the utility of the contributions. That is, when both worker and organization feel they are getting value received for value paid, then equilibrium exists.

[1] For a more complete discussion of these theories see Bennis (1966).

The Levinson theory of exchange is very similar, except that the model is based more upon motivational aspects of work rather than upon the economic aspects of work. Thus, the worker must feel that he is getting as many psychological reward units from the organization as he is giving up in order for balance to exist.

GROUP THEORIES

The group theories are those which take the point of view that association and human affiliation are primary work motivators, and that the degree of cooperation and cohesiveness in one's primary group is very critical to a successful organization. Mayo, of course, was probably among the first to emphasize the importance of viewing any organization as a *social* system, rather than as an economic system.

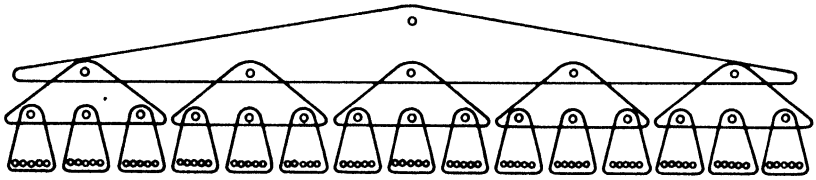


Figure 16.7. The "linking-pin" concept. (From R. Likert. Motivational approach to organizations. In M. Haire, ed. *Modern organizational theory.* Wiley, New York, 1959, 194.)

To Mayo, the essential element is the interaction between the organization and its members was that of *cooperation*. However, in order for such cooperation to take place, management had to be made aware of the importance of the social aspects of work and its influence upon the success of any organization. Likert's theory is very similar in that it also emphasizes the importance of a cohesive primary work group, but Likert feels that cooperation between groups can be accomplished by what he has called "linking pins," rather than by having to train or convince all of management of the importance of social work group forces. The linking pin concept simply means that there should be at least one member of every organization unit who is *also* a member of the organizational unit immediately higher in the organizational hierarchy.

VALUE THEORIES

Several theories are predicated upon the notion that the basic values of the organization are generally quite different from the values held by the individuals within the organization, and that this conflict of value systems is what is critical to determining the effectiveness and the stability of the organizational system. Argyris, for example, takes the point of view that the basic values of organizations are "bureaucratic" values, and that these values are shallow and impersonal. They lead to mistrust, conflict, lowered efficiency and poorer overall effectiveness. Managers should be trained in the development of competence in understanding and appreciating new values—values which include the expression of individual feelings, personal trust, and personal concern. It is basically a very humanistic theory; one which draws heavily from Maslow's need hierarchy (see Chapters 11 and 14).

The Blake and Mouton managerial theory is substantially different. It seems rooted in the notion demonstrated by the Ohio State leadership studies that the two major dimensions of managerial "style" are (1) concern for people and (2) concern for production. According to Blake and Mouton, one can find eight different styles of management, as defined by the particular combination of these two dimensions. They suggest that optimal style involves maximizing *both* dimensions, and suggest further that such a style can only be developed through extensive organizational training.

STRUCTURAL THEORIES

The essential element of the structural theory or point of view proposed by Shepard and others is that organizations should *not* be viewed as being mechanistic in nature. An organization is not a machine and it is, therefore, inappropriate to apply mechanistic principles to it. Instead, organizations should be viewed as "living" organisms which are dynamic, adaptive to change and, in general, highly responsive to the world around them. Table 16.1 shows the major distinctions between the two types of system approaches to organizations. Note that the mechanical approach is a much more rigid and controlled approach to organizational effectiveness than is the organic system approach.

SITUATIONAL THEORIES

Several theories of organizational behavior have taken the specific situation as the focal point for resolving the conflict between individual and organizational goals. For example, McGregor suggests a process he calls *integration* which involves a complicated process of working out these conflicts for any particular situation as a joint endeavor between the manager and his subordinates. Leavitt takes the point of view that the organization must fit the task or the situation—that is, one should take the task as being given and structure the organization to fit it, rather than take the organization as being given and attempt to shape the task to fit the organization. He feels that within one large organization there may be many subunits, each with quite different organizational forms.

TABLE 16.1 *Differences Between Organizations Viewed as Mechanical Systems and Organizations Viewed as Organic Systems*

Mechanical Systems	*Organic Systems*
Individual skills	Relationships between and within groups
Authority-obedience relationships	Mutual confidence and trust
Delegated and divided responsibility rigidly adhered to	Interdependence and shared responsibility
Strict division of labor and hierarchical supervision	Multigroup membership and responsibility
Centralized decision making	Wide sharing of control and responsibility
Conflict resolution through suppression, arbitration, or warfare	Conflict resolution through bargaining or problem-solving

SOURCE: W. G. Bennis. Organizational developments and the fate of bureaucracy. *Industrial Management Review*, Spring 1966, 7(2), 41–55.

Finally, Thompson and Tuden suggest that the basic pivotal element for organizational structures should be the kind of decision process involved (see Table 16.2). They postulate that the two major dimensions of controversy or conflict in any type of decision are (1) the causal factors leading up to the decision and (2) the alternative outcomes or courses of action available. Depending upon whether there is agreement or disagreement in either or both of these dimensions, one can conceive of four somewhat different "decision situations," as shown in Table 16.2.

They believe that a different type of decision strategy, and, therefore, a different type of organizational structure, is appropriate for each of the four separate kinds of decisions. These four strategies and their corresponding structures are shown in the table.

TABLE 16.2 *Four Classes of Decisions Postulated by Thompson and Tuden*

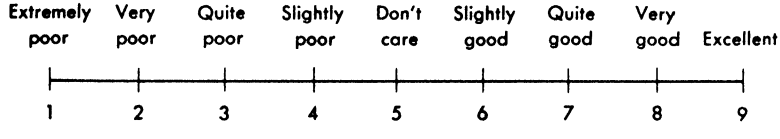
Beliefs About Causation	*Preference About Possible Outcomes*	
	Agreement	*Nonagreement*
Agreement	Computation in *bureaucratic* structure	Bargaining in *representative* structure
Nonagreement	Majority judgment in *collegial* structure	Inspiration in *"anomic"* structure

SOURCE: W. G. Bennis. Organizational developments and the fate of bureaucracy. *Industrial Management Review*, Spring, 1966, 7(2), 41–55.

RECENT RESEARCH ON ORGANIZATIONAL BEHAVIOR

Shartle and his coworkers have developed a framework for studying organizational behavior which is based upon the unit of a "value act." If an act is simply a unit of observable behavior, then a value act is a second unit of observable behavior *which is an evaluation of the first unit or act*. According to Shartle (1958), value acts of people and of groups can be clustered and scaled into distinct value dimensions; further, it should be possible to study organizations in terms of value "profiles."

In their major study, Shartle, Brumback, and Rizzo (1964) collected several thousand statements of behavior which were supposedly good or poor in nature. These statements were edited and the number of items reduced to slightly over 200 for the final questionnaire form. A separate form was made up for each of three classes of organizations—military, business, and public schools. Each form was then given to several hundred subjects for rating. Subjects were asked to rate how "good" or how "poor" they felt each behavior was for that type of organization using the nine-point scale shown below.



A sample of behaviors taken from the form for business firms follows:

1. Each employee is put on his own.
2. The board of directors holds the president solely responsible for the firm.
3. The firm donates money only when such gifts appear to benefit the firm directly or indirectly.
4. Executives in the firm are rotated from one job to another.
5. The firm prefers to sell more goods to its present customers rather than simply to increase the number of customers.
6. The firm plans to set up many branch offices so that more customers can be conveniently reached.
7. The firm prefers that its employees live in company housing.
8. Clever strategy is necessary for one to get promotions.
9. The firm emphasizes hiring office personnel who are dignified and responsible.
10. Length of service in the firm is the principal qualification for promotion.

Of course each response on the part of the rater is a value act in that it is an evaluation of another unit of behavior. Subjects' responses were intercorrelated and statements on each questionnaire were grouped using factor analysis to obtain underlying "value dimensions" for that type of organization.

The resulting dimensions for each type are shown below, along with several items which help define each factor.

1. *The Business Firm*

B-1. *Degree of organizational magnitude, expansion, and structure*
The firm solicits contracts from all sources, government as well as others, the firm has salesmen in all states soliciting business.

B-2. *Degree of internal consideration for welfare, health, and comfort*
The firm insists that each employee carry hospitalization insurance; executives look out for the welfare of the individual employees.

B-3. *Degree of competition, strategy, and shrewdness*
Because of special favors they have received, the police force gives the firm extra attention; the firm will absorb a competitor if it can.

B-4. *Degree of ethical and social responsibility*
The firm attempts to aid in solving international problems; the firm has a chaplain who leads devotional services for employees.

B-5. *Degree of quality of products and services*
The firm is extremely particular in checking every detail of the finished product; the firm keeps the quality of its services high even when it loses business.

B-6. *Degree of change*
A procedure or policy is never changed in the firm until a better one has been proved; the firm is always very cautious in making changes.

B-7. *Degree of organizational control over member identifications*
Employees act as if their lives belong to the firm; the firm encourages employees and their families to join community activities that will help the firm.

B-8. *Degree of external political participation*
The firm asks its employees to keep quiet about political views; the firm allows its name to be listed as one of the supporters of a political party.

B-9. *Degree of member equality and recognition*
The firm's officials are called by their first names; executives treat all the employees as their equals.

2. *The Military Service*

M-1. *Degree of internal efficiency and harmony*
The enlisted men dislike the commander; occasionally, military personnel have to cover up for the actions of their superiors.

M-2. *Degree of internal consideration for happiness, friendliness, and comfort*
Officers find time to listen to subordinates; the commander knows everyone in the unit by name.

M-3. *Degree of military authority and individual status*
Officers in the service have more comfortable working conditions than do enlisted men; military personnel are asked on short notice to work overtime.

M-4. *Degree of moral and religious emphasis*
The service has chaplains who lead devotional services for the military personnel, the service encourages military personnel to go to a church of their own choice.

M-5. *Degree of internal authority and organizational control*
Military personnel must wear uniforms at all times; the service will not let any unmarried military personnel live off base.

M-6. *Degree of military task orientation*
The service asks its civilian employees to keep quiet about their political views; any mission, no matter how trivial, is completed.

M-7. *Degree of personal sociability and equality*
Officers do personal favors for subordinates; officers treat all subordinates as their equals.

M-8. *Degree of individual welfare and development*
Officers look out for the welfare of individual subordinates; the service offers opportunity to young men to develop individual leadership.

3. *The Public High School*

E-1. *Degree of progressiveness, newness, and scope of service*
The school provides a course on family and marriage; the school has the very latest textbooks and equipment; the school offers educational services to various firms in the community.

E-2. *Degree of internal consideration and staff cooperation*
The principal refuses to explain his actions; the school board, the principal, and the teaching staff do not know what each other is doing

E-3. *Degree of athletic participation*
The school holds a pep assembly before each home athletic game; the school stresses the athletic program.

E-4. *Degree of emphasis on academic achievement and competition*
The teachers encourage competition among students; the school stresses scholastic achievement above all else.

E-5. *Degree of quality and financial support*
The school seeks financial support from all sources, government as well as others; taxes in the community are kept high because of the school.

E-6. *Degree of internal authority control*
The Lord's Prayer is recited at the start of every large student assembly [This item would no longer be valid.]; the principal encourages the use of uniform procedures.

E-7. *Degree of member equality*
The school employs mainly male teachers; the school obtains many more scholarships for boys than for girls.

E-8. *Degree of student services and teacher quality*
The school operates a cafeteria; the school expects the teachers to set an example for the students.

E-9. *Degree of administrative emphasis on teacher standards*
The principal requires the teachers to have definite standards of performance; the teachers are encouraged to get graduate degrees.

E-10. *Degree of academic emphasis and absence of structure*
The school does not teach bright students all they can learn about a subject; the school does not fail students about to graduate.

Examination of the above factors shows certain similarities and certain differences between the three types of organizations in terms of their value profiles. For example, a dimension of organizational consideration for member welfare was present in all three types, as was a dimension of control over members. Also, all three types of organizations exhibited an individual-member equality dimension and a competitiveness dimension. On the other hand, the dimensions of external political participation and of organizational change were unique to the business firm. Unique military factors were task orientation and individual status. Unique public high school factors were emphasis on teacher standards and athletic participation.

In summary, the work of Shartle appears, at this point, to hold considerable promise as an approach to gaining further insight into fundamental patterns of organizational behavior. Hopefully, further research along these lines will be forthcoming.

Dunteman (1966) has reported a study of considerable magnitude in which he attempted to explore some of the relationships among 84 different variables pertaining to different organizational characteristics, management attributes, incentive conditions, worker characteristics, personnel performance, and organizational functions. The data were gathered by an 84-item multiple-choice questionnaire sent to a representative sample of 234 manufacturing firms located throughout the United States. The correlations among the 84 variables were factor analyzed, and 14 dimensions of organizational attributes and behavior were isolated and interpreted. The fourteen factors were:

1. Size of organization
2. Economic growth
3. Tardiness versus family responsibility
4. Pay-skill level
5. Personnel tenure
6. Ownership and concern for organizational interests
7. Work-force reduction and job mechanization
8. Technical personnel and controls versus protection against human liabilities
9. Minority-group composition
10. Improvement of working conditions
11. Retail sales personnel and authority—conflict behaviors
12. Community and employee support versus work-output restriction
13. Employee selectivity
14. Allocations to labor versus product development

In a similar study done on 188 manufacturing firms in a southern metropolitan area, Palmer (1961) clustered 35 different organizational variables and identified 8 different organizational dimensions.

1. Retirement welfare
2. Cooperation with survey
3. Size of work force
4. Thrift benefits
5. Cost of sickness versus use of machinery
6. Job aversion
7. Insurance benefits
8. Product theft versus discounts on products

It is important to note the basic difference in philosophy between studies of the Dunteman and Palmer type as opposed to that of Shartle. Shartle's point of view has the implicit assumption that organizations are best studied through the kinds of behavior or policies that the organization exhibits (acts) and how "good" or "bad" they are. It is an evaluative, judgmental approach which, in a sense, says that you must take society (that is, people) into account in studying organizations. The Dunteman-Palmer approach simply groups organizations on the basis of a number of more or less physical attributes. While both methods no doubt contribute to our knowledge of organizations, the approach of Shartle would somehow seem to be the more psychologically richer of the two methods.

Indeed, Vroom (1966) has recently argued against studies of the Dunteman-

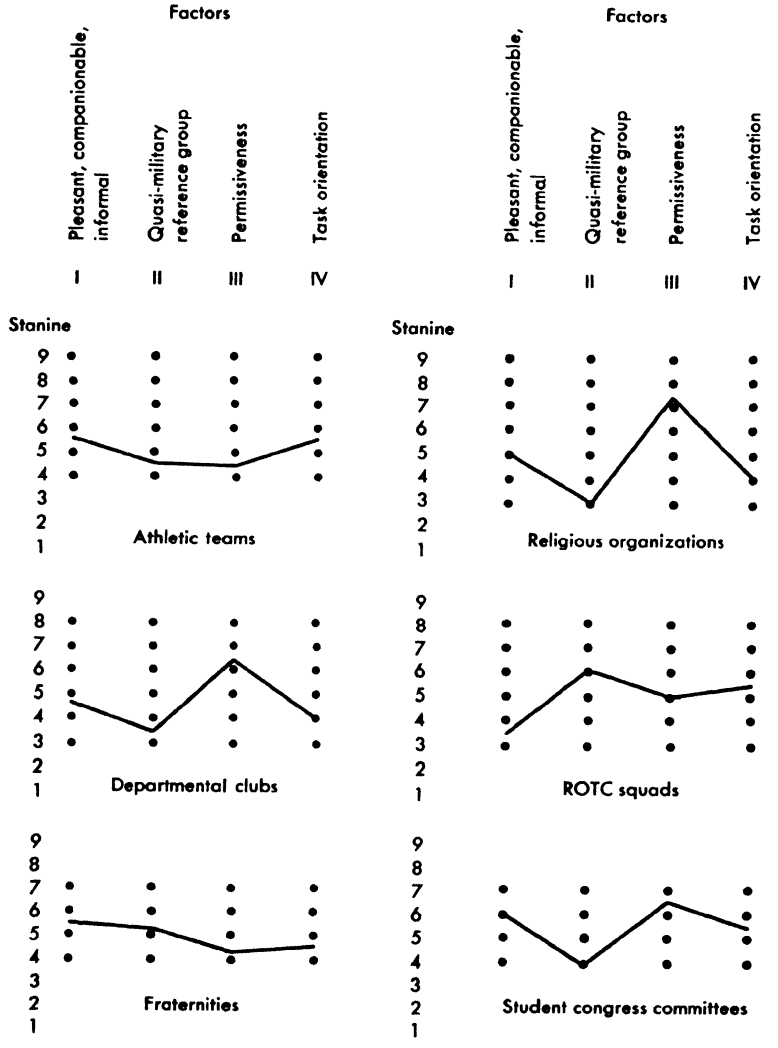


Figure 16.8. Profiles of six sets of groups on four orthogonal factors. (From F. Findikan and S. B. Sells. Study of organization structure and similarity of campus student organization. *Organizational Behavior and Human Performance*, 1966, 1, 169-190.)

Palmer type. He calls these "static correlational studies." He suggests an alternative procedure which he calls "dynamic correlational studies" in which one looks at the correlations between changes in organizational variables over time, rather than in simply looking at absolute values at one specific moment in time.

Findikan and Sells (1966) adopted a somewhat different method in looking at dimensions of organizations. They administered the Group Dimensions Description Questionnaire of Hemphill and Westie (1950)—an instrument designed to measure and quantify organizational dimensions—to 967 undergraduates at Texas Christian University who were members of 60 different campus student organizations. The 13 scales on the questionnaire were factor analyzed and four factors were obtained. The groups to which the students belonged were then classed into six types, and factor profiles for each type were computed. These profiles are shown in Figure 16.8.

Student congress committees, fraternities, and athletic teams all scored high on the first factor (pleasant-companionable) while ROTC scored lowest. Factor two was a dimension that represented high structure and high control over members. ROTC groups and fraternities scored high, while religious organizations and departmental clubs scored low. The third factor, permissiveness, found religious groups scoring high and fraternities, athletic teams, and ROTC groups scoring low. The final factor, task orientation, found ROTC groups, student congress committees, and athletic teams scoring high, while departmental clubs and religious organizations showed a low degree of task orientation.

Until quite recently there has been virtually no research dealing with the topic of international organizations—no doubt due to the tremendous practical problems inherent in such research. However, Triandis (1967) reports an initial study on the interpersonal aspects of organizational relationships in Greek-American organizations in which he found that interpersonal stereotypes were markedly influenced by the amount of contact between organizational members.

COMMUNICATION IN ORGANIZATIONS

It is probably safe to state that communication is one of the most basic processes in organizations. Indeed, any type of cooperative enterprise would seem to require communication in some form, even if it is some type of nonverbal communication system. The task of studying the characteristics of real-life communication patterns, systems, procedures, etc., in any large-scale organization can be an exceedingly difficult and complex task. There are a number of techniques which have been used for studying organizational communication—each with its own peculiar advantages and its own share of disadvantages.

"LIVING-IN" PROCEDURE

As the name implies, the experimenter becomes an observer who actually joins and attempts to become part of the organization under study. He tries to establish himself as an accepted member of the group so that his presence will not be perceived as an "alien" outside influence which could itself have a biasing influence on the behavior of the system under study. While the method has merit, it is obviously very time-consuming. Days, weeks, and even months may be required before the

group members become adapted and at ease in the presence of the observer, and, therefore, begin to behave as if he were not there.

INDIRECT ANALYSIS

This method studies communication by using other, more easily observed measures as a base, and then inferring from them how much communication takes place between various organizational units. For example, one approach is simply to obtain measures (using logs or self-reports) of how much time various units spend with each other. If one then makes the assumption that those units which spend the most time together also communicate the most together, one can develop a picture of the communication structure of the organization. Unfortunately, these indirect measures provide little insight into the *content* of the information itself.

DUTY STUDY

In the duty study method, the experimenter or observer selects a particular spot or location in the network and stations himself there to observe all communications which flow past this particular point. In this method a great deal of information can be obtained about the content, quality, and quantity of information at any selected point, but one gets no information about the overall *pattern* of information flow in the organization.

CROSS-SECTION ANALYSIS

Cross-section analysis is in a sense the opposite technique from duty study. This method, at a given moment in time, draws a sample of communications from points throughout the entire organization. Thus, at 10:15 A.M. the experimental team might walk into the situation and collect and/or record all communication taking place at that moment. By doing this a number of times at unexpected intervals, a random sample of communications may be obtained which would provide an overall picture of the communication pattern *and* also provide some indication of the type and content of information flow at each point in the system.

ECCO ANALYSIS

The term *ECCO* is an abbreviation for "episodic communication channels in organization." This method was first proposed by Davis (1953). It is simple in that it focuses on a particular piece of information and watches it progress throughout the entire organization until it reaches its terminal point. The information can be experimentally inserted at any point and subsequently followed, or one can observe regular information units that are part of the daily routine. The advantage of this method is that not only does it provide information about the communication pattern, but it also gives information about time delays, points of blockage, etc.

DECENTRALIZATION OF ORGANIZATIONS

As business firms or corporations grow in size and diversification of product output, they must inevitably come to grips with the problem of whether or not to reorganize

their structure in keeping with this expansion. The most typical reorganization question that is faced in such growth situations is whether or not the firm should "decentralize."

The centralization-decentralization problem is an extremely complicated one. In the first place, there is no clear-cut definition to the decentralization process. In general, it refers to a procedure of "cutting-up" an organization into more autonomous units which are then often located some distance apart geographically. There are a wide variety of ways in which the organizational chart of a large firm can be "sliced" into separate units. Thus, the subdivision may take place by separating the total organization by horizontal slicing (for example, taking the top managerial units and moving them out of the main plant to a downtown corporate headquarters) or vertical slicing (for example, separating the marketing, production, and service divisions and locating them in different places). In a major decentralization, aspects of both vertical and horizontal separation are usually present.

The arguments in favor of decentralization are usually based upon an efficiency concept. Often marketing and distribution demands, labor pool requirements, etc., are such that a firm may feel a strong need to decentralize in order to grow and to remain competitive. Costs can often be reduced by decentralization.

Decentralization also raises a number of problems. Clearly, communication between subunits may become more difficult in a subdivided firm. The advantages of autonomy also have corresponding disadvantages, such as the feeling of isolation. The employees in a certain plant may feel that "the brass in the main office just doesn't understand the problems that we have out here!" Also, in certain cases decentralization may lead to both duplication of effort and to unnecessary competition between subunits. Because of the difficulty in communication between subunits, it is easily possible that these kinds of things may occur.

Morale may often be affected, although little research has been done to determine whether this is the case. With a decentralized firm it may be very difficult for an individual employee to get any real feeling for the goals and objectives of the overall organization, thus making it hard for him to assess whether his goals are indeed compatible with those of the firm.

There should be more attention brought to bear on the decentralization problem in the future. It is an important area—one well worthy of the attention of industrial psychologists.

REFERENCES

Bakke, W. E. Concept of the social organization. In M. Haire (ed.). *Modern organizational theory*. Wiley, New York, 1959.

Bavelas, A. Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 1950, 22, 725–730.

Bennis, W. G. Organizational developments and the fate of bureaucracy. *Industrial Management Review*, 1966, 7, 41–56.

Davis, K. A. A method of studying communication patterns in organizations. *Personnel Psychology*, 1953, 6, 301–312.

Dubin, R. Stability of human organizations. In M. Haire (ed.). *Modern organizational theory*. Wiley, New York, 1959.

Dunteman, G. H. Organizational conditions and behavior in 234 industrial manufacturing organizations. *Journal of Applied Psychology*, 1966, 50, 300–305.

Findikan, F., and S. B. Sells. Study of organizational structure and similarity of campus student organization. *Organizational Behavior and Human Performance*, 1966, 1, 169–190.

Guetzkow, H., and W. R. Dill. Factors in the organizational development of task-oriented groups. *Sociometry*, 1957, 20, 175–204.

Heise, G. A., and G. Miller. Problem solving by small groups using various communication nets. *Journal of Abnormal and Social Psychology*, 1951, 46, 327–336.

Hemphill, J. K., and C. M. Westie. The measurement of group dimensions. *Journal of Psychology*, 1950, 29, 325–342.

Leavitt, H. J. Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 1951, 46, 38–50.

Leavitt, H. J. Toward organizational psychology. Walter V. Bingham Memorial lecture, Carnegie Institute of Technology, March 23, 1961.

Likert, R. Motivational approach to organizations. In M. Haire (ed.). *Modern organizational theory*. Wiley, New York, 1959.

Palmer, G. J., Jr. Incentive conditions and behavior in 188 industrial manufacturing organizations. Technical Report No. 3, 1961, Project NR 170-478, Tulane University, Contract Nonr 475(08), Office of Naval Research.

Sayles, L. R., and G. Strauss. *Human behavior in organizations*. Prentice-Hall, Englewood Cliffs, N.J., 1966.

Shartle, C. L. A theoretical framework for the study of behavior in organizations. In A. W. Halpin (ed.). *Administrative theory in education*. University of Chicago, Midwest Administration Center, Chicago, 1958.

Shartle, C. L., G. B. Brumback, and J. R. Rizzo. An approach to dimensions of value. *Journal of Psychology*, 1964, 57, 101–111.

Tannenbaum, A. A., and S. E. Seashore. Some changing conceptions and approaches to the study of persons in organizations. Paper presented at International Association of Applied Psychology, Ljulhyana, Yugoslavia, 1965.

Triandis, H. C. Interpersonal relations in international organizations. *Organizational Behavior and Human Performance*, 1967, 1, 26–55.

Vroom, V. H. A comparison of static and dynamic correlational methods in the study of organizations. *Organizational Behavior and Human Performance*, 1966, 1, 55–70.

JOB ANALYSIS AND EVALUATION 17

A job analysis is an accurate study of the various components of a job. It is concerned not only with an analysis of the duties and conditions of work but also with the individual qualifications of the worker. Whereas the industrial engineer is concerned primarily with the job aspects, the industrial psychologist tends to emphasize the man aspects of the job.

Studying the world of work makes it obvious that jobs are performed under vastly different temperatures, postures, and hazards, as well as innumerable other variations in work effort.

Although it is true that different meanings are assigned to the terms job analysis, job specification, and job evaluation, for the practical purposes of industrial psychologists these terms may be regarded as relatively similar.

USES AND VALUE OF JOB ANALYSIS

Lawshe and Satter (1944) propose four major uses of job analysis: the derivation of training content, the setting up of personnel specifications, the improvement of job efficiency, and the establishment of wage structures.

After an intensive survey of 401 articles in the literature on job analysis, Zerga (1943) concludes that there are approximately 20 uses for this type of information. He mentions the following specific uses:

1. Job grading and classification
2. Wage setting and standardization
3. Provision of hiring specifications
4. Clarification of job duties and responsibilities
5. Transfers and promotions
6. Adjustment of grievances
7. Establishment of a common understanding between various levels of workers and management
8. Defining and outlining promotional steps

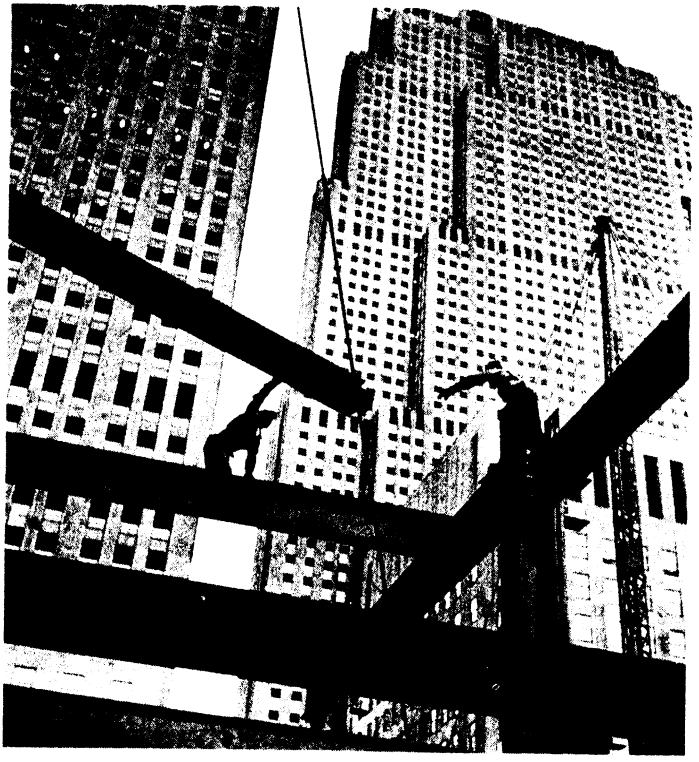


Figure 17.1. Hazards in work: Be sure the pieces fit. (Courtesy Standard Oil Co. [N.J.].)

9. Investigating accidents
10. Indicating faulty work procedures or duplication of effort
11. Maintaining, operating, and adjusting machinery
12. Time and motion studies
13. Defining limits of authority
14. Indicating cases of individual merit
15. Indicating causes of personal failure
16. Education and training
17. Facilitating job placement
18. Studies of health and fatigue
19. Scientific guidance
20. Determining jobs suitable for occupational therapy

A careful examination of this list shows the great importance of having a job analysis made. Furthermore, this technique should not be regarded as being suitable only for large organizations. A good job analysis can be very important in easing much of the friction that often arises in the typical small office employing only a few girls. It is not unusual to find that in such an office the girls continually quarrel as to whose job it is to get the bills ready for mailing; everyone insists that she was not hired for that job and she sees no reason why it should fall on her. A frequent solution is to have the girls take turns, but this hardly ever helps because the girl who has to do this work on a particular day will accuse the other girls of refusing to help

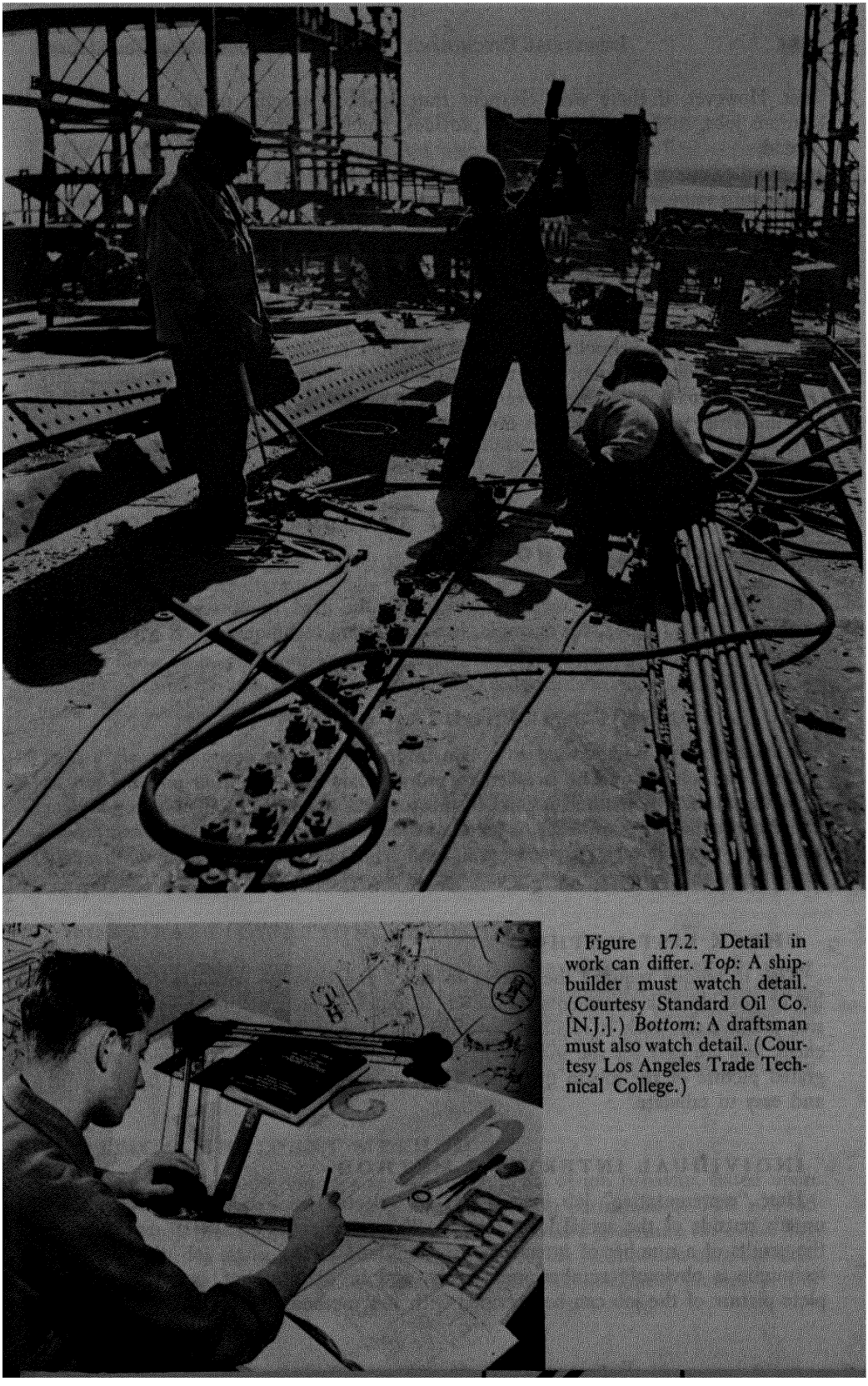


Figure 17.2. Detail in work can differ. *Top:* A shipbuilder must watch detail. (Courtesy Standard Oil Co. [N.J.].) *Bottom:* A draftsman must also watch detail. (Courtesy Los Angeles Trade Technical College.)

her. However, if there were clear-cut statements of specific requirements for the various jobs, such as a job analysis provides, much of this friction would be eliminated.

Another source of friction in offices is the lack of definite limits of authority. Thus a girl with seniority, even if it is only of two weeks, will attempt to assume authority over a new employee; the latter soon discovers that she has far too many bosses. Here again a job analysis is valuable.

In many respects a job analysis is a vital part of working efficiency, besides promoting smooth working relationships among the employees. It can be regarded as the foundation upon which a system of efficiency is built. Effective training programs, good personnel policies, attempts to reduce fatigue, and many other measures leading to efficiency cannot be introduced effectively unless the basic and preliminary step of job analysis has been taken. Prior to the development of psychological tests for industrial use, a job analysis must be made.

JOB ANALYSIS METHODS

There are many different methods which may be used to obtain data for a job analysis. Morsh (1964), for example, has listed nine techniques, each with its own special advantages.

QUESTIONNAIRE METHOD

This method is usually used to obtain information about occupations via a mail survey. The job incumbent is asked to provide data about himself and his job in his own words. The method is good for people who write easily but not so good for collecting data from low-level workers who have little facility for self-expression. Also, it is often a very time-consuming and laborious process to analyze the data obtained in this manner.

CHECK LIST METHOD

This technique requires the worker to check the tasks he performs from a long list of possible task statements. However, in order to prepare the check list, extensive preliminary work is required in collecting appropriate task statements. While check lists are easy for the incumbent to respond to, they do not provide an integrated picture of the job in question. They are easily administered to large groups and easy to tabulate.

INDIVIDUAL INTERVIEW METHOD

Here "representative" job incumbents are selected for extensive interviewing—usually outside of the actual job situation. The interview is usually structured, and the results of a number of interviews are combined into a single job analysis. The technique is obviously cumbersome, costly, and time-consuming, but a very complete picture of the job can be obtained with this method.

OBSERVATION INTERVIEW METHOD

The observation interview actually takes place right on the job. The interviewer collects data from the incumbent, using normal interview methods, as the incumbent performs his work. The interviewer observes and questions the worker in an attempt to get complete job description data. Like the individual interview, it is a slow and costly method which may also interfere with normal work operations. However, it generally produces a good and complete job description.

GROUP INTERVIEW METHOD

The group interview is similar to the individual interview except that a number of job incumbents are interviewed simultaneously. Under the guidance of the interviewer, the interviewees recall and discuss their work activities. The interviewer then combines their comments into a single job description. The advantage over the individual method is the savings in time obtained by the group method.

TECHNICAL CONFERENCE METHOD

This method uses "experts" rather than actual job incumbents as a source of information. These experts are usually supervisors who have extensive knowledge of the job in question. They meet with the job analyst and attempt to specify all the characteristics of the job. The problem with this method is that the experts may not actually know as much about the job as the analyst would hope, since they do not actually perform the task themselves. Thus, their judgments are only estimates based upon their background experience.

DIARY METHOD

Here job incumbents are required to record their daily activities each day using some type of logbook or diary. The method is good in that it systematically gathers a great deal of information, but it can also take a great deal of time on the part of the worker if the recording forms are not kept simple.

WORK PARTICIPATION METHOD

With this procedure the job analyst actually performs the job himself. By doing the work himself he is thus able to obtain firsthand information about what characteristics comprise the job under investigation. The technique is fairly effective for simple jobs, but complex jobs usually require that the job analyst be extensively trained prior to his session of work activity. The method is clearly time-consuming and costly.

CRITICAL INCIDENT METHOD

This involves the collection of a series of statements of job behavior, based upon direct observation or memory, about good and poor job performance. This method was discussed in the chapter on performance evaluation as a procedure for evaluating job incumbents. In job analysis such incidents can provide information about *critical* aspects of the job, but the method does not provide an integrated picture of the entire task.

Form USES-546
(2-44)

U. S. DEPARTMENT OF LABOR
BUREAU OF EMPLOYMENT SECURITY
UNITED STATES EMPLOYMENT SERVICE

Budget Bureau No. 44

JOB ANALYSIS SCHEDULE

1. Job title .....
2. Number .....
3. Number employed M ..... F .....
4. Establishment No. .....
6. Alternate titles .....
5. Date .....
.....
Number of sheets ...
.....
8. Industry .....
.....
9. Branch .....
7. Dictionary title and code .....
10. Department .....

11. WORK PERFORMED:

(CONTINUE ON SUPPLEMENTARY SHEETS)

Analyst .....
Reviewer .....

Figure 17.3. Job analysis form used by the United States Employment Service.

SOURCES OF WORKERS

12. Experience: None ..... Acceptable .....

13. Training data: Minimum training time—(*a*) Inexperienced workers.
(*b*) Experienced workers.

Training	Specific Job Skills Acquired Through Training
In-plant (on job) training	
Vocational training	
Technical training	
SRW Eng. General education	
Activities and hobbies	

14. Apprenticeship: Formal ..... Informal ..... Length required .....

15. Relation to other jobs:

(*a*) Promotions from and to, transfers, etc.: .....

(*b*) Supervision received: General ..... Close ..... By ..... (Title)

(*c*) Supervision given: None ..... Number supervised ..... Titles .....

The following items must be covered on supplementary sheets.

PERFORMANCE REQUIREMENTS

16. Responsibility (consider material or product, safety of others, equipment or process, cooperation with others, instruction of others, public contacts, and the like).

17. Job knowledge (consider pre-employment and on-the-job knowledge of equipment, materials, working procedures, techniques, and processes).

18. Mental application (consider initiative, adaptability, independent judgment, and mental alertness).

19. Dexterity and accuracy (consider speed and degree of precision, dexterity, accuracy, coordination, expertness, care, and deftness of manipulation, operation, or processing of materials, tools, instruments, or gages used).

COMMENTS

20. Equipment, materials, and supplies.

21. Definition of terms.

22. General comments.

GPO 883461

Figure 17.3. Job analysis form (continued).

SOME EXAMPLES

The best results in making a job analysis are usually obtained with a combination of various parts of all the methods just mentioned. The form used by the United States Employment Service, which incidentally has done a great deal of valuable work in this field, is presented in Figure 17.3.

A somewhat different form is suggested by Viteles' Guide to Job Analysis (1932) which lists the following eleven items:

1. Identification of the job
2. Number employed
3. Statement of duties
4. Machines used
5. Analysis of operations
6. Conditions of work
7. Pay and nonfinancial incentives
8. Relation to other allied jobs (place in the organization)
9. Opportunities for transfer and promotion
10. Time and nature of training
11. Personal requirements:
    a. General (e.g., age, marital status)
    b. Physical
    c. Educational
    d. Previous experience
    e. General and special abilities
    f. Temperamental and character requirements

Another form (Shartle, 1952) includes the following major divisions:

1. Job name or title
2. Classification title
3. Industry, plant, department, division, section
4. Number employed
5. Hires, quits, job absences
6. Work performed
7. Supervision given and received
8. Level of difficulty
    a. Responsibility
    b. Knowledge
    c. Initiative
    d. Mental alertness
    e. Judgment
    f. Dexterity
    g. Accuracy
    h. Experience and education required
    i. Other factors
9. Pay rate
10. Hours and shifts
11. Output standards
12. Job combination or breakdown
13. Job relationship
14. Machines, tools, equipment, material used
15. Working conditions
16. Social environment
17. Physical demands
18. Worker characteristics

19. Previous experience required
20. Training required
    a. General level
    b. Vocational training
    c. Technical training
    d. On-the-job training
21. Selection methods

The United States Air Force has been exceedingly active in the area of job analyses over the past ten-year period. The method developed by the Air Force combines the features of the check list with those of the questionnaire and the observation interview (Morsh, 1964). Their procedure for developing a job analysis inventory involves the following steps:

1. Development of a preliminary task inventory of from 200 to 300 task statements using job descriptions, training materials, experts and all other possible sources.
2. Administration of preliminary inventory to experienced job incumbents for purposes of review and modification of task statements.
3. Administration of revised inventory to large samples of job incumbents. These men are asked to respond to all statements and to add any which they feel have been omitted.

Inventories constructed in this manner have been found to possess average reliabilities of about 0.70.

A comparison of these four outlines for a job analysis shows considerable overlapping. The point here is that for any organization the specific items to be included in the study may vary. The decision as to which type to use, or the development of a new form that will more readily conform to the needs of a particular company, will depend primarily on which of the many potential uses of a job analysis is considered the most important.

SOME RESEARCH ON JOB ANALYSIS

CHECK LISTS

Dunnette and Kirchner (1959) have developed a check list of job activities for differentiating different kinds of sales jobs; Dunnette and England (1957) have developed a comparable check list for use among engineering jobs. Similarly, the United States Department of Labor has published a list of "estimated worker trait requirements" for 4000 jobs listed in the *Dictionary of Occupational Titles*.

FACTOR ANALYSIS

One of the areas in which it is most difficult to carry out job analyses is with high-level managerial positions. Hemphill (1960) has approached this problem by factor analysis. He studied the basic characteristics of the positions of 93 different business executives to try to determine the essential dimensions of executive positions.

Prien (1963), in a study somewhat similar to Hemphill's, identified seven dimensions of supervisory jobs. He labeled these:

1. Manufacturing process supervision
2. Manufacturing process administration
3. Employee supervision
4. Manpower coordination and administration
5. Employee contact and communication
6. Work organization, planning, and preparation
7. Union management relations

Palmer and McCormick (1961) also used the factor analysis method to get job dimensions for a stratified sample of 250 people in a steel-producing firm. They identified four major job dimensions:

1. General decision making and mental activity
2. Sedentary versus physical work activity
3. Communication and information processing
4. Knowledge of tools and mathematics

DIARY METHOD

Another method for analyzing jobs is through a procedure known as the self-recording or *diary* method. This requires each job incumbent to keep a daily log of his activities and the amount of time spent on each during the day. In this fashion one is able to get a very comprehensive picture of the job in question. The major difficulty with the method is that it requires a great deal of time on the part of the job incumbent. He may start complaining that he spends more time writing in his diary than he does doing anything else! Thus, very simple, time-saving report forms are a necessity for this method. It is particularly useful for high-level jobs, such as management or scientific positions. Some recent examples of job analyses carried out with this procedure are the studies of Dubin and Spray (1964), Hinrichs (1964), and Horne and Lupton (1965).

The data collection form used by Hinrichs is shown in Figure 17.4. His study had three major objectives:

1. To develop and evaluate a work sampling methodology utilizing a self-recording technique
2. To provide data to the company on how its professional employees spend their time as a base for programs to attain better manpower utilization
3. To compare questionnaire estimates of time allocation with data obtained through work sampling

He was particularly interested in the amount and types of communication carried out by professional and scientific personnel (his 232 subjects were primarily chemists and chemical engineers).

Each participant filled out the form five times for eleven consecutive work days (times were varied randomly each day to get a random sample). The major results are shown in Table 17.1.

The striking finding was that even though these were primarily scientific personnel, the biggest single slice of their time (61 percent) was spent in communi-

NAME

DATE

TIME

1. WHERE ARE YOU? 1 ☐ Your Desk/Office; 2 ☐ Your Laboratory; 3 ☐ Elsewhere in Your Division; 4 ☐ Other Division; 5 ☐ Other Affiliate; 6 ☐ Other

WHAT TIME IS IT NOW?

WHAT ARE YOU DOING?

2. NOT COMMUNICATING – 1 ☐ Analysis or Calculations Not in Draft or Final Form to be Transmitted to Others; 2 ☐ In Transit; 3 ☐ Filing; 4 ☐ Looking for Info.; 5 ☐ Waiting; 6 ☐ Lab Work; 7 ☐ Other

IF YOU CHECK ANY OF THE ABOVE DO NOT GO ON – WIND YOUR WATCH AND SET IT FOR THE NEXT SCHEDULED TIME

3. COMMUNICATING – IF YOU CHECK ANY OF THE FOLLOWING CONTINUE IN THAT COLUMN TO REVERSE SIDE

1 ☐ Listening, Speaking	2 ☐ Writing 3 ☐ Dictating 4 ☐ Editing Written Work of Others 5 ☐ Editing Written Work of Your Own 6 ☐ Plotting or Listing Data in Draft or Final Form to Be Transmitted to Others	7 ☐ Reading Something Written By Others
4. 1 ☐ On Telephone 2 ☐ Meeting – Sch'd. Periodic Intervals 3 ☐ Special Meeting – Not Periodic but Planned at Least One Day Ahead 4 ☐ Contact, Gathering or Bull Session Not Planned One Day Ahead	1 ☐ Letter 2 ☐ Book, Article, or Manual 3 ☐ Note or Informal Memo 4 ☐ Formulas, Data 5 ☐ Formal Memo 6 ☐ Progress Report	1 ☐ Letter 2 ☐ Book, Article, or Manual 3 ☐ Note or Informal Memo 4 ☐ Formulas, Data 5 ☐ Formal Memo 6 ☐ Progress Report
5. How Many People Are Included in This Communication, Including Yourself? 1 ☐ 2 3 ☐ 7-10 2 ☐ 3-6 4 ☐ Over 10	**Is This Material Prepared at Periodic Intervals; For Example, Weekly, Monthly, Quarterly?** 1 ☐ Yes 2 ☐ No	**Is This Material Prepared at Periodic Intervals; For Example, Weekly, Monthly, Quarterly?** 1 ☐ Yes 2 ☐ No
6. Who Started This Communication? 1 ☐ Yourself 2 ☐ Someone Above Your Level 3 ☐ Someone at Your Level 4 ☐ Someone Below Your Level 5 ☐ Don't Know or Not Applicable	**Is This a Redraft?** 1 ☐ First Draft 2 ☐ Second Draft 3 ☐ Third Draft 4 ☐ Fourth Draft or More 5 ☐ Don't Know	
7. Are the Other People Primarily From: 1 ☐ Outside the Company 2 ☐ Other Affiliate 3 ☐ Research Company 4 ☐ Don't Know	**Is This Written Matter Directed Primarily to People From:** 1 ☐ Outside the Company 2 ☐ Other Affiliate 3 ☐ Research Company 4 ☐ Don't Know	**Are You Reading Material Prepared by People From:** 1 ☐ Outside the Company 2 ☐ Other Affiliate 3 ☐ Research Company 4 ☐ Don't Know

8. What Is the Level of Most of the Participants (Other Than Yourself)? | **To What Level Is the Written Matter Primarily Directed?** | **What Is the Level of the Writer or Writers?**

1 ☐ Above Director Level
2 ☐ Director Level
3 ☐ Assoc. or Asst. Director or Sr. Res. Assoc.
4 ☐ Section Head or Assoc.
5 ☐ Group Head or Senior Eng./Chemist
6 ☐ Professional – Non-Supervisory
7 ☐ Non-Professional
8 ☐ Mixed, Not Primarily Any One of the Above
9 ☐ Don't Know or Not Applicable

COMPLETE ALL OF THE REMAINING QUESTIONS

What Is the Main Subject of This Communication? (Answer either 9 or 10)

9. TECHNICAL
1 ☐ Technical Service to Affiliates or Other Divisions
2 ☐ Technical Service to Non-Affiliates
3 ☐ Exploratory Research or Engineering
4 ☐ Development or Normal Research or Engineering
5 ☐ Other Technical

10. NON-TECHNICAL
1 ☐ Employee Matters: Interviewing, Employment, Training and Employee Matters *Other* Than Supervision of Technical Work.
2 ☐ Budget and Finance: Budget Review, Cost Data on Non-Technical Work
3 ☐ Materials: Supplies, Purchases, Sales, Inventory, Inspection, Ordering
4 ☐ Other Non-Technical Company Business
5 ☐ Personal: All Personal Time

11. Does the Subject of This Communication Require Action? 1 ☐ By You 2 ☐ By Another Participant 3 ☐ By Someone Not Participating 4 ☐ By Several Persons 5 ☐ No Action to Be Taken 6 ☐ Don't Know	**12. Does This Communication Deal With An Idea Which:** 1 ☐ You Alone Originated 2 ☐ You Originated With Others 3 ☐ Someone Else Originated 4 ☐ Don't Know	**13. How Long Has This Communication Been Taking Place?** 1 ☐ Less Than 5 Minutes 2 ☐ 6-15 Minutes 3 ☐ 16-30 Minutes 4 ☐ 31-60 Minutes 5 ☐ Over 60 Minutes
14. What Is Your Major Function In This Communication? 1 ☐ Giving Information 2 ☐ Receiving Information 3 ☐ Exchanging Information	**15. Do You Think This Communication Is** 1 ☐ Necessary to Your Own or Someone Else's Work 2 ☐ Of Only Minor Importance 3 ☐ Not Necessary to the Work of Anyone in the Company 4 ☐ Don't Know	**PLEASE WIND YOUR WATCH AND SET IT FOR THE NEXT SCHEDULED TIME.**

TABLE 17.1 *Major Breakdowns of Time Distribution (Data are percent of eight hour day)*

	Non-supervisory	*Supervisory Ladder*			*Technical Ladder*		*Weighted Average Company-wide*
		First Level	*Second Level*	*Third Level*	*First Level*	*Second Level*	
Not communicating (total)	44	26	19	13	43	40	39
Communicating (total)	56	74	81	87	57	60	61
Listening, speaking	30	48	57	62	28	33	35
Writing	16	17	14	13	17	16	16
Reading	10	9	10	12	12	11	10
Oral communications:							
Unplanned informal contacts	17	26	26	23	12	14	19
Meetings (scheduled 1 day ahead)	8	14	23	31	10	13	10
Telephone	5	8	8	8	6	6	6
Number of individuals covered	59	59	58	19	17	20	232

SOURCE: J. R. Hinrichs. Communications activity of industrial research personnel. *Personnel Psychology*, 1964, *17*, 193–204.

cating, with oral communications alone taking up over one-third (35 percent) of the day. Time spent in communicating orally increases noticeably as one progresses from the nonsupervisory category (30 percent) to third-level supervisor (62 percent).

JOB EVALUATION

Industry is often concerned with making job evaluations rather than job analyses. The essential difference between these two is that job evaluation has as its function the establishment of equitable wage and salary rates. Job evaluations may be obtained by ranking one job as a whole against another job, or by classifying jobs according to previously determined standards. These methods attempt to evaluate the job as a whole. The more common approach, however, is to evaluate job against job by breaking each into its elements. A rating system is usually used, such as that devised by Kress (1939). In this system there are four major items and each item has subdivisions. Job evaluation is achieved by assigning one of five ranks to each of these items. Each job is rated for:

1. Skill
   a. Education
   b. Experience
   c. Initiative and ingenuity

2. Effort
   a. Psychological demand
   b. Mental or physical demand
3. Responsibility
   a. For equipment or process
   b. For material or product
   c. For safety of others
   d. For work of others
4. Job conditions
   a. Working conditions
   b. Unavoidable hazards

Points are assigned to each of the subitems and the total points for a job are then transferred to the many values which establish the wage for that job.

Rothe, after interviewing hundreds of supervisors, has proposed six basic managerial operations (1951): plan, decide, organize and delegate, communicate, lead, and analyze. He believes that decide, organize, and lead are predominantly personality characteristics while the other three are "intellectual." These characteristics are each rated on a sixteen-point scale in relation to the job complexity requirements and can result in a profile-type description of the man's characteristics as well as the job requirements.

The primary reason for including Rothe's research is to illustrate how breaking a job down into its operational functions can lead to a description of the job. Techniques such as this are applicable for jobs from menial tasks to the executive level.

THE EVALUATION OF JOB EVALUATION

Job evaluation techniques have been subjected to research and evaluation, as have all other topics, concepts, and tools in industrial psychology. Always the important questions are asked: What technique or method leads to more valid and reliable results? What are the advantages and disadvantages of one method when compared with another? Industrial psychology is never content to let an armchair proposal become an undisputed fact. This easy but often faulty practice does not meet its standards.

Lawshe and his coworkers have done an excellent job in a series of studies in job evaluation. Lawshe and Satter (1944) did a factor analysis on data obtained from job ratings of hourly paid jobs in three different plants. Skill demands (characteristics possessed by individuals) and job characteristics or aspects of the job were the primary factors uncovered. The former were found to vary from 77.5 percent to 99 percent in the different plants studied and of course are the most important factor.

Lawshe (1945) has also proposed an abbreviated scale consisting of "experience or learning time," "hazards," and "initiative." He indicates that this briefer scale yields results that would include 62 percent of the jobs in the same labor grade and an additional 37.2 percent displaced by only one labor grade.

Lawshe and Maleski (1946) investigated the primary factors operating in a salary rating plan. Skill demands accounted for 95.6 percent of the variance in the total

point ratings. Supervisory demands accounted for 3.7 percent and job characteristics accounted for 0.7 percent.

An abbreviated scale was proposed and included "experience," "complexity of duties," and "character of supervision." About 96 percent of the variance may be attributed to these three items; the other eight items in the longer scale contributed the remaining 4 percent.

Lawshe and Alessi (1946) factor analyzed a different point rating system from the one just reported and found that three primary factors accounted for 96 percent of the variability in total point rating. Skill demands (general), job characteristics, and skill demands (specific) are the factors involved. Using an abbreviated scale made up of "responsibility," "manual skill," and "working conditions" would have yielded results that would have displaced only three jobs by as much as seven cents per hour.

Lawshe and Wilson (1946) analyzed the job evaluation data based upon still another system. Using the factor analysis technique they reached similar findings. It was possible to select three of the original five elements in the scale and these correlated + 0.99 with the original scale.

The essence of Lawshe's work in this area would indicate that abbreviated scales save time and yield results in job evaluation that closely approximate the original but more lengthy techniques.

Chesler (1948) conducted a study to determine not only the reliability of a specific job evaluation manual but also the degree to which different types of job evaluation systems give the same results. In an adequately controlled experiment he compared the results of a 12-item system with two systems including 5 factors, two point rating systems including 15 factors, a point rating system with 13 factors, and a ranking and grade system. The 12-item system included:

1. Work experience
2. Essential knowledge and training
3. Dexterity
4. Character of supervision received
5. Character of supervision given
6. Number supervised
7. Responsibility for funds, securities, and other valuables
8. Responsibility for confidential matters
9. Responsibility for getting along with others
10. Responsibility for accuracy—effect of errors
11. Pressure of work
12. Unusual working conditions

Chesler found that the job analysts who rated the 35 jobs studied obtained similar ratings. The reliability coefficients ranged from + 0.93 to + 0.99. Further, intercorrelation among these different company job evaluation systems ranged from + 0.89 to + 0.97. The obvious conclusion is that different job evaluation systems, when used by trained raters, will lead to similar results. If this is generally true, then it would appear that most job evaluation systems will lead to approximately similar results and that no one system has a distinct advantage over another. Keeping Lawshe's results in mind, one might conclude that the shorter the scale the better.

Chesler (1948), using an abbreviated scale including only four factors—work experience, character of supervision received, character of supervision given, and

responsibility for confidential matters—substantiates the findings of Lawshe and his associates. Abbreviated job evaluation scales are justified from the standpoint of technical and scientific accuracy and economy.

Satter (1949) reports the results of applying two measurement techniques to the problem of building job evaluation scales. He compared the method of "paired comparisons" with the development of a scoring key applied to job specifications. The two methods yielded results which are very similar. The choice of which one is used depends, therefore, upon considerations other than accuracy or validity of measurement.

The scoring key method can be developed in a shorter period of time. The method of paired comparisons is useful when a comparatively large group of judges is available, or where a comparatively small number of new jobs need to be slotted into an already established wage structure.

It appears that the particular number of factors included in a job evaluation scale or the system of scoring does not lead to appreciably different results, provided the job evaluator is trained and knows the meaning of objectivity.

Gomberg (1951) in a critical review of job evaluation from the trade unionist's point of view regards job evaluation as a subordinate tool in collective bargaining. His position is that job evaluation measures, to a limited extent, job content and not job worth. He considers the establishment of wages the responsibility of collective bargaining and does not concede that wages can be established solely via the job evaluation technique regardless of the system used.

Gomberg's view is presented to shed some light on the realities that exist. Too often one studies a technique from only a single point of view. Later, and upon application, one is surprised to find that others do not accept the "rigorous by-products of science."

PSYCHOLOGICAL CONTRIBUTIONS TO JOB ANALYSIS AND EVALUATION

Flanagan (1949b) has proposed a technique known as "critical incidents." His view is that this technique integrates the problem of job definition, selection and classification, and the development of criteria measures, and further makes it possible to conduct research on the criterion problem on a sound and rational basis. This procedure establishes the critical requirement through direct observation by participants or supervisors. A critical requirement is defined as a requirement which is crucial in the sense that it has been responsible for outstandingly effective or definitely unsatisfactory performance of an important part of the job. According to Flanagan, a critical requirement differs from the requirements which appear to be important but in practice have no important effect on performance.

Flanagan (1949a) considers job requirements as stated in terms of critical requirements of the job. These are determined by the collection of reports of behavior which were critical and made a difference between success and failure in the observed work situation.

Unions tend to be suspicious of company rating programs. They generally advocate the seniority principle and prefer collective rather than individual dealing. Grievance and arbitration cases have grown out of contested rating reports.

Habbe (1951) has summarized the procedures used in nine companies. He reports in detail how and why the programs were begun, how they were operated, their values and limitations. The companies are:

1. Aldens, Inc.
2. Atlantic Gelatin Division (General Foods Corporation)
3. Atlantic Refining Company
4. Berger Brothers
5. Inland Manufacturing Division (General Motors Corporation)
6. Mission Appliance Corporation
7. Owens-Illinois Glass Company
8. Standard Oil Company of California
9. Company A

Possibly the greatest value of Habbe's report is the importance attached to training the evaluator not only for rating but also for reporting back to the employee. The communication of results is essential, but it is both a complex and a difficult process, as Covner points out (1953). He favors the nondirective interviewing techniques to prepare the way for understanding the ratings, acceptance of them, and constructive action based upon them. Too many companies have given up rating systems when their raters were not prepared to know how to handle those employees whose self-estimate differed from the estimate by superiors. Habbe's report also includes exact copies of the forms used by the various companies. Anyone planning to introduce rating forms would do well to see this collection conveniently gathered in one report.

CONTRIBUTIONS OF HISTORICAL INTEREST

As mentioned previously, the psychologist's interest in job analysis is the man aspect of the job, and two contributions along these lines have been Viteles' job psychograph (1932) and the occupational ability pattern devised by the Employment Stabilization Research Institute of the University of Minnesota. Although both have more historical than practical significance since neither is now widely used, they do represent interesting departures on the theme of job evaluation.

Viteles' job psychograph consists of a complete statement of the personnel requirements for a job. Theoretically it involves an analysis and specification of special abilities necessary for success. A uniform list of traits is presented and each trait is rated on a five-point scale in accordance with its degree of importance for the specific job.

The job psychograph presents certain difficulties. For example, each trait listed requires specific definition. A major although surmountable difficulty in using this technique is that the person doing the rating must be trained in the understanding and knowledge of the specific traits. Furthermore, up to the present time psychologists have not been able to develop, to their complete satisfaction, valid measures of many of the traits listed. Proficiency, alertness, and initiative are a few examples.

A job psychograph for a power machine operator is shown in Figure 17.5.

Another example of a job psychograph is shown in Figure 17.6. A comparison of these two psychographs shows some differences of the traits involved.

There is an important implication for vocational guidance and selection in the job psychograph: a person without experience in a specific job can be rated for the

	1	2	3	4	5	Remarks
1. Energy		X				
2. Rate of discharge			X			
3. Endurance		X				
4. Control	X					
5. Coordination A				X		
6. Coordination B					X	
7. Initiative		X				
8. Concentration			X			
9. Distribution (of attention)			X			
10. Persistence			X			
11. Alertness		X				
12. Associability		X				
13. Visual discrimination				X		
14. Auditory discrimination	X					
15. Tactual discrimination	X					
16. Kinesthetic discrimination			X			
17. Space perception				X		
18. Form perception		X				
19. Accuracy			X			
20. Visual memory		X				
21. Auditory memory	X					
22. Kinesthetic memory		X				
23. Understanding		X				
24. Understanding (quickness)	X					
25. Observation			X			
26. Planfulness		X				
27. Intelligence	X					
28. Intellect	X					
29. Judgment	X					
30. Logical analysis	X					
31. Language ability	X					
32. Executive ability	X					

Key: 1. Negligible
2. Barely significant
3. Significant
4. Of great importance
5. Of utmost importance

Figure 17.5. Job psychograph of a power-machine operator. (From J. L. Otis and K. R. Smith. Job psychograph in a job analysis. *Occupations*, 1934, *12* (10), 47–54.)

possession of the various traits and the resulting profile matched with the profile for a specific job. When there is a resemblance, job and applicant can be brought together.

Paterson, Dvorak, and others connected with the Employment Stabilization Research Institute of the University of Minnesota, where the occupational ability pattern was devised, were interested in carrying the job analysis technique one step farther. They assumed that abilities on a job can be measured by a representative sampling of psychological tests and, further, that the battery (collection) of tests was such a sampling. With these assumptions they gave a battery of tests to groups of individuals who were successful in specific occupations and obtained scores and profiles of the most typical test performances of people on the job. The occupational ability pattern thus attempts to avoid trait specifications such as Viteles proposes, and it also tries to avoid evaluative and subjective opinions, even though they are

DEPARTMENT	Division: Auditing				
Accounting	1	2	3	4	5
Rate of discharge					XX
Coordination A				X	
Initiative		X			
Concentration		X			
Distribution of attention				X	
Persistence				X	
Alertness		X			
Auditory discrimination	X				
Tactile discrimination	X				
Space perception	X				
Form perception	X				
Accuracy					XX
Visual memory				X	
Auditory memory	X				
Kinesthetic memory		X			
Understanding		X			
Understanding (quickness)	X				
Observation					XX
Planfulness		X			
Intelligence	X				
Judgment	X				
Logical analysis	X				
Language ability	X				
Executive ability	X				

Key: 1. Negligible 3. Significant
2. Barely significant 4. Of great importance
5. Of utmost importance

Figure 17.6. Job psychograph of an auditor. (From B. J. Dvorak. *Differential occupational ability patterns*. University of Minnesota Press, Minneapolis, copyright 1935, renewed 1963 by the University of Minnesota, 11.)

obtained from experts. The tests attempt to sample the abilities and aptitudes of the individual on a rather wide range and cover such material as verbal intelligence, finger dexterity, spatial relations, and eye–hand coordination. A more complete discussion of psychological tests was included in Chapter 4.

There is an even more obvious relationship between this technique and vocational guidance and selection than there was with the job psychograph. A person can be given the battery of tests and his profile then compared with the profile typical of various occupations. He can then be given the job whose profile has the greatest

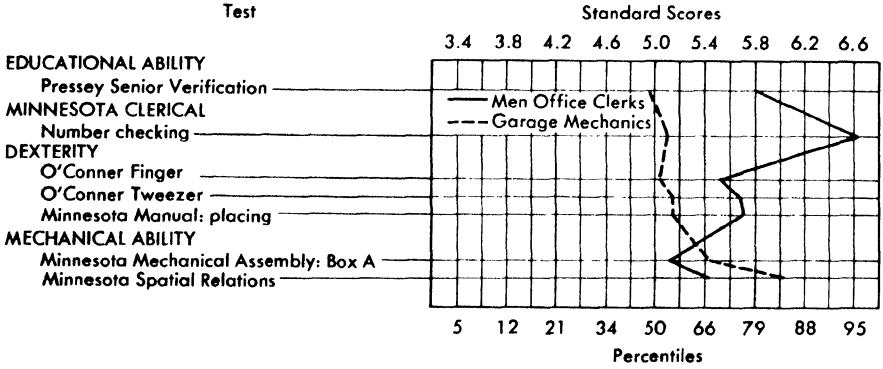


Figure 17.7. Various occupational ability patterns. (From B. J. Dvorak. *Differential occupational ability patterns*. University of Minnesota Press, Minneapolis, copyright 1935, renewed 1963 by the University of Minnesota, 25.)

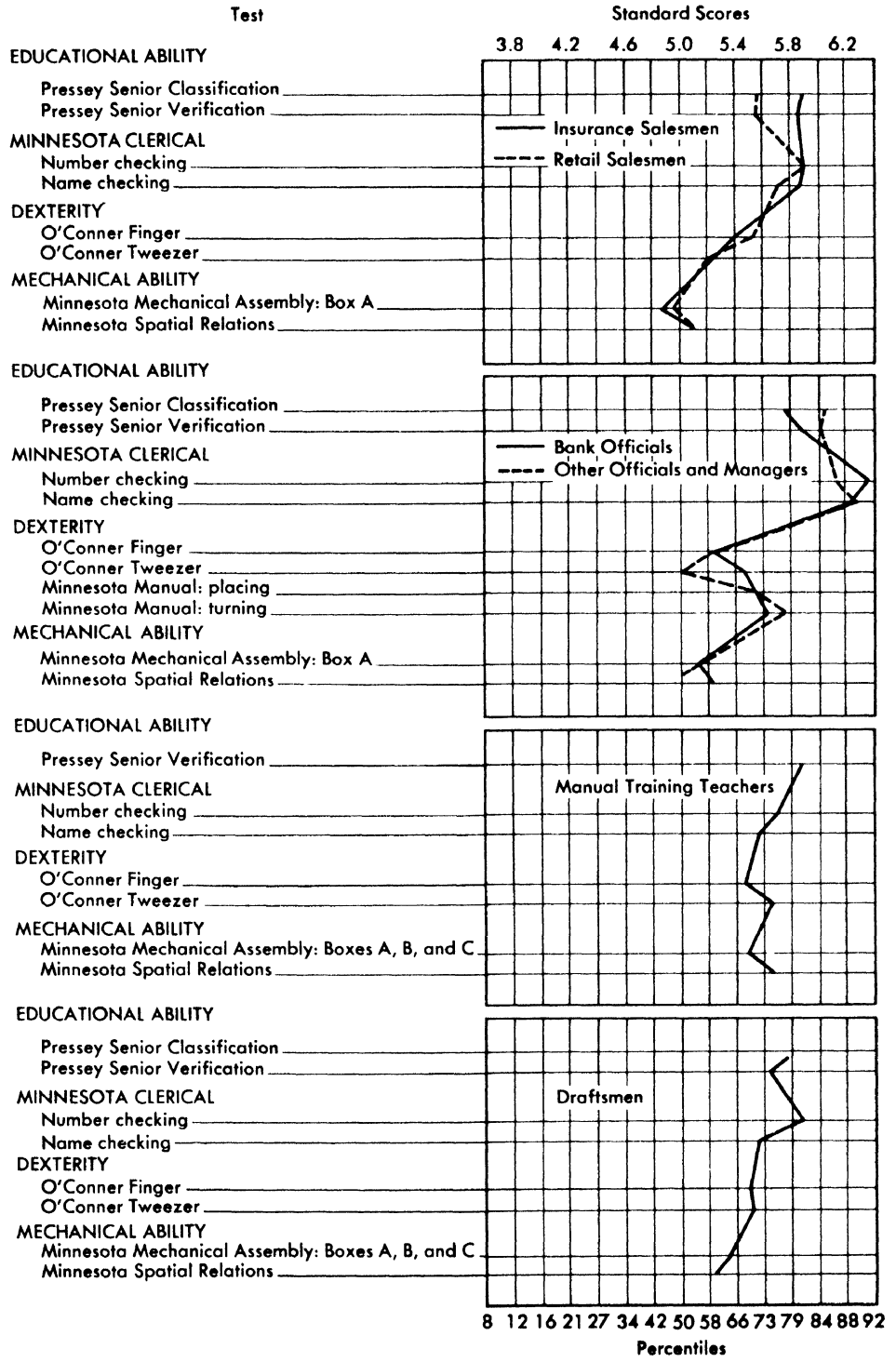


Figure 17.8. Various occupational ability patterns. (From B. J. Dvorak. *Differential occupational ability patterns.* University of Minnesota Press, Minneapolis, copyright 1935, renewed 1963 by the University of Minnesota 34.)

resemblance to his own. Sample occupational ability patterns are presented in Figures 17.7 and 17.8.

OCCUPATIONAL INFORMATION

As a result of the many job analyses that have been made in many industries, a huge body of knowledge known as *occupational information* has accumulated.

One of the very valuable contributions of the United States Employment Service has been the preparation of the *Dictionary of Occupational Titles* by the United States Employment Service. First published in 1939, it was revised and updated in 1949 and again in 1965. An indication of how much occupational information has been accumulated over that time period can be seen simply by comparing the three editions. Note the interesting drop in the number of defined occupations between 1949 and 1965.

D.O.T. Edition	*Number of Defined Occupations*	*Number of Defined Job Titles*
1939	17,452	29,744
1949	22,028	40,023
1965	21,741	35,550

The contents of all definitions in the 1965 edition were verified by over 75,000 job observations. In its most recent edition, the D.O.T. consists of two volumes. Volume I is entitled "Definition of Titles" and Volume II is "Occupational Classification." It has become the standard reference source not only for the U.S. Employment Service but also for other governmental agencies and the Canadian government.

Standardization is necessary in describing jobs and occupations if the maximum utilization of the available information is to be achieved. The D.O.T. has been a great aid in this direction. For example, the occupational information gathered in an industrial program carried out by the U.S. Employment Service has led to interesting and somewhat unexpected results, such as the fact that a stone planer is able to do the job of a metal planer. The job analysis for these two jobs showed that the skills, aptitudes, and experience for both of them were essentially the same. The fact that the stone planer worked on different materials in a different industry was not as important as might be assumed. In another instance a close relationship was found between the job of a plasterer and that of an asbestos pipe coverer; the metamorphosis was completed easily by teaching the plasterer how to sew. Both the Army and the Navy during World War II found that knowledge of occupational information expedited transfers from civilian to military occupations.

Another very important source of occupational information is issued every two years by the United States Department of Labor under the title *Occupational Outlook Handbook*. Its basic purpose is to provide employment information in various occupations for use in occupational guidance, that is, to provide the prospective job hunter with information about opportunities in different occupations. Also, many state employment services have descriptive brochures which provide information on occupations within the state civil service system.

Dr. Carroll L. Shartle has compiled a valuable source book entitled *Occupational Information* (1952). He has presented the available material in such a manner as to show clearly the need for occupational information on the part of industry, education, and the community. The book also contains pertinent facts on occupation trends and the labor supply and demand.

Occupational information based upon job analyses has resulted in information about "occupational families," or groups of related occupations with similar characteristics. This material is useful for business organizations that are considering the transfer and upgrading of personnel and for communities seeking new industries. "Occupational family" information can also help the individual who is seeking a job.

Another value of occupational information is the increased knowledge it gives people about jobs. There are very few people who can name more than a few hundred occupations, and yet there are thousands of different ones.

An article by Shartle (1946) serves as an illustration of the lack of knowledge that most people have concerning specific jobs. The need for occupational information is well brought out by examining the list of varied positions held by psychologists. The average person, and to some extent even the sophisticated college student, usually believes that a psychologist either teaches at a college or gives tests. The following list indicates some of the positions held by qualified psychologists in various agencies, institutions, and organizations. This list makes it obvious that some of the positions are not clear; that is, they lack definition and description. A list of positions is useful, but it must be supplemented by additional relevant material.

College teaching
College counseling
Educational research
Public schools—psychologist
Clinical psychologist
Child guidance
Psychologist—institution for feeble-minded
Psychologist—hospital for insane
Psychologist—penal institution
Court psychologist
Personnel examiner
Rehabilitation training officer
Personnel psychologist
Consulting psychologist
Public opinion survey psychologist
Psychometrician
Employment interviewer
Employment counselor
Occupational analyst
Vocational adviser
Vocational counselor
Research psychologist

CLUSTERING OR GROUPING OCCUPATIONS

Earlier we mentioned that factor analysis was being used to try to identify underlying dimensions of various jobs or job types. Another use of factor analysis as applied to job analysis is for grouping jobs into meaningful subcategories or clusters. In many cases it is important to know the similarity between different jobs and different job "classes" in order to develop a truly functional job classification system.

Orr (1960) took 4000 jobs, selected from the D.O.T., for which ratings on eight different aptitudes had been previously obtained by Trattner, Fine, and Kubis (1955). The aptitudes and their definitions are given in Table 17.2.

TABLE 17.2 *The Aptitudes and Their Definitions*

Aptitude
Intelligence: General learning ability: the ability to "catch on" or understand instructions and underlying principles; ability to reason and make judgments.
Verbal: Ability to understand meanings of words and ideas associated with them, and to use them effectively; to comprehend language and understand relationships between words and to understand meanings of whole sentences and paragraphs.
Numerical: Ability to perform arithmetic operations quickly and accurately.
Spatial: Ability to comprehend forms in space and understand relationships of plane and solid objects; may be used in such tasks as blueprint reading and in solving geometry problems; frequently described as the ability to "visualize" objects of two or three dimensions.
Form Perception: Ability to perceive pertinent detail in objects or in pictorial or graphic material; to make visual comparisons and discriminations and see slight differences in shapes and lengths.
Clerical Perception: Ability to perceive pertinent detail in verbal or tabular material; to observe differences in copy, to proofread words and numbers, and to avoid perceptual errors in arithmetic computation.
Motor Coordination: Ability to coordinate eyes and hands or fingers rapidly and accurately in making precise movements with speed; ability to make a movement response accurately and quickly.
Finger Dexterity: Ability to move the fingers, and manipulate small objects with the fingers, rapidly or accurately.
Manual Dexterity: Ability to move the hands easily and skillfully; ability to work with the hands in placing and turning motions.

SOURCE: D. B. Orr. A new method for clustering jobs. *Journal of Applied Psychology,* 1960, *44,* 44–59.

Submitting these data to factor analysis revealed six different clusters or groupings of jobs:

Cluster I. This cluster was characterized by high-level technical, supervisory, and mechanical jobs. The aptitudes required are Intelligence, Verbal, Numerical, Spatial, and Form Perception. Clerical was rated low.

Cluster II. This cluster was composed of extremely low-level jobs, primarily unskilled.

Cluster III. This cluster was composed largely of fairly high-level skilled jobs, mostly mechanical and artisan types. The emphasis was on Spatial Form Perception, Manual Dexterity, and Finger Dexterity.

Cluster IV. The jobs in this cluster were of a very high level with respect to their requirements of Intelligence and Verbal. Supervisory jobs were contained in this cluster, and there was some emphasis on clerical abilities.

Cluster V. The emphasis here was on clerical jobs and supervisory jobs at a lower level than those in Cluster I. The aptitudes required in the greatest amounts were Intelligence, Verbal, Numerical, and Clerical.

Cluster VI. This cluster featured mechanical-manual jobs of a medium grade of skill. Intelligence, Finger Dexterity, and Form Perception were required, but no aptitude was needed in more than an average amount.

Gonyea and Lunneborg (1963) also report a clustering of 22 occupations using similar methodology. Table 17.3 shows the results of their analysis. They found their 22 occupations fell into five groups:

Group A: Included Buyer, Office Manager, Personnel Manager, Interior Decorator, Accountant, Insurance Salesman, and Secretary. This was called a "Business" group.

Group B: Aviator, Automobile Mechanic, Surveyor, Radio Operator, Policeman, and Engineer. This was called a "Masculine" group.
Group C: Artist, Writer, Interior Decorator. This was called an "Esthetic" group.
Group D: Policeman (again), Social Worker, Lawyer, Physician, Personal Counselor, and Teacher. This was called a "Service" group.
Group E: Medical Laboratory Technician, Chemist, Physician, Engineer, Automobile Mechanic, Radio Operator. This was considered a "Scientific" group.

TABLE 17.3 *Results of Factor Analysis of Perceived Similarity of 22 Occupations*

Occupation	*Cluster*				
	A	B	C	D	E
Accountant	0.59	0.00	−0.05	−0.11	0.14
Artist	−0.08	0.08	0.76	0.07	0.04
Automobile Mechanic	0.12	0.56	−0.05	−0.10	0.32
Aviator	−0.16	0.78	0.06	0.04	0.14
Buyer	0.65	0.07	0.08	−0.09	−0.16
Chemist	−0.07	0.12	0.04	0.08	0.66
Engineer	−0.01	0.45	0.03	0.01	0.38
Insurance Salesman	0.54	0.12	−0 01	0.14	−0.16
Interior Decorator	0.39	0.03	0.57	−0.12	−0.01
Lawyer	0.11	0.05	−0.03	0.53	0.04
Medical Laboratory Technician	0.08	0.08	−0.06	0.03	0.74
Office Manager	0.64	0.00	−0.12	0.05	0.06
Personal Counselor	0.28	−0.09	0.06	0.49	0.02
Personnel Manager	0.53	−0.03	−0.07	0.23	0.03
Physician	−0.09	−0.04	−0.06	0.50	0.57
Policeman	−0.05	0.53	−0.17	0.59	−0.10
Radio Operator	0.07	0.61	0.04	−0.09	0.25
Secretary	0.51	−0.07	0.12	−0.01	0.06
Social Worker	0.16	0.01	0.11	0.58	−0.03
Surveyor	0.06	0.53	0.09	0.07	0.09
Teacher	0.07	−0.08	0.20	0.47	0.16
Writer	0.10	−0.06	0.64	0.13	−0.06

SOURCE: G. G. Gonyea and C. E. Lunneborg. A factor analytic study of perceived occupational similarity. *Journal of Applied Psychology,* June 1963, *47*(3), 166–172.

The clustering of occupations using procedures such as factor analysis appears to be a very rich way of developing job families which can then be used as a basis for other kinds of research. We need to know a great deal more about jobs than we do at present. Unless we can define jobs and tasks adequately so that we can develop a formal language to describe their similarities and their differences, the entire field of industrial psychology is going to be at a disadvantage.

REFERENCES

Berkshire, J. R., and R. W. Highland. Forced choice performance rating—methodological study. *Personnel Psychology,* 1953, 6, 355–378.

Chesler, D. J. Reliability of abbreviated job evaluation scales. *Journal of Applied Psychology*, 1948(a), 32, 622–628.

Chesler, D. J. Reliability and comparability of different job evaluation systems. *Journal of Applied Psychology*, 1948(b), 32, 465–475.

Covner, B. J. The communication of merit ratings. *Personnel*, September 1953.

Dubin, R., and S. L. Spray. Executive behavior and interaction. *Industrial Relations*, 1964, 3, 99–108.

Dunnette, M. D., and G. W. England. A checklist for differentiating engineering jobs. *Personnel Psychology*, 1957, 10, 191–198.

Dunnette, M. D., and W. K. Kirchner. A checklist for differentiating different kinds of sales jobs. *Personnel Psychology*, 1959, 12, 421–430.

Dvorak, B. J. *Differential occupational ability patterns*. University of Minnesota Press, Minneapolis, 1935.

Estimates of worker trait requirements for 4000 jobs as defined in the Dictionary of Occupational Titles. U.S. Department of Labor, Bureau of Employment Security, Washington, D.C., 1958.

Flanagan, J. C. Critical requirements: A new approach to employee evaluation. *Personnel Psychology*, 1949(a), 2, 419–425.

Flanagan, J. C. A new approach to evaluating personnel. *Personnel*, July 1949(b), 26, 35–42.

Ghiselli, E. E. The forced choice technique in self description. *Personnel Psychology*, 1954, 7, 201–208.

Gomberg, W. A trade unionist looks at job evaluation. *Journal of Applied Psychology*, 1951, 35, 1–7.

Gonyea, G. G., and C. E. Lunneborg. A factor analytic study of perceived occupational similarity. *Journal of Applied Psychology*, June 1963, 47(3), 166–172.

Habbe, S. Appraisal of job performance. *Studies in Personnel Policy #121*. National Industrial Conference Board, New York, 1951.

Hemphill, J. K. *Dimensions of executive positions: A study of the basic characteristics of the positions of ninety-three business executives*. Ohio State University, Bureau of Business Research, Columbus, 1960.

Hinrichs, J. R. Communications activity of industrial research personnel. *Personnel Psychology*, 1964, 17, 193–204.

Horne, J. H., and T. Lupton. The work activities of "middle" managers—an exploratory study. *Journal of Management Study*, 1965, 2, 14–33.

Kress, A. L. How to rate jobs and men. *Factory Management*, 1939, 97, 60–65.

Lawshe, C. H., Jr. Studies in job evaluation. II. The adequacy of abbreviated point ratings for hourly paid jobs in three industrial plants. *Journal of Applied Psychology*, 1945, 29, 177–184.

Lawshe, C. H., Jr., and S. L Alessi. Studies in job evaluation. IV. Analysis of another point rating scale for hourly paid jobs and the adequacy of an abbreviated scale. *Journal of Applied Psychology*, 1946, 30, 310–319.

Lawshe, C. H., Jr., and A. A. Maleski. Studies in job evaluation. III. An analysis of point ratings for salary paid jobs in an industrial plant. *Journal of Applied Psychology*, 1946, 30, 117–128.

Lawshe, C. H., Jr., and G. A. Satter. Studies in job evaluation: I. Factor analysis of point ratings for hourly paid jobs in three industrial plants. *Journal of Applied Psychology*, 1944, 23, 189–198.

Lawshe, C. H., Jr., and R. F. Wilson. Studies in job evaluation. V. An analysis of the factor comparison system as it functions in a paper mill. *Journal of Applied Psychology*, 1946, 30, 426–434.

Morsh, J. E. Job analysis in the United States Air Force. *Personnel Psychology*, 1964, 17, 7–17.

Orr, D. B. A new method for clustering jobs. *Journal of Applied Psychology*, 1960, 44, 44–59.

Otis, J. L., and K. R. Smith. Job psychograph in job analysis. *Occupations*, 1934, 12(10), 47–54.

Palmer, G. J., and E. J. McCormick. A factor analysis of job activities. *Journal of Applied Psychology*, 1961, *45*, 289–294.

Prien, E. P. Development of a supervisor position description questionnaire. *Journal of Applied Psychology*, 1963, *47*, 10–14.

Radom, M. Picking better foremen. *Factory Management & Maintenance*, October 1950.

Rothe, H. F. Matching men to job requirements. *Personnel Psychology*, 1951, *4*, 291–301.

Satter, G. A. Method of paired comparisons and a specification scoring key in the evaluation of jobs. *Journal of Applied Psychology*, 1949, *33*, 212–221.

Shartle, C. L. Occupations in psychology. *American Psychologist*, 1946, *1*, 559–582.

Shartle, C. L. *Occupational information*, 2nd ed. Prentice-Hall, Englewood Cliffs, N.J., 1952.

Trattner, M. H., S. A. Fine, and J. F. Kubis. A comparison of work requirement readings made by reading job descriptions and by direct job observation. *Personnel Psychology*, 1955, *8*, 183–194.

Viteles, M. S. *Industrial psychology*. Norton, New York, 1932.

Zerga, J. E. Job analysis, a résumé and bibliography. *Journal of Applied Psychology*, 1943, *27*, 249–267.

18 ACCIDENTS, SAFETY, AND FATIGUE

Accidents not only are costly to our industrial economy but they also result in anguish, pain, or even death to the unfortunate individuals involved. Sometimes accidents are attributed to physical causes in the environment, and sometimes they are considered to be caused by any of a number of items related to the human factor. The tendency has been to minimize the role of "luck" in accidents and to look for a specific cause. The chart in Figure 18.1, which was prepared by the Metropolitan Life Insurance Company, lists the major causes of industrial accidents. However, this chart deals merely with the physical aspect of accidents and does not take into consideration the psychological factors at work in the individual at the time of the accident. The human factor as a contributing cause of accidents is emphasized in Figure 18.2 which shows the primary causes of accidents among the employees of the Cleveland Railway Company; this study was done in cooperation with the Metropolitan Life Insurance Company (Accident-Prone Employee, 1945).

This particular study, similar to many of those which preceded it, brings out some of the personal factors involved in accidents.

> Analysis of the accident records led to the discovery that a relatively small group of motormen were involved in a large percentage of all accidents reported, although operating under practically identical conditions as the other larger groups of trainmen whose accident records were, in most cases, excellent. This finding apparently confirmed statements of psychologists and psychiatrists that accidents do not distribute themselves by chance but that they happen frequently to some men and infrequently to others as a logical result of a combination of circumstances. Those individuals who because of certain mental, psychological or physical defects fail to control a situation leading to an accident when it arises, usually become involved, while those possessing the necessary physical and mental requirements show little susceptibility to accident.

These findings led to the belief that the incidence of accidents was an individual problem and that accidents could be reduced by studying the individuals who had them and removing the causes.

The treatment of people who were likely to have accidents was divided into three main categories: medical assistance, personality readjustment, and operating defects.

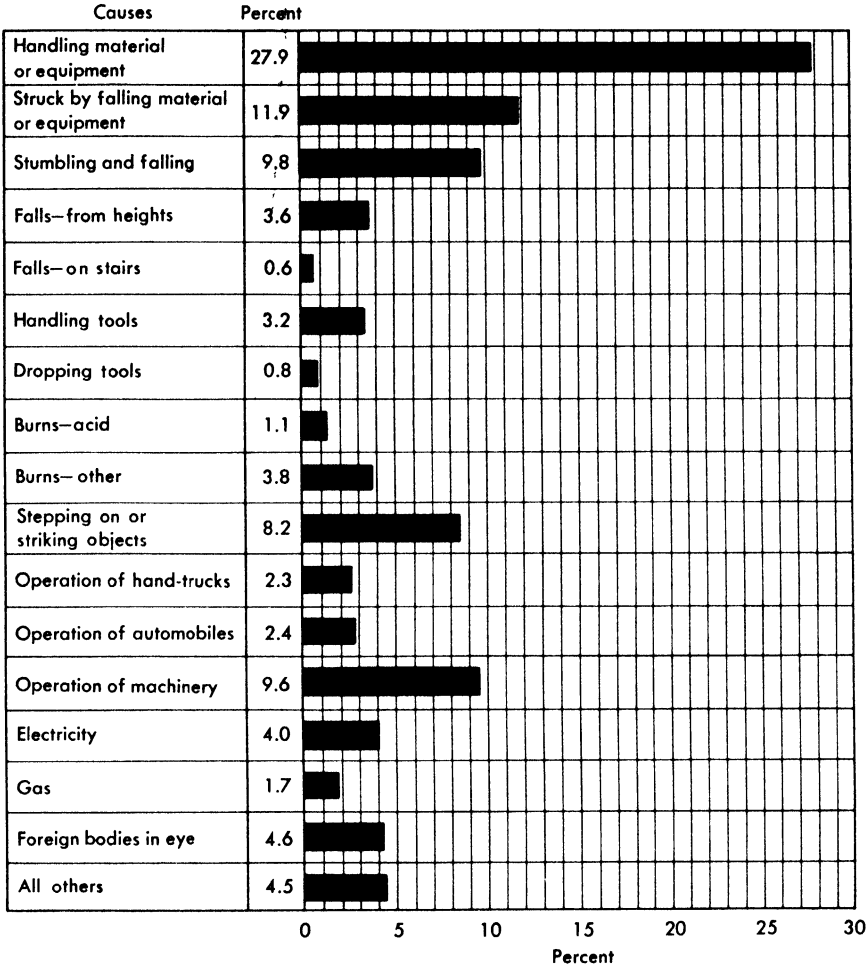


Figure 18.1. Causes of accidents. (Courtesy Metropolitan Life Insurance Company.)

MEDICAL ASSISTANCE

Physical difficulties were found to be of primary importance in 12 percent of the cases. Defective vision, high blood pressure, and chronic appendicitis are examples of the type of disability which, when corrected, result in reduced accident rates.

PERSONALITY READJUSTMENT

Faulty attitude, irresponsibility, and other personality problems were involved in 22 percent of the cases. Although the railway company handled the "therapy" in a rather crude manner, successful changes in the individual employees were reported. The company had the superintendent talk to the employee and give him "a clear understanding of his personality difficulties and . . . point out to him frankly why they interfered with satisfactory performance on the platform and what

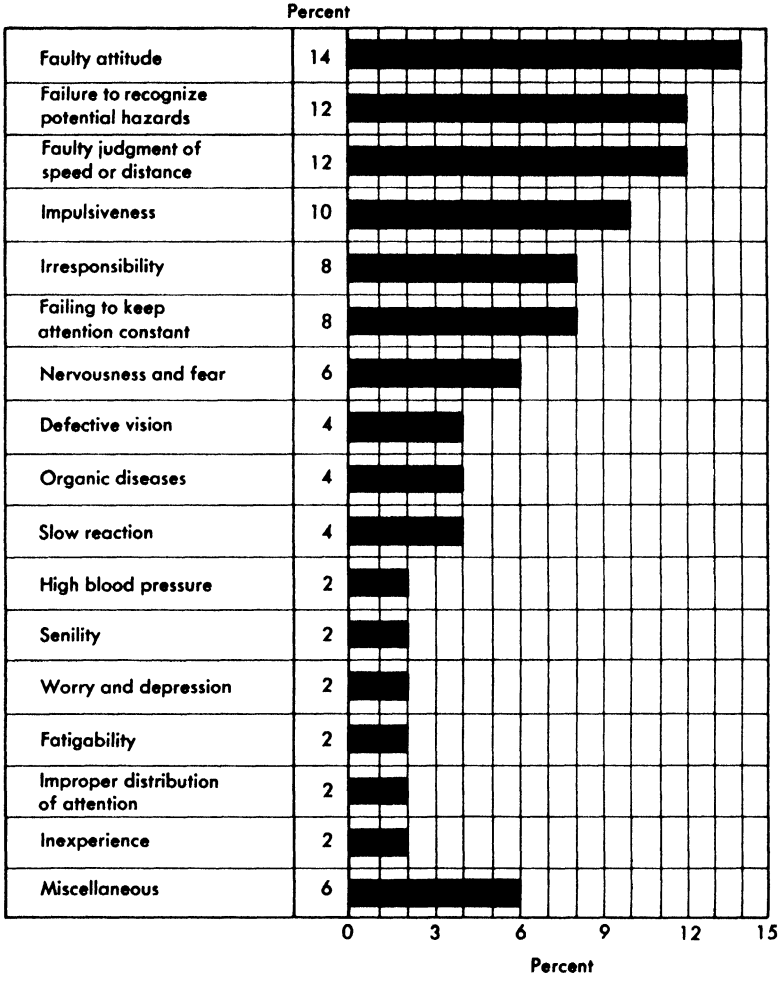


Figure 18.2. Primary causes of accidents. (Courtesy Metropolitan Life Insurance Company.)

the probable outcome of them would be unless a change in general outlook and behavior on the job were shown." Most of these men were placed on probation.

OPERATING DEFECTS

The remaining 66 percent of the cases were due to poor job performance, i.e., failure to recognize potential hazards, faulty judgment of speed and distance, and lack of attention. These shortcomings were overcome by retraining the men.

The results of this study were gratifying to the railway company. Thirty-three (or 75 percent) of the men disappeared from the "accident-prone" group because their rate fell below the one accident per 1000 car-miles operated. The accident rate for the entire group dropped from 1.31 accidents per 1000 car-miles to 0.75, a decrease of 43 percent.

ACCIDENT PRONENESS PRINCIPLE

With slight modification the statement "Accidents do not distribute themselves by chance but . . . happen frequently to some men and infrequently to others as a logical result of a combination of circumstances" has become a generalization. This statement embodies the principle of accident proneness.

Accident proneness is a hypothesis about human behavior which says that accident behavior is not a random phenomenon. Rather, it is a consistent characteristic which can be predicted. The accident proneness concept does not hold that accidents are caused strictly by chance. One can represent this concept with the simple equation

$$A_T = a_e + a_p$$

where A_T = total number of accidents
a_e = accidents caused by chance factors
a_p = accidents attributable to personal characteristics

Those who believe in accident proneness believe that a_p is an important source of accidents.

Mintz and Blum, who have critically evaluated the literature in the field, have come to the conclusion that accident proneness is overrated (1948). A correct interpretation demands that the second half of the statement be worded as follows: Accidents happen frequently to some men, that is, *some men have more accidents than would be expected by chance*, and infrequently to others, that is, *other men have fewer accidents than would be expected by chance.*

In accordance with chance expectancy, some men will have no accidents, some will have one accident, some two accidents, and some three or more accidents. In other words, chance will result in a distribution of accidents, and it is wrong to assume that in accordance with chance expectancy all men should have equal numbers of accidents.

To establish or support the principle of accident proneness, three methods can be used. In one method, the distribution of the total number of accidents in a population is compared with the distribution that would be expected if only chance factors operated. A comparison of these two distributions makes it possible to determine whether accidents happened more frequently to some men than would be expected by chance. The second method is to study the individuals and the number of accidents they have in two successive periods. A tendency for these people to have similar numbers of accidents can be regarded as evidence in favor of accident proneness. The third method is to compute the correlation coefficient of the accident records of a group for two periods.

Mintz and Blum (1949) found a considerable amount of careless reporting, illogical reasoning, and unfamiliarity with the statistical theory underlying accident proneness. Their analysis shows that 60 to 80 percent of the accidents reported appear to be attributable to unpredictable factors and the remaining 20 to 40 percent to the component of accident liability which includes both personal characteristics and environmental conditions contributing to accident records. Thus accident proneness does not seem to be the major factor in understanding or attacking the problem of accidents.

The most frequently offered evidence for the existence of accident proneness is

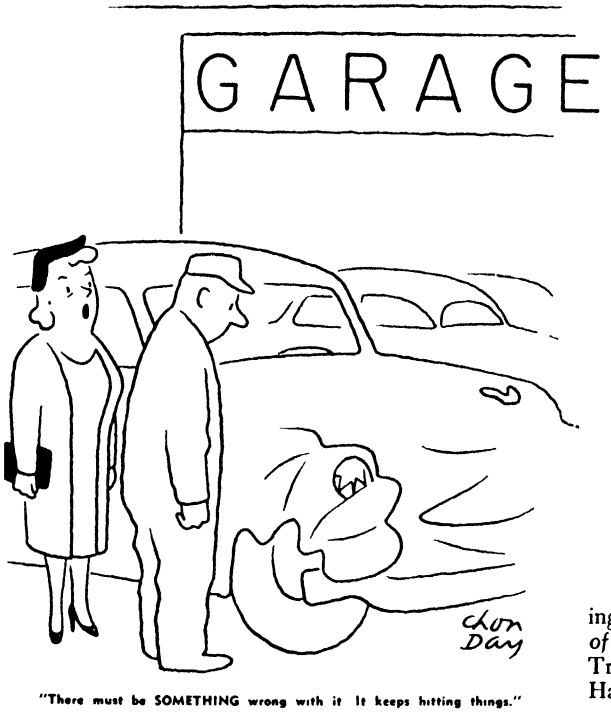


Figure 18.3. (From "The Passing Scene." *The Travelers 1954 book of street and highway accident data.* Travelers Insurance Companies, Hartford.)

the fact that a small percentage of the population has a large percentage of the total number of accidents. Such statements, by themselves, do not substantiate accident proneness. The obvious error in them is made clear by the following hypothetical situation: Two hundred employees have 100 accidents. If every employee who has an accident has only one, there will be the opportunity for only 100 employees to have an accident record, and yet this leads to the conclusion that 50 percent of the employees have 100 percent of the accidents. This situation is extremely unreal, for there is no reason to believe that each employee should have one accident. According to chance,[1] approximately 121 in a population of 200 should have no accidents, 61 people should have one accident, 15 people should have two, and 3 people should have three accidents. On this basis, the expectancy is that 9 percent of the population will have 39 percent of the accidents and that 39.5 percent of the population will have 100 percent of the accidents. This distribution is due solely to chance expectancy; there is no evidence of accident proneness.

If a deck of cards is dealt to four people it sometimes happens that one of them will get six, seven, or more hearts instead of three or four. Such a distribution is normally attributed to chance.

Statements to the effect that 10 percent of a population have 30 percent of the accidents or that 25 percent have 75 percent of them must be regarded as inconclusive unless the total population and the total number of accidents are known. Only when we have these figures can we establish the extent to which the distribution of accidents is due to chance expectancy and the extent to which other factors such as accident proneness enter,

[1] These data have been computed by A. Mintz by means of the Poisson distribution or "law of small numbers."

An early study which is often referred to and which finds evidence for accident proneness was conducted by Greenwood and Woods (1919). The data in this study are rather complete, which is more than can be said about many of the more recent studies. These authors develop the statistical formula based upon a theory of accident proneness and this must be recognized as a contribution. Table 18.1 presents some of the data from this study.

If the principle of accident proneness is to be upheld, more people should have no accidents than would be predicted by chance. In three of the five groups this holds true, but in the other two it does not. In the three groups where it does hold true, the percentages are 6, 3, and 10 percent. These percentages are to be attributed to accident proneness in certain people, but they must also be attributed to other causative factors such as length of employment, job hazard, training, etc. In any event, accident proneness as a cause for accidents is apparently not so formidable as some authorities have made it. Although Greenwood and Woods do establish the

Table 18.1 *Accident Distribution*

Group	*Total Population*	*Total Number of Accidents*	*Percent of People Having No Accident*	*Percent of People Having No Accident on Chance Expectancy*
1	750	432	53	56
2	580	278	61	62
3	648	303	68	62
4	584	253	73	70
5	414	200	71	61

Source: M. Greenwood and H. M. Woods. The incidence of industrial accidents upon individuals with special reference to multiple accidents. *Industrial Fatigue Research Board No. 4*, 1919.

principle of accident proneness, they do not indicate the degree to which it is responsible for accidents; many others who have done research in this field imply that the extent of accident proneness has been exaggerated.

Another erroneous use of accident proneness is seen in the arbitrary classification of people who have more than the average number of accidents as accident prone. Such a method was used in the Cleveland Railway study, and at least one textbook defines accident proneness as being present in people who have two or three times as many accidents as the average person has. In Table 18.1 the average number of accidents per person is approximately 0.5; hence in an arbitrary classification a person having one or two accidents would be accident prone. This is not necessarily true. The hypothetical situation previously mentioned shows clearly that some people may have two or three accidents solely on the basis of chance expectancy. This critical attitude toward the concept of accident proneness is in part supported by Cobb (1940) and Johnson (1946). More work along these lines should be encouraged.

Arbous and Kerrich have also made an exhaustive search of the literature on the topic of accident statistics and accident proneness (1951). Like Mintz and Blum, they believe that the knowledge of this concept has hardly proceeded further than the early studies of Greenwood and in some respects the topic has actually suffered

a reverse because of the misunderstanding. Their article is an exceedingly well-written and readily understandable one and should be a standard reference for the more serious student in this area. They raise such important points as the correlation between minor accidents for two successive periods, the correlation between major accidents for two successive periods, the correlation between minor and major accidents, and the correlation between different types of accidents. They conclude with the comment: "'Accident-proneness Percy' is a figment of the imagination resulting from wishful thinking." This points to a view of accident prevention since it means that an individual's liability or proneness to accidents (when such a thing exists) in one set of circumstances will give little indication of proneness in another. The cogent remark that might best summarize the point of view of Arbous and Kerrich is as follows: "This does not mean that accident proneness does not exist, but that so far we have not succeeded in defining it, assessing its dimensions and constituent elements, nor evolved a technique for putting it to practical use."

TOWARD A THEORY OF ACCIDENTS

Drake (1937) has proposed an interesting theory to explain the influence of accident proneness. According to him, "Where the perception level is equal to, or higher than the motor level, the employee is relatively a safe worker. But where the perception level is lower than the motor level, the employee is accident prone and his accident proneness becomes greater as this difference increases."

Drake demonstrated the feasibility of this theory in a study made on a group of factory workers. He gave them three motor or manipulative tests and two tests of visual inspection and sorting. The raw scores were converted into percentiles and the difference in motor level as compared with perception level was obtained. The study showed that employees who were faster on the motor tests than on the perception tests had more accidents and, conversely, those who were faster on the perception tests than on the motor tests had fewer accidents. Drake reports that selecting new employees for comparable tasks on the basis of performance on these tests reduced accidents 70 percent. He believes that this theory is applicable to all types of accidents including automobile accidents. Although the present authors do not know of any check on these claims, they regard them as interesting. However, an obvious limitation of this theory is the lack of consideration given to the individual's personality and to contributing emotional factors.

A slightly different approach to the problem of accidents has been proposed by Kerr (1950). Provided one does not take seriously the concept of accident proneness as applied to factory departments, the work of Kerr is worthy of consideration. He found that accidents occur with greatest frequency in those factory departments with low intracompany transfer mobility, low promotion probability, and high noise level. Accident severity was found in departments with such characteristics as predominance of males rather than females, low promotion probability, low suggestion record, nonyouthfulness of employees, and high average tenure of workers.

The tendency for departments lowest in promotion probability to be high in both severity and frequency of accidents may be related to indifference to the work environment and may contribute to accidents, according to Kerr. The correlation between accident frequency and severity was + 0.64. This fact points to the need to know which record is considered the criterion. Many studies merely accept the criterion that is most readily available and then go on to prove or disprove a point—

which may be impossible to prove or disprove. An exaggeration of this is offered by the cartoon presented in Figure 18.3.

ROLE OF MACHINES IN ACCIDENTS

Earlier we stated that the equation

$$A_T = a_e + a_p$$

was one way of viewing the accident proneness concept. Actually, such an equation is far too simple; for example, it ignores the role of the machine or environment in accidents. Some cars are more dangerous than others—some machines are more risky to operate. Accidents can be caused by these factors, which we might call *system* factors. Thus

$$A_T = a_e + a_p + a_s$$

Also, there may well be an *interaction* of people types with system types, such that some people are suited for some systems while others are suited for other systems. Thus, a very short man in a low car may tend to have more accidents than if he were in a high car, while a tall man might have the reverse tendency (we do not really know, of course). We could represent this in the accident equation as

$$A_T = a_e + a_p + a_s + a_{p \times s}$$

Undoubtedly other factors can be put into the equation as well.

Actually, the role of systems in accidents has probably not been given enough attention. Admittedly we have seen some noticeable advances in car safety features in recent years with padded dashes, seat belts, etc., but even so it is apparent that the public is not overly concerned with safety items when buying. No one seems to want an ugly, safe car. Thus designers do not feel compelled to sacrifice glamour features to bring about safety—sales might suffer as a consequence! Such safety standards as are finally going into effect are the result of governmental directives, not of the consumer's rush to the "safer" car. These governmental directives may be more political than factual. If they are, we ask the question, "Who benefits and who suffers?"

ACCIDENT DATA

Possibly one of the greatest sources of limitation in studying accidents is the accuracy, or rather the lack of it, in the reporting of accidents. Anyone who has done any work at all in industry knows that many variables enter into the decision to record an accident. Fear of punishment or the likelihood of gaining an insurance claim very often distort accident records in either or both directions. Researchers should seriously study the manner of ascertaining an accident and its recording before accepting the data on face value and using statistics to explain what may not exist in the first place. This simply means that some accidents are not recorded and others are. Somehow this should be known to the researcher before he embarks on a theory to explain the phenomenon.

Accident proneness as a means of understanding the occurrence of accidents is useful provided that we recognize clearly that human factors and the physical environment may contribute to accidents in the absence of accident proneness.

ACCIDENT REDUCTION

Raising a question as to the importance of accident proneness does not mean that we should take the fatalistic attitude that little can be done to reduce the number of accidents caused by the human factor. Nothing could be farther from the truth. The fact that accident distribution is in part determined by chance means that many unrelated factors are operating, not that accidents do not have causes; it means that there are multiple causes and all of them must be considered. The principle of accident proneness merely gets at one of the causes, that is, the human factor in individuals who are accident prone. The human factor must also be considered in accidents to individuals who are not necessarily accident prone. The environment, the physical factors of machine design and wear, and innumerable other factors must also be included. Recognizing that there are many elements which contribute to an accident and considering differences in accident proneness as a minor rather than a major element will make it possible to accomplish more in reducing accidents. Even a person who is accident prone can cut his accident record by attacking the other factors which contribute to his accidents. Many people who are not accident prone may have one or more accidents. These people too can reduce their accidents by altering the physical conditions conducive to them. For example, accidents are reduced by installing safety islands on heavily traveled roads and providing safety guards on moving machinery.

There are two approaches to accident reduction as related to the human factor. The first is usually referred to as the actuarial method and the second involves educational campaigns.

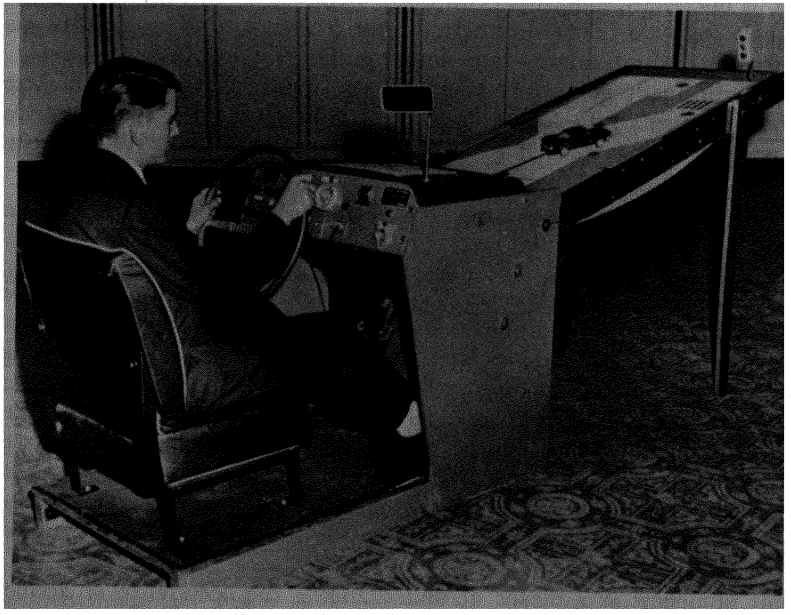


Figure 18.4. Driver training apparatus.
(Courtesy American Automobile Association.)

THE ACTUARIAL METHOD

A good example of the actuarial method is the tremendous amount of data on automobile accidents which have been compiled by various individuals and agencies. For some excellent summaries of this material the reader is referred to both De Silva (1939) and Nader (1965). In brief, the actuarial method involves studying accident statistics to determine, based upon actual data, those things which seem to be related to accident frequency. For example, McFarland and Moseley's (1954) data presented in Table 18.2 clearly indicate that accident repeaters committed many more violations than did accident-free drivers. The number of violations of accident repeaters is much greater than proportional expectations. McFarland is inclined

TABLE 18.2 *Motor Vehicle Violations Recorded for Driver Sample*

	Number of Drivers	
	Accident Free N = 59	*Accident Repeater* N = 88
Minor Offense		
Leaving vehicle running and unattended	1	1
Driving within 8 feet of streetcar stopped for passengers	7	20
Not reasonably right for vehicle coming in opposite direction	4	15
Not keeping right half of road when view obstructed	2	5
Crossing throughway without stopping	7	16
Failure to obey traffic signal	1	11
Speeding	24	50
Left of streetcar	0	2
Violation of traffic rules	6	8
Mechanical defect	0	1
Without proper lights	1	5
Without proper brakes	0	2
Without proper muffler	0	1
No vehicle inspection sticker	4	19
Improper operation	5	7
Negligent collision	3	5
Serious Offense		
Operating under influence of liquor	8	13
Operating so as to endanger lives and safety	8	15
Going away after injury to property	1	7
Going away after injury to persons	0	1
Operating after license suspension	0	7
Operating without proper registration	0	10
Operating without being properly licensed	4	9
Executive Order #35 (wartime speed)	3	2
Violation of compulsory insurance law	0	6
Operating without authority	0	4

SOURCE: R. A. McFarland and A. L. Moseley. *Human factors in highway transport safety.* Harvard University Press, Cambridge, Mass., 1954.

to believe that a man drives as he lives. Speed in private driving and violations which reflect attitude toward authority seem to be characteristic of repeaters.

However, Forbes (1939) has presented rather convincing data which show that accident "repeaters" account for only a small portion of the total traffic accident problem. His sample consisted of 29,351 drivers, and he found that *less than 4 percent* of the accidents in a three-year period involved repeaters from the prior three-year period! Thus, accident involvement, contrary to McFarland, does not appear to be a highly stable characteristic of individual drivers.

For years the Travelers Insurance Companies of Hartford, Connecticut, have been



Figure 18.5. (From "The Passing Scene." *The Travelers 1954 book of street and highway accident data.* Travelers Insurance Companies, Hartford.)

publishing booklets on auto accidents which should be distributed to every driver and pedestrian. The figures given in these booklets are important; some of them are summarized below:

1. Speeding is the greatest single cause of deaths and injuries. It accounts for 46 percent of the deaths and 43 percent of the injuries.
2. Approximately one-quarter of the injuries and one-third of the deaths to pedestrians are caused by crossing between intersections.
3. About 85 percent of the accidents occur in clear weather.

4. The figures covering accident experience of male and female drivers would seem to settle once and for all the question as to whether women are better or worse drivers than men. The weight of evidence is all in favor of the women!

Actually, no country-wide figures are available to tell us how many more male drivers there are than female drivers, how much more mileage the average male covers in a year than does the average female, or the relative traffic hazards which each sex faces.

Studies have indicated that on a typical day men account for about eight times as many "car miles" as women, that men drive faster on the average (partly from necessity, as in the case of men using their cars for business purposes), that men often are forced by the nature of their business to drive when weather and road conditions are at their worst, and that men do more of their driving in congested, hazardous areas.

Until more data become available, it will be impossible to compare the safe driving abilities of men and women.

Females in the under-26-years groups are safer drivers, that is, have fewer accidents, than males of this age group. Again, it may be that they do not drive on dates and may drive less miles.

5. Most cars involved in crashes are apparently in good condition.

6. The hours between 4 and 6 P.M. are the most hazardous. People hurrying home from work as darkness falls may explain this.

7. Saturday is the worst day for auto accidents. Saturday and Sunday combined account for 39 percent of the deaths.

Recently California (1964) has undertaken an exceedingly comprehensive study of automobile accidents in that state. They found that only about 60 percent of the driving population is "conviction-free" over any three-year period, and that about 85 percent is "accident-free" over the same time span. California uses the "point" system in which drivers accumulate points for different violations. Over a 36-month period *less than half* of the male driving population failed to accumulate at least one point. Women were more successful; 70 percent of the female drivers escaped points.

SAFETY EDUCATION

Accidents can be reduced not only by actuarial studies of the accident-prone employee but also by safety education conducted by management and employee groups. Foremen's meetings can be effective in reducing accidents, as Figure 18.6 indicates.

Accident prevention in industry is the responsibility of both labor and management. The United States Department of Labor (1947) has issued a "Guide to Industrial Accident Prevention Through a Joint Labor-Management Safety Committee" which suggests an eight-point program for a safety committee:

1. Make immediate and detailed investigation of accident reports.
2. Develop accident data to indicate accident sources and injury rates.
3. Develop or revise safe practices and rules to comply with plant needs.
4. Inspect the plant to detect hazardous physical conditions or unsafe work methods.
5. Recommend changes or additions to protective equipment and devices to eliminate hazards.
6. Promote safety and first aid training for committee members and workers.
7. Participate in advertising safety and in selling the safety program to workers.
8. Conduct regular scheduled meetings.

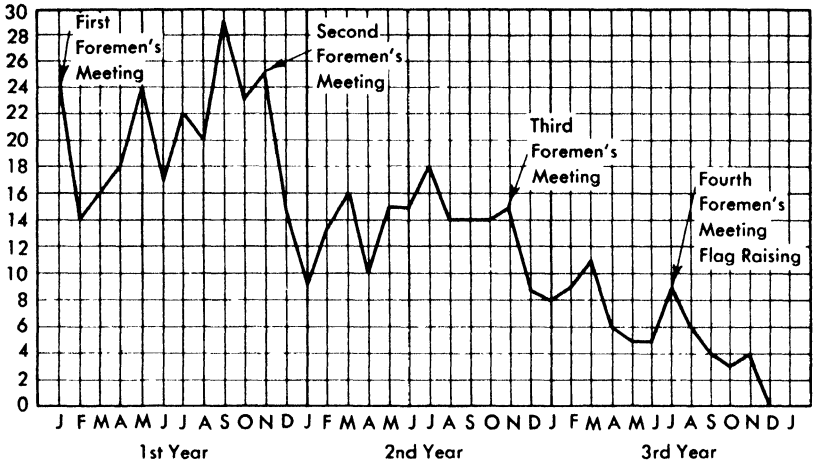


Figure 18.6. Foremen's meetings reduce accidents. (Courtesy Metropolitan Life Insurance Company.)

Valuable work is being performed by the Committee of Highway Safety Research of the National Academy of Sciences (1952). This committee serves as a center for the collection of information related to accidents and is also busily engaged in outlining desirable research programs. The following summary statements may be made from the committee's data.

1. Seventy-five to 90 percent of automobile accidents are due wholly or in part to human errors or lapses.
2. The accident records kept have value as a general index but are of questionable value as a criterion for research on accident causes.
3. Many drivers having multiple accidents in one period are accident free in a later period.
4. Factors leading to accidents are largely correctable or may be compensated for by greater caution.
5. Drivers between 18 and 25 years of age have far more than their proportionate share of accidents.
6. The effect of intensive safety education and enforcement "campaigns" has been shown to be immediate but temporary.
7. Psychophysical driver tests are probably of more value as educational than as selection devices.
8. Clinics for accident repeaters have been shown to have value. Whether this is due to increased knowledge, attitude change, or motivation is not known.

PREDICTING ACCIDENTS

Harris (1950) studied two groups of workers who were equated except for number of accidents. Using items from the Bernreuter, a personality inventory, and such projective techniques as the Rosenzweig Picture Frustration and the multiple-choice Rorschach, he found no personality differences in the two groups. This study tends to be typical of researches investigating personality characteristics and accident data.

Maybe an entirely different approach is needed. Possibly the identification of such driver types as the "overcautious," the "you-can't-do-this-to-me," and the "show-off," if detected, might lead to a reduction of the horrible toll of accidents.

Uhlaner, Goldstein, and Van Steenberg (1952) in a study of safe motor vehicle operation found 21 driving habits rated to be important. Table 18.3 may be a check list differentiating safe and accident-record drivers.

Too many people have constantly made the assumption that driving skill is related to safe driving. Actually, this may not be true at all. Such a problem is worthy of serious investigation. Many of the tests included in batteries have constantly been found to have zero correlation with safe driving. Among the useless tests should be included those for color vision, simple reaction time, and visual acuity above a certain level. Simple reaction time rarely is a factor in the prevention of an accident; compound reaction time is a more appropriate concept. Possibly the main value of a knowledge of reaction time is the educational one of knowing that a driver cannot stop on the proverbial dime.

TABLE 18.3 *Twenty-One Behavior Statements Rated "Very Important" and "I Know I Can Rate on This" by 49 Percent or More of Drivers and Supervisors*

	Percent Checking			
	"Very Important"		*"I Know I Can Rate on This"*	
Behavior Statement	A[a]	S[b]	A[a]	S[b]
Breaks the speed limit	59	66	52	60
Drives too fast for road conditions	68	89	51	69
Doesn't stay on his side of the road	73	91	54	49
Ignores stoplights or signs	77	89	61	63
Doesn't give the right of way to other drivers	65	60	58	60
Passes on curves and hills	76	83	49	51
Doesn't signal for stops or turns in advance	57	71	60	63
Doesn't check brakes before driving	71	77	51	51
Doesn't slow down at intersections when he has the right of way	52	63	57	60
Follows other vehicles too closely	59	77	60	69
Pulls away from the curb without looking back for oncoming traffic	66	80	51	51
Takes chances when driving	50	71	51	54
Gets into accidents with other vehicles	64	63	50	74
"Horses around" when he's driving	62	80	50	54
Shows off when driving	61	74	49	60
Doesn't cut wheels to curb when parking on a hill	66	80	49	49
Backs up without looking behind	72	80	53	57
Drives with dirty windshield	55	57	58	71
Swings too wide on turns	55	60	50	54
Drives faster than the other traffic	54	49	51	51
Fails to turn in his vehicle for repairs promptly	60	60	50	77

[a] Associates [b] Supervisors

SOURCE: J. E. Uhlaner, L. G. Goldstein, and N. J. Van Steenberg. Road-user characteristics: Development of criteria of safe motor-vehicle operation. *Highway Research Board Bulletin 60,* National Academy of Sciences, National Research Council, Washington, D.C., 1952, 12 (Table 8).

Attitude as related to safe driving and driving judgment appears to be more promising as an instrument to predict safe driving than most of the psychophysical measurements. In a study performed for the United States Army by Lauer (1952), a battery of tests for driver selection was found to be predictive of efficiency and included the following:

1. Attention to detail—as measured by determining the number of 0's with broken lines
2. Driver self-description blank describing the driver's background and personality-attitude
3. Driving "know-how" consisting of 48 items involving knowledge of driving beyond the wheel
4. Emergency judgment and a pictorial presentation of traffic situations with the subject indicating his solution
5. Two-hand coordination and motor coordination tests
6. Word-matching. A visual perception test in which a word at the left is matched by one of five choices at the right. The task requires vision to read the word correctly as the size is successively reduced past the threshold of readability

This approach of Lauer *et al.* is delightfully refreshing and may lead to more positive results than those in the past. The Uhlaner, Goldstein, Van Steenberg driving habits test may, however, be the most direct approach. This would require training and testing of knowledge of driving. It might effectively instill desirable safety habits necessary for safe driving. This approach may prove to be more important than an attempt to isolate the psychophysical skills involved.

Flanagan (1953) suggests the study of near-accidents as a means of reducing accidents. A study conducted for the Air Force found that the main causes were: personnel acts, 61 percent; mechanical malfunction, 34 percent; design and procedure, 32 percent. (Some accidents had more than one cause factor.) The clear-cut implication is that human behavior is the greatest single cause of near-accidents, and while it is true that in this instance human behavior also prevented the accident, nevertheless every accident, at the moment prior to the occurrence, was the near-accident. The study of a near-accident can lead to the prevention of accidents if the causes of near-accidents can be more clearly determined. Such a view would apply to the industrial scene as well as to the automobile accident.

Goldstein (1962) has reviewed the research literature relating to human variables in highway accidents. He classes the predictor variable into twelve different categories:

1. Driving behavior variables (prior accidents, violations, etc.)
2. Visual ability measures
3. Reaction time measures
4. Psychomotor measures (apparatus)
5. Psychomotor measures (paper and pencil)
6. Sensory-perceptual measures
7. Cognitive measures
8. Personal, emotional, and attitude measures
9. Background and sociological measures
10. Fatigue
11. Age
12. Alcohol

After examining the studies in each category to see how well they were able to predict accidents, Goldstein comes to the following conclusions:

1. Accident records do not measure a very stable human performance characteristic; accident status in one period is only slightly related to accident status in another.
2. Accident records seem to be only slightly predictable from measures of other, stable, human characteristics such as visual acuity, reaction time, and sensory, psychomotor, cognitive, and attitudinal measures.
3. Accident repeaters apparently do not account for much of the total of traffic accidents on record.
4. Two age groups contribute disproportionately to the accident total: (1) those below 25 and (2) those above 65.
5. Alcohol has a bad effect on driving performance at much lower blood level concentrations than is generally recognized; namely, 0.05 percent or even lower.
6. Alcohol is a contributing factor in 25 to 50 percent of fatal traffic accidents, on the part of drivers and/or pedestrians.
7. Drivers with extremely poor attitudes of aggressiveness, social irresponsibility, and/or who are highly unstable, apparently have more accidents than those who are responsible, stable, and less aggressive than average.
8. Students who elect driver education courses apparently have different attitudes and personality characteristics as a group from those who do not choose such courses.

INSURANCE COMPANY TESTING PROGRAMS

In spite of the rather dismal view presented by Goldstein (1962), there is a growing tendency among major American insurance companies to turn to psychologists for aid in deciding (1) whom to insure and/or (2) in what risk category to place a new client. Automobile insurance has *not* been a particularly profitable business in recent years. With the ever increasing cost of the average claim, a great pressure has been put on insurance companies to raise rates. However, this is an unpopular practice and bad for the company image. Also, in many states it is a lengthy process to get a rate increase approved by the state insurance commissioner. Therefore, the possibility that psychological assessment procedures might be able to identify high-risk drivers at the time of application is a highly appealing notion.

Presently, the Kemper Insurance Company and Farmers Mutual Reinsurance Company of Iowa are both using tests to determine in what rate category a new applicant will be placed. Haner (1963) has reported some of his research for the latter company and the procedure seems to be successful. Using an attitude inventory, Haner classed applicants into four risk groups: (I) adjudged good risks, (II) intermediate risks, (III) poor risks, and (IV) very poor risks. He then examined later driving performance to determine how successful his categorization scheme was in predicting who would have claims against the company. Some of his results are shown in Tables 18.4 and 18.5. The results clearly support the use of the attitude measure. Because of this type of success other companies are now beginning to try psychological tests to help reduce their loss ratio. (State Farm Insurance Company, for example, currently has a five-year research program along these lines.)

FATIGUE

Fatigue is a subject that interests both the employer and the employee. Since a decline in the curve of production is often attributed to the worker's fatigue, the

TABLE 18.4 *Risk Categories as Adjudged by the Inventory Related to Seriousness of Injuries Incurred in Closed Primary Negligence Claims*

	Serious Injuries	*Moderately Serious Injuries*
Group	*Claims*	*Claims*
I	0	18
II	6	46
III	14	56
	20	120

SOURCE: C. F. Haner. Use of psychological inventory in writing insurance for youthful male drivers. *Traffic Safety*, 1963, 7, 5–14.

TABLE 18.5 *Risk Categories as Adjudged by the Settlements of Claims*

Group	*Mean Settlement*	*Settlement Per Car Year Ratio*
I	$331.50	100
II	306.66	124
III	488.67	263
I	236.23	100
II	544.63	254
III	530.54	336
IV	620.80	365

SOURCE: C. F. Haner. Use of psychological inventory in writing insurance for youthful male drivers. *Traffic Safety*, 1963, 7, 5–14.

employer would like to eliminate this decline by eliminating fatigue. The employee recognizes fatigue as a feeling of tiredness or pain; he regards it as intrinsically unpleasant, and so he too is interested in its decrease or elimination. For these reasons, industrial fatigue, despite its elusive nature, has been the subject of many investigations, from both a physiological and a psychological point of view. Many topics in industrial psychology are directly or indirectly concerned with fatigue. One of the goals of time and motion study is to reduce the effects of fatigue in workers by introducing more efficient work methods. Evidence indicates that proper training results in correct work methods which can do much to prevent the onset of fatigue. It is also recognized that effective motivation can be successful in eliminating fatigue or at least in reducing its effects. A relationship has been found between accident rate and fatigue.

The word "fatigue" is widely used and misused. Practically all physiologists and psychologists recognize the diverse factors involved in this concept. Sometimes fatigue of a physical nature is differentiated from fatigue of a psychological nature.

This leads to the use of the term "fatigue" as differentiated from the term "mental fatigue." Monotony, boredom, and other similar phenomena in industry are in some respects similar to fatigue, at least as far as reduced production is concerned. Although such terms are sufficiently different so that anyone knows in personal situations when he is fatigued or tired as compared with when he is bored, in industrial situations there is often confusion between these two concepts. The industrial psychologist studies fatigue not so much to understand its physiological basis as to be able to climinate as many of its effects as possible, and thereby maintain or increase production and job satisfaction.

WHAT IS FATIGUE?

Investigators working in this area are in general disagreement as to the nature of fatigue, and many are perplexed by the diverse character of the concept. According to Dill (1933):

> We can say that fatigue is not an entity but merely a convenient word to describe a variety of phenomena. The common fallacy of supposing that the word fatigue corresponds to a definite thing has been a source of much confusion. Fatigue from short bursts of activity, whether by the whole body or by isolated muscular groups, is characterized by increase in lactic acid and temporary inability to continue. Fatigue from depletion of fuel reserves does not occur commonly in man, but when it does, chemical analysis of the blood reveals a low level of blood sugar. Fatigue from working in a hot environment has several manifestations, the most simple to measure being the increase in heart rate. Finally in the instance of two individuals doing the same task, one may become more fatigued than the other because the poor nervous coordination of the unskillful man makes it necessary for him to expend more energy than the other. In general, fatigue from any of these causes is greater the more nearly the individual approaches his capacity for work.

In an intensive study called *Fatigue and Hours of Service of Interstate Truck Drivers* (1947), the term "fatigue" is defined as an altered psychological and physiological state in relation to the status of recovery or normal capacity. Muscio (1921) defines fatigue as "a condition caused by activity in which output produced by that activity tends to be relatively poor and the degree of fatigue tends to vary directly with the poorness of output." Muscio, however, is himself dissatisfied with this definition because fatigue cannot be directly measured and because diminished output may also be a result of distraction. He concludes that the term "fatigue" should be banished from scientific terminology. But banning the word would still leave the phenomenon that by any other name would still be fatigue. It would be pointless to argue with a fatigued worker that there is no such thing as fatigue because fatigue does not stand the rigors of scientific tests and standards.

Whatever fatigue is, it is safe to say that any muscular work, even that involved in sitting in a chair, will result in fatigue provided the work of the muscles and the resulting expenditure of energy are at a faster rate than is recovery. The major difficulty in understanding the true nature of fatigue is caused by the fact that many factors in addition to muscle use contribute to its onset. The length of the work period, the speed of work, the extent of the musculature involved, and the tensions accompanying the task are a few of these factors. Further complications are caused by the fact that industrial fatigue rarely concerns a single muscle but rather the individual as a whole. Depending on the strength, stamina, and preparation of the individual, fatigue is manifested at different rates in different people.

In some respects no one can ever hope to eliminate fatigue from work unless the work can proceed at the exact rate at which recovery takes place and all the other factors contributing to fatigue can be eliminated. And yet some people set their pace on the job so that recovery seems to take place faster than the onset of fatigue. Half seriously, this might explain the stenographer who comes to work in the morning very fatigued but somehow manages to gain energy during the day so that she is able to keep her date in the evening, only to be fatigued the next morning. This must not be regarded as fatigue; it is a phenomenon related to fatigue that will be discussed later in this chapter.

Psychologists can help to eliminate much of the unnecessary fatigue that is caused by unfavorable conditions on the job. Inefficient work spaces, inefficient work methods, inefficient tools, and inefficient people to do the work cause much unnecessary fatigue. These causes can be attacked and conditions changed so as to eliminate or reduce much fatigue on the job.

A simple approach to an understanding of fatigue has been the study of it in the laboratory. An early study of this nature was conducted by Mosso (1915) when he constructed an ergograph to record and measure the work done by the muscles in flexing a finger. By attaching a weight to a string tied to a finger and strapping the hand so that only this finger was allowed to move, it was possible to obtain a sample work or fatigue curve. Using this technique, Mosso demonstrated the rate and extent of the work that could be done by a single finger lifting a specific weight. He was also able to demonstrate that if a fatigued muscle is forced to work, the period of recovery is prolonged. Thus if a few minutes is required for recovery after a 10-minute work period but the work period is nevertheless increased to 15 minutes, recovery requires more than one and a half times the time required after the 10-minute work period. Although Mosso's ergograph does study fatigue, it has the obvious disadvantage of doing so under extremely unnatural conditions; furthermore, no job in industry is likely to require the movement of only one finger.

Vernon (1924) developed a variation of the dynamometer which enables one to study the effects of work when the larger muscles of the body are involved. His apparatus required the subject to pull with both hands on a handle attached to a spring balance; the pull on the balance was recorded by a pointer on a revolving drum. With this apparatus Vernon was able to show that the strength of the pull varies with the height of the handle. He also demonstrated that the introduction of rest pauses increases the capacity for work, and further, that a change of posture during the rest pauses helps recovery to a greater extent than merely stopping work. Vernon's study has a more direct bearing on various industrial jobs and is a strong argument in favor of introducing rest pauses during the work period. It also recognizes the benefits to be derived from encouraging employees to change their posture from that normal during the work period. Many other studies of a similar nature have been made in the laboratory, but each has the disadvantage that the subject is not actually working on a job. Inferences drawn from the laboratory studies may not apply to industry because differences in work speed, pace, and continuity may exist.

INDUSTRIAL STUDIES OF FATIGUE

Crowden (1932) classifies muscular work in industry into three general types. The first is heavy muscular work that is too strenuous for a steady, continuous rate of work to be maintained. Examples of this type of work are loading trucks, building

roads, and possibly mining coal. The second type consists of moderately heavy work that is continuous and in which the rate of expenditure of effort is much lower than in the first type and is somewhat balanced by the rate of recovery. Examples of this type of work include machine tending and many other kinds of factory work. The third type of muscular work in industry is the light, speed work which involves a relatively small expenditure of energy but which often requires a postural strain that causes unnecessary fatigue. Office work is a typical example.

Crowden found that in a fifty-yard barrow run the worker expends approximately 8 percent of his energy in raising and lowering the handles, 22 percent in attaining a wheeling speed and in stopping, and the remaining 70 percent in the run itself. This study shows the tremendous inefficiencies that would result from interfering with the run once it is started. In studying the energy cost of moderately heavy work, Bedale (1924) found that carrying a load with a yoke as a milkmaid did is the most economical method from the point of view of expenditure of bodily energy. Any method of carrying weights which requires postural strain and displacement of the body when walking is more costly. Crowden reports that light speed work involves little expenditure of energy, but that there may be considerable fatigue because of the cramped or uncomfortable position maintained by the worker. Measurement of the actual expenditure of energy does not give a measure of fatigue with Crowden's method because his measures basically the oxygen consumption in excess of normal.

The tremendous difficulty of the problem of fatigue is vividly illustrated by the research done by the United States Public Health Service on the relationship between fatigue and hours of work of interstate truck drivers. This is one of the most valuable studies in this field. The purposes of this study were (1) to determine if various periods of truck driving would produce demonstrable and significant psychophysiological changes; (2) to investigate the nature of these changes; and (3) to discover whether a characteristic pattern of psychophysiological response occurs after long hours of driving, i.e., the syndrome "drivers' fatigue."

The sixteen factors which were considered as contributing to fatigue in truck drivers were:

1. Performance of a skilled operation requiring a high degree of alertness and attention
2. Nervous strain due to driving under adverse conditions
3. Muscular exertion in loading and unloading and in the repair and maintenance of vehicles
4. General irregularity of habits as a result of long-distance hauls
5. Failure to obtain satisfactory rest or sleep during rest periods or when off duty
6. Physical condition
7. Constant use of the eyes, frequently under unfavorable conditions such as glare, etc.
8. Social factors in the environment or occupational tradition possibly promoted by enforced absence from home
9. Monotony inducing sleepiness
10. The consumption of coffee and alcohol
11. Exposure to all types of weather conditions
12. Exposure to toxic fumes and gases
13. Economic insecurity, that is, fear of losing one's job, especially in the case of older men
14. Noises
15. Vibration
16. Sedentary occupation—the effect of posture

In this study a total of 889 drivers in three cities—Baltimore, Nashville, and Chicago—were given a comprehensive battery of tests. The investigators took no

sides in the controversy as to which type of test best measured fatigue, but used both simple and complex performance tests and also included nonperformance tests in their battery.

Whereas performance tests measure ability to do a given task, nonperformance tests measure bodily states over which the subject has little or no voluntary control. An impartial observer will readily concede that each of the two types of test used in the measurement of fatigue has its advantages and disadvantages and that both should be used. However, they have been the subject of considerable debate among those attempting to measure fatigue. One of the advantages of a performance test is that it directly measures a function in relation to the specific task—hand steadiness, for example. Another is that it can detect relatively slight degrees of fatigue more rapidly than most nonperformance tests. The disadvantage of performance tests is that they can be influenced by the motivation and attitude of the subject.

The advantage of using a nonperformance test is that it is based entirely on chemical and physical changes which, in most cases, cannot be brought about deliberately by the subject. The disadvantage is that emotional states which have nothing to do with fatigue may, in certain instances, cause similar chemical and physical changes and so the test may measure not fatigue but an undetermined emotional state.

The complete battery included the following:

1. Psychological tests
   a. Spatial perception (the estimation of known sizes)
   b. Manual steadiness
   c. Precision of movement (aiming)
   d. Reaction-coordination time
   e. Reaction time
   f. Speed of tapping and work decrement
   g. Strength of grip
   h. Static equilibrium (postural steadiness)
2. A series of tests using the De Silva driver vigilance test apparatus, consisting of accelerator-brake-foot reaction time, steering efficiency, and a combination of starting efficiency plus brake reaction
3. Tests to measure resistance to and recovery from glare
4. Tests to measure the speed of eye movement (see Figure 18.7)
5. Tests to determine the critical fusion frequency following exposure of the eye to flicker at two levels of illumination
6. Snap acuity test
7. White blood cell count
8. Concentration of potassium and total base in blood serum
9. Determination of carbon monoxide content of blood

In addition to these tests, a thorough medical examination was given. Also included in the data was an occupational and habit study of the drivers.

The major findings of this study do not present the clear-cut conclusive evidence that one would like; but there is no real reason to believe that such information can be forthcoming with the measures of fatigue that are known at the present time. The investigators state: "It appears that a reasonable limitation of hours of service of interstate truck drivers would reduce the number of drivers on the road with low functional efficiency. This, it might reasonably be inferred, would act in the interest of highway safety." The test results showed that the men who had not been driving just prior to being tested had the highest average efficiency, those who had driven

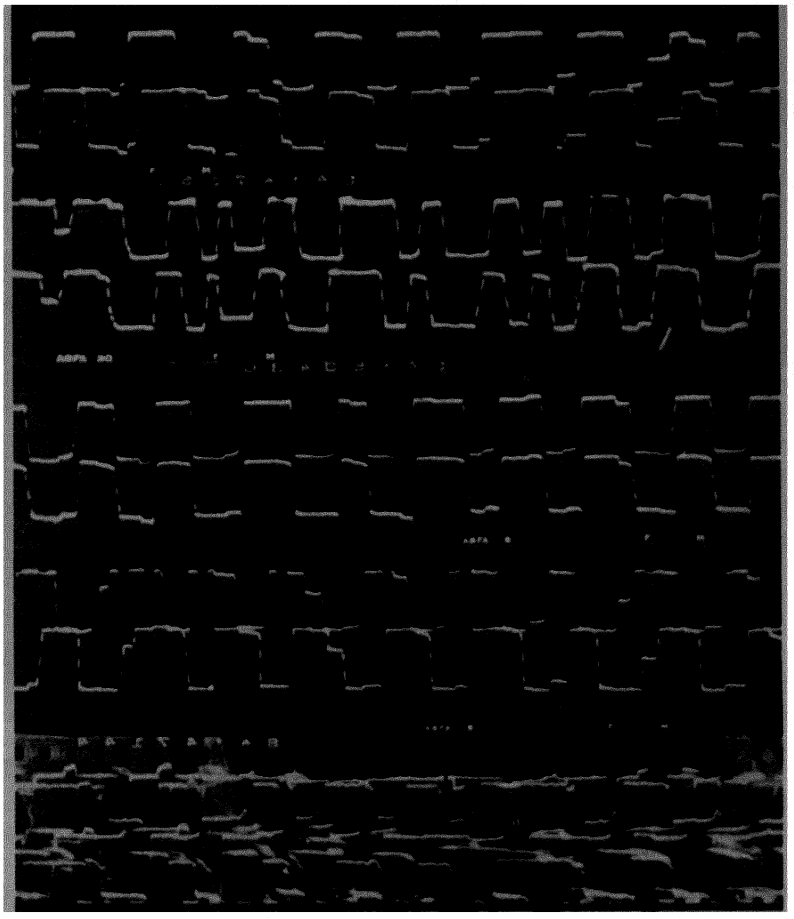


Figure 18.7. Ophthalmograms of truck drivers. The dotted lines are saccadic movements and the solid lines are fixations. (From *Fatigue and hours of service of interstate truck drivers.* Public Health Bulletin No. 265, Federal Security Agency, U.S. Public Health Service, Government Printing Office, Washington, D.C., 1941.)

less than ten hours had the next highest average efficiency, and those who had driven over ten hours had the lowest average efficiency, in the following seven functions:

1. Speed of tapping
2. Reaction-coordination time
3. Simple reaction time
4. Manual steadiness
5. Body sway
6. Driving vigilance
7. Ability to distinguish flicker

The men who had been driving prior to being tested performed less efficiently on the average than those who had not driven, on tests of the following three functions:

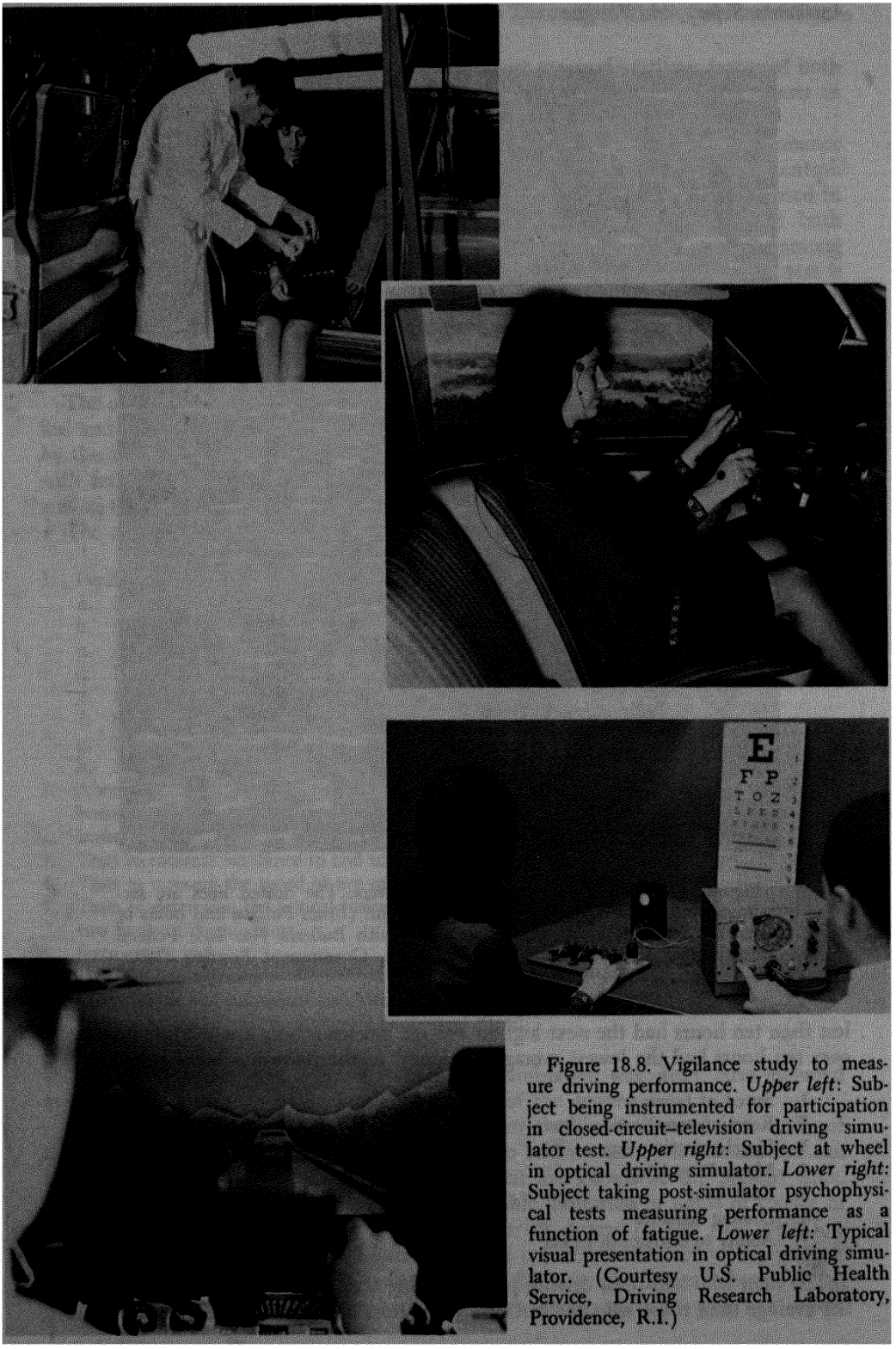


Figure 18.8. Vigilance study to measure driving performance. *Upper left*: Subject being instrumented for participation in closed-circuit–television driving simulator test. *Upper right*: Subject at wheel in optical driving simulator. *Lower right*: Subject taking post-simulator psychophysical tests measuring performance as a function of fatigue. *Lower left*: Typical visual presentation in optical driving simulator. (Courtesy U.S. Public Health Service, Driving Research Laboratory, Providence, R.I.)

(1) aiming, (2) resistance to glare, (3) speed of eye movement. The data on these three tests, however, do not differentiate consistently between drivers who worked from one to ten hours and those who worked over ten hours. In items 1–7 this difference does occur. Heart rate decreased slightly with hours of driving. The average white cell count was higher in men who had driven than in men who had not driven since sleeping. No relationships were found between hours of driving and ability to estimate the size of known objects, differential white cell counts, hemoglobin content of the blood, acidity of urine, specific gravity of urine, visual acuity, and the total base and potassium concentration of blood serum.

There was a relationship between the men's subjective estimate of fatigue and the objective measure provided by some of the tests.

Driver reactions under prolonged loss of sleep are reported by McFarland and are discussed in Chapter 19.

Ryan (1947) in studying the concepts of work and efficiency rightfully prefers to consider the relationships between cost and energy of input and output. In his book he admits, "These chapters on the fundamental problems of efficiency [as he sees them] have raised more questions than they have answered." For practical purposes it does appear as if the concept of fatigue, though interesting from a physiological and laboratory point of view, leads to little progress in better understanding the involvements in the day-to-day work situation of a typical man.

RELATED PHENOMENA

If fatigue is regarded as elusive and difficult to measure, mental fatigue must be regarded as much more so. After preparing a lengthy assignment, college students often insist that they cannot possibly read another page; they are exhausted and must go to bed. If, at that particular moment, the telephone rings and an attractive date is in the offing, the fatigue caused by this "strenuous" mental activity goes out the door with the student. The question is: Was there any mental fatigue in the first place?

Mrs. Gilbreth, an applied psychologist, once said in a speech that girls with dates suffer less from fatigue than girls who have no dates. It is not necessary to wax philosophical in a book on industrial psychology, but mental fatigue implies that such "stuff" exists, apart and distinct from the physiological realm. This, according to present-day science, cannot be taken too seriously. Undoubtedly, the major component of mental fatigue is attitudinal; the physiological component is surely not sufficient to resemble the fatigue measured by available performance or non-performance tests.

In an experiment on mental fatigue (Huxtable, White, and McCarton, 1946), three subjects worked for a 12-hour period on four successive days multiplying four-place numbers by four-place numbers. Each problem was solved without the aid of pencil and paper and only the answer was recorded. If anything could cause mental fatigue and even a certain amount of physical fatigue, an experiment such as this should have produced it. However, little evidence to support this was uncovered. Even though the subjects had had extensive practice prior to beginning the experiment, all three showed an increase of speed in mentally multiplying four-place numbers. It was found that the effect of continued mental work was associated with an increase in the percentage of error. One subject scored 38 correct answers out of 80 on the first day, and 33 correct out of 115 on the fourth day. The other two gave

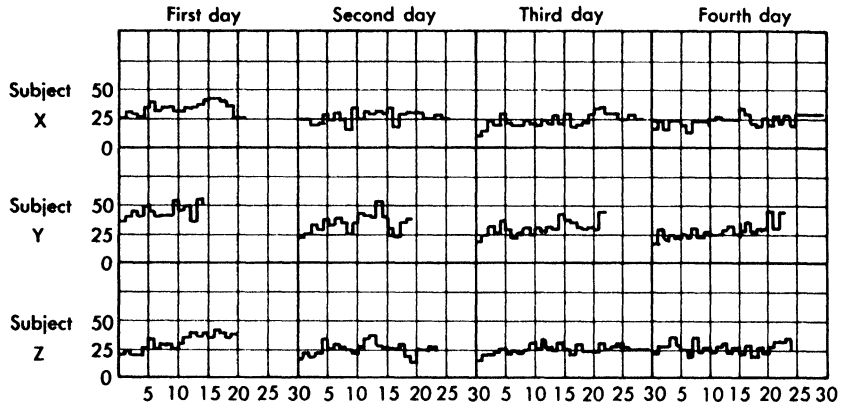


Figure 18.9. Work curves of subjects performing a difficult mental task. (From *Psychological Monographs*, 1946, 59(5), 14. By permission of the American Psychological Association.)

similar results. The work curves for all three subjects (Figure 18.9) are relatively flat and do not show the theoretical characteristics of either fatigue or monotony curves.

"Despite conclusive, objective evidence that feelings of intense dislike, ennui, boredom to the extent of headache and illness, dizziness, nervous instability and physical and muscular weariness were accompanied by some decrease in mental efficiency, one would not venture a guess as to what extent the decrease was attributable to divided attention or to mental fatigue per se."

In addition to this rigorous mental task, the subjects took a battery of psychological tests before and after the 12-hour multiplication period. The results of these tests were not consistent. For example, at the close of the first day all three subjects made lower scores after the 12-hour period of mental work; this suggests the possibility of "transferred fatigue." But on the second day all three made slightly higher scores after the ordeal, and on the third day the results were erratic and contradictory. With reference to the nonperformance tests, all three subjects displayed few conclusive signs of unusual physical fatigue as revealed in measures of abdominal and costal respiratory movements, metabolic rate, pulse, temperature, weight, recording of blood content, and urine analysis. The authors concluded: "Despite subjective states of physical fatigue and overwrought nerves recognized by all three subjects, the records showed little indication of physical fatigue except for the rise of metabolic rates during the third and fourth days of the fatigue series. Even these metabolic rates had returned to normal on the morning of the day following the experiment." The notes of one of the subjects best summarize the subjective results. She wrote, "Isolation, hard work, etc., made entire four days seem like one long nightmare to me. I marvelled that the other two girls seemed to hold up so well. Would not repeat these four days for ten thousand dollars, I believe."

This experiment and its findings must create considerable doubt as to the existence of mental fatigue as an entity. As a feeling, it undoubtedly exists. However, no physiological changes took place and measures of performance did not show any decrement.

However, Geldreich, conducting an experiment in mental tasks, obtained rather

different results (1953). He subjected ten persons to the simple task of manual naming one of five colors—red, yellow, green, blue, and white. The actual task lasted 55 minutes and was repeated over a number of days. Experimental design was planned to include all necessary controls. Geldreich found that the production in the last five minutes was 14 percent less than in the first five minutes. He also found increases in heart beat, respiration rate, blood pressure, and skin conductance during the performance of this task.

This study tends to support the view that simple mental tasks result in not only a work decrement but also physiological changes; in general this disagrees with the findings reported in the previous study. Thus the last word and final answer to the existence and effects of mental fatigue remain to be offered.

Possibly the largest component of mental fatigue is the tension and attitude usually accompanying a task that an individual regards as a challenge. Writing an examination paper for three hours, were no tensions involved, would be no more fatiguing than writing a letter to a friend. Tension is not only a psychological but also a physical entity. A person can continue extremely difficult mental tasks without any dire effects much longer than is ordinarily believed. The Huxtable experiment clearly illustrates this point, but the Geldreich experiment on a much simpler task definitely did show work decrements. The clue to the differences in results may be that for college students color naming is not as challenging as mental multiplication. Motivation and boredom are sometimes overlooked as variables in experimentation. Results from many different sources indicate that physiological differences may be a result of motivation or emotion as well as of physical tasks.

A study on the subjective feelings of tiredness during each half-hour of an eight-hour workday (Griffith, *et al.*, 1950) for manual workers ($N = 232$), office workers ($N = 73$), and supervisors ($N = 75$) reveals considerable similarity in the morning periods and some similarity in the afternoon periods. The peculiarity of maximum "tiredness" prior to the lunch hour raises the question that the feeling of tiredness may not be an indication of fatigue at all. One should expect more fatigue, if it exists, in the afternoon or at least at the day's end. Figure 18.10 presents the curves as found by Griffith, *et al.*

Another phenomenon related to fatigue is monotony or boredom. The outstanding characteristic of monotony is that it is dependent upon the *individual* rather than the job. The girls assembling relays in the Hawthorne Studies showed little evidence of monotony despite the extremely repetitive nature of their job. A story is told about a champion orange packer—in other versions it is an electric bulb packer or someone else. (Chances are it never happened, but it illustrates the point.) In a certain state a contest was held to determine the fastest orange packer. He was to be designated "king" or "champion" with much fanfare. After a series of preliminaries, quarterfinals, and semifinals, the final was held and the champion selected. This man worked at a fast, steady, almost furious pace. But of course he finally stopped work long enough to be interviewed, and he supposedly reported that he found the job very challenging. For one thing, he did not regard oranges as all alike; they differed in size, color, and texture. Furthermore, when he saw an empty crate in front of him, he had a strong desire to fill the emptiness by putting in the first layer as rapidly as possible. When this layer was finished, the motivation was strong to complete the second layer and finally the last, so that the crate could be taken away and a new one brought to him. To this person, if he ever existed, orange packing was not a monotonous job.

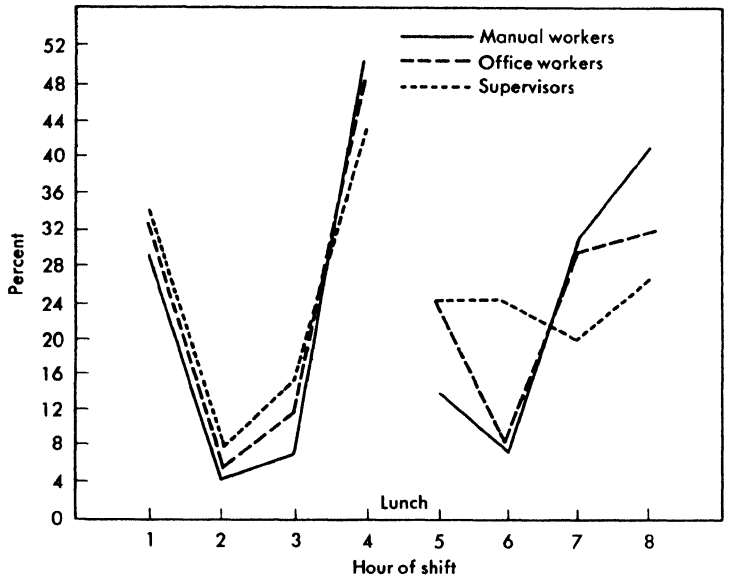


Figure 18.10. Percent of manual, office, and supervisory employees reporting maximal feelings of tiredness at each hour of each half of the eight-hour work shift. (From J. W. Griffith, *et al.* Changes in subjective fatigue and readiness for work during the eight-hour shift. *Journal of Applied Psychology*, 1950, 34, 163–166.)

Much has been written about the effect of specialization and simplification of work. People are supposed to prefer varied rather than uniform or repetitive tasks. This is not true to the extent that is ordinarily believed. The average individual gives lip service to the importance of and need for variation in work and life in general, but he merrily performs as many tasks as possible in a routine fashion. For example, people who ride on the subway in New York City have little need to travel on the same train each morning, since trains are likely to run just a few minutes apart. Yet many people who claim they desire variability in their work walk to the station by exactly the same route and enter the same door of the same car, day in and day out. People are likely to eat in the same restaurant every day; many prefer to sit at the same table each time. Examples of the desire and preference for uniformity of tasks are endless. Many people prefer a job with a minimum of responsibility. Jobs that are varied sometimes require decisions that may get the person into trouble, whereas routine jobs are "safe" jobs. For some people, a uniform task is a boon and not a boomerang. The only job that is monotonous is the one which the worker who does it regards as monotonous, and this is true regardless of the occupational level.

As we saw earlier, fatigue, assuming that it does exist in industry, can be reduced by shortening the working day, by introducing rest pauses, and by providing more efficient environmental factors on the job. Monotony can be reduced by a more careful selection of personnel—by attempting to hire the person with the necessary intelligence for the job and rejecting those with too much or too little. Consideration of the job in relation to the personality of the individual often makes for a situation in which there is no clash between the individual and his tasks.

Just as rest pauses reduce fatigue, so are they useful in reducing monotony. A rest pause often provides a necessary break in activity and just the change the worker needs to combat monotony.

From the point of view of reducing monotony, job changing is important. Employees very often change jobs, even against company rules, in order to combat monotony. Good supervision demands that such changes be noted, studied, and possibly encouraged. In many cases the rule against changing jobs is the result of an incomplete definition of efficiency in the first place. Many jobs can be made more interesting provided meaning is attached to the work. By this we do not mean assuming a "Pollyanna" attitude and attempting to create meaning where no meaning exists; workers are too smart for these tactics. However, explaining to the worker what his task is in relation to the organization as a whole often gives him a sense of the meaning of his job that he could not have gained by himself.

Another way to combat monotony is to provide social and recreational activities. Some organizations carry this to such an extreme that a man's job sometimes seems merely a fill-in between one bowling contest and the next. However, there is no doubt that many workers look forward to the few minutes they can spend in the canteen smoking, talking, and drinking soda.

HOURS OF WORK AND PRODUCTION

The crux of the problem in industry is not really the presence of fatigue or monotony but the relationship between an employee's hours of work and his production. There is widespread misunderstanding and confusion about this relationship. It is generally believed that the way to increase production is to increase the number of working hours. During periods of unemployment it is widely believed that the way to spread employment is to shorten the workweek. Both these views are "common sense"—and faulty. Stated most simply, increasing the hours of work usually decreases production. The corollary is also true: Decreasing the hours of work usually increases production.

As we all know, a tremendous increase in production occurred during the war years. This developed in spite of the increased hours of work, rather than because of them. The real problem is not the number of hours of work in relation to production, but the relationship between the real and the nominal hours worked. Bold recognition must be given to the fact that in any workweek a difference exists between the actual hours of work and the nominal hours of work. Further, studying the relation between actual and nominal hours in a workweek uncovers that each does not increase or decrease in a constant time manner.

Many prominent labor-management leaders, who are quite important and generally well informed, indicate their lack of awareness of this phenomenon. For example, in 1950 C. E. Wilson (1950), formerly president of General Electric, called for a temporary 45-hour week to boost defense production without cutting consumer goods. In addition, Mr. Wilson saw this technique as a way to avoid inflation. As might be expected, CIO and AFL leaders rejected the proposal that overtime pay start at 45 hours rather than 40 hours.

It does appear as if the workweek length is a "football." A more calm consideration reveals that a certain number of actual hours are worked in a nominal week, and

that the optimal relation establishes the highest number of actual work hours in relation to the total or nominal week.

The length of the working day has become an emotional problem. Unions have used a shorter workday as an indirect means of increasing wages. Some businessmen have expressed great concern over what the employee will do with this new leisure time; others believe that a shorter workday will force them out of business. Both groups tend to overlook the important point—how many hours a person actually works during the workday. Very often, a decrease in the nominal hours of work does not in any way affect the actual hours worked. Nominal hours are defined as those between punching "in" and "out" on the time clock. There is a difference between these hours and the actual hours worked; any employer or employee knows this. Unproductive working time, rest pauses, tardiness, early stoppages, absenteeism, and changes in work pace must all be considered. It is these factors, as well as others, which create the discrepancy between the nominal and the actual hours of work.

It is a fact that as nominal hours increase the proportion of actual hours worked to nominal hours decreases. Similarly, as nominal hours decrease the proportion of actual hours to nominal hours increases. This principle must be remembered, for not until it receives the proper attention will much erroneous thinking be prevented. Knight (1939) reports a number of typical illustrations of this principle. In one instance when nominal hours were reduced from 63¼ to 54, actual hours fell only from 56 to 51. In another case, when nominal hours were reduced from 62.8 to 56.5, actual hours *rose* from 50.5 to 51.2. In a third case, the amount of time lost through sickness was 2.8 percent of a 46-hour week. This rose to 3.85 percent when the nominal hours were increased to 54, but dropped to 2.7 percent when the hours were reduced to 46. In still another study reported by Knight, the nominal hours were reduced from 74½ to 63½ and the actual hours fell from 66 to 54.4. However, since the hourly output increased by 21 percent, the total output remained unchanged. The reduction of nominal hours continued, and finally there was an increase of 13 percent in weekly output, although the actual hours had been reduced 18½ hours.

In one other study, 2⅔ hours overtime was added to the then normal 10-hour day. The hourly output fell by 6.5 percent on the day overtime was worked and 3.9 percent on the following day.

A survey made by the U.S. Department of Labor (1944) in 12 metal-working plants found that the 40-hour week and the 8-hour day yield the highest output for each hour worked. More than 40 or 48 hours a week resulted in additional output, but with constantly decreasing efficiency and with increasing absenteeism as the hours were stepped up.

Employees doing light work under wage incentive systems and with weekly schedules ranging between 55 and 58 hours have achieved the equivalent of approximately two hours' output for every three hours worked over 48. On heavy work, the ratio was more nearly one hour's output for every additional two hours' work. This survey also gathered data indicating that the 7-day week as a steady program is uneconomical and may actually result in lower production than the 6-day week.

The primary effect of lengthening the workday for employees working a 5-day week is to wipe out the midweek spurt in production. Analysis of daily production records in several plants on a 40- to 48-hour schedule shows a buildup of hourly efficiencies toward a peak on the third or fourth day of the week, and a slight drop

thereafter. When the workday was lengthened to 9½ hours or more, this peak disappeared. The data indicate clearly that workers adjust themselves to a longer workday by slowing down.

This study also found that when a sixth workday was added to bring the work week up to 58 or 60 hours, the result was likely to be a steady decline in the efficiency level every day, with the peak points occurring at the beginning of the week (Monday or Tuesday).

In a further study reported by the U.S. Department of Labor (1947) it was found that, all else being equal, the 8-hour day and 40-hour week are best in terms of efficiency and absenteeism and that more hours are less satisfactory. Since this was a study of 78 cases covering 2445 men and 1060 women in 34 plants, it must be inferred that the investigators did not have ample opportunity to study shorter workweeks. They also found that longer hours yielded higher output but at increased unit costs. Possibly in a wartime economy when production at all costs is the theme such output can be justified. In an efficient and normal economy it is difficult to justify such a position. Accompanying longer workweeks were increases in absenteeism and injuries.

The 6-hour workday unbroken by a lunch hour but with a slight interval for refreshments is not an industrial impossibility. A study reported by the U.S. Department of Labor (1933) describes the experiences of a factory which changed from three 8-hour shifts to four 6-hour shifts. This led to a reduction in earnings for most employees. Four hundred and twenty women were interviewed. Of this number, 265 had worked under both systems and 77 percent of this group preferred the 6-hour shift because it gave them "more time for the home," "more leisure," and "less fatigue."

During the war, when many people were concerned with the optimal workweek from the point of view of maximum production, the Industrial Relations Section of Princeton University canvassed a group of key industrial relations executives in representative companies throughout the country. The prevailing judgment of these executives (1942) was that the 8-hour day and 48-hour week constituted the best schedule for sustained production in our war industries. This study gathered striking evidence which indicated that a workweek of more than 48 hours, and particularly over 54, resulted in a reduction in individual output and an increase in the number of days missed from work. The effects of the long hours piled up slowly and did not become evident in the first few weeks under the extended schedules. The survey also found that an increase in the number of days worked per week from five to six seemed to have less effect on productivity than an increase in hours per day.

The increase in absences with the 8-hour day and 6-day week is probably due more to a desire for leisure or recreation than to the accumulation of physical fatigue. Women in particular want time off on a weekday for shopping and household duties; they probably achieve their best performance in a 5-day week, even if it entails longer daily hours.

Some years ago, when the employees of a large department store in New York City returned to work after a strike, they voted 822 to 97 for shorter hours with the same pay. The settlement plan they voted for called for a 5-day 40-hour week; they turned down a 5½-day 42-hour week with a weekly wage rise of $1.50. In other words, employees who do not earn too much money in the first place prefer a 5-day week to a 5½-day week.

The question of nominal hours in relation to actual hours of work should not

be an emotional issue but rather a subject calling for careful study of the available facts. Most of the published material in the field indicates that shortening the workweek does not necessarily interfere with total production; in fact it may serve to increase total production.

Furthermore, this material indicates that overtime results in inefficiency. With the general practice of paying for overtime at 1½ to 2 times the normal rate, the cost of production must mount even higher when it is recognized that production during the overtime period is less than it is during the normal workday. Moreover, overtime has been found to lower production both the day the overtime occurs and the following day. A worker consciously or unconsciously paces himself when he knows he must put in extra hours that day. Telling employees, without warning, that they are expected to work overtime creates dissatisfaction. On the other hand, if the announcement is made a day in advance, there is usually an increase in absenteeism. No matter how one looks at it, overtime means an increase in the hourly working week and often serves little purpose, although the employee may not object to it because it is an opportunity to earn more money. In a study conducted over a three-year period, Scheidt (1937) reported that a 6-hour day raised employee efficiency from 15 to 30 percent above what it was for an 8-hour day that included two hours of overtime. The employees were 17 women who set type by machine and 52 men who set it by hand. Overtime is not recommended as a means of increasing production, unless it is infrequent and on a volunteer basis.

THE ULTIMATE WORKWEEK

In view of the current trend in the workweek, it appears safe to say that we are on the brink of the 30-hour week. This workweek, with our present means of production, is entirely within the realm of possibility. It is to be hoped that the decision to adopt it will not involve the emotional battles characteristic of all previous reductions of the workweek, but rather that management and labor will approach the question from a factual basis and decide it solely on whether it will result in decreased production and increased costs, or whether production can be kept up and costs maintained or lowered. Many people may be surprised to learn that in manufacturing as well as retail organizations such a workweek can yield maximum production and satisfaction.

There are some who are ostensibly worried about the shorter workweek for fear the employee will not know what to do with his spare time. If the company and the community are really concerned with this problem and attack it intelligently, it provides no cause for worry. Other people insist that an honest day's work stretches from sunrise to sunset and that everyone should work an "honest day." In certain respects, this book is dedicated to the joys of working, but it also acknowledges that too much of anything can give cause for worry. The marvels of our productive ability have contributed toward the shorter workday and thus enabled us to transfer our interest to other things. An "honest day's" work is desirable, but there is no reason to believe that an 8-hour day—which was once 12 hours, then 10—represents the acme of honesty. If a college professor had to teach 8 hours a day, he would have no time to write books (maybe that would be a good thing!).

UNPRODUCTIVE WORKING TIME

In a survey and analysis of the working day, Angles (1930) found that over 30 percent of the working day was unproductive. Collecting materials, delivering finished goods, sharpening and preparing tools, and consultations absorbed this 30 percent.

The present authors' observations of professional workers indicate that unproductive working time is much greater, amounting possibly to 50 percent of the working day. Professional workers—be they researchers, writers, or statisticians—present weird problems to a director of research. The badge of distinction of this group seems to be the privilege of arriving between 10 and 20 minutes late. Coffee and newspaper time is a necessary allowance several times each day. Informal consultations with fellow workers take their toll. Taking a leaf out of the executive's book, the professional worker finds that the one-hour lunch period is totally inade-

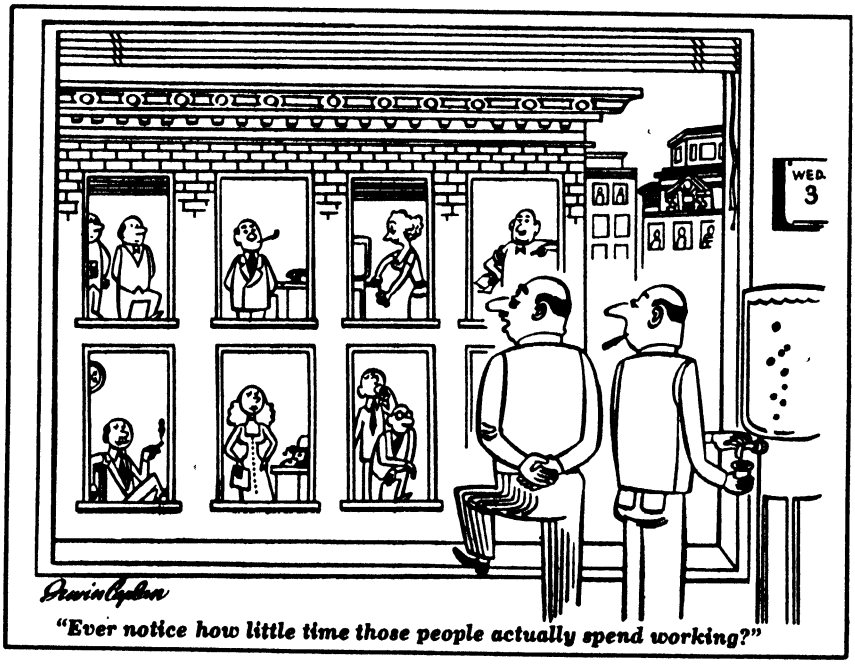


Figure 18.11. (From *The American Legion Magazine*. By permission of the magazine and the cartoonist, Irwin Caplan.)

quate; he usually manages to get around this by going to lunch at a different time every day, hoping thus to prevent a check from being made on his habits. In some cases it is almost three o'clock before the day's work is begun. Many professional people avoid a sense of guilt by bringing a briefcase in which they take work home, but the work is rarely done until the deadline for the finished job is upon them.

This is not a tirade against professional workers; it is merely an illustration of the inroads that unproductive working time can make. The employer, factory worker,

clerk, and professional worker must acknowledge that a considerable part of the day is spent in unproductive activity. It is because of this that the nominal workweek can justifiably be reduced without in any way affecting production.

Many attempts have been made to increase production by cutting down the unproductive time, but this will not be discussed here because it is not directly related to work fatigue or monotony.

REST PAUSES

The Hawthorne Studies indicated that rest pauses are valuable aids to increasing production. However, the length of the rest pause and what the employee shall do during it deserve consideration.

From the point of view of management, it is best to give the employee a regular rest pause, for if management does not do so, the employee will take it anyhow. Some employees spend more effort and energy in "soldiering" on the job than in doing the work they are supposed to do. A typist may not leave her desk or take her fingers away from her machine, but she may be indulging in a rest pause; she may be typing personal letters or other material not connected with her job. The professional worker who is suddenly attacked by drowsiness during the afternoon learns to rest his head on his hand with his index finger propping one eye open as if he were in the process of "creating." The man who works at a machine often finds it necessary to adjust the machine—it suddenly "doesn't seem to be working just right." In other words, during a normal working day employees will take a rest pause. When management offers such a rest or forces them to take one, much is gained in the form of goodwill and a reduction of guilt feelings, plus the alleviation of fatigue or monotony if it exists.

Depending upon the type of work, the number and duration of rest periods vary. For most jobs a 10- to 15-minute rest in midmorning and midafternoon seems to bring the best results. To determine definitely whether one or two rest periods are necessary, data should be gathered on production and the employees should be interviewed. Such brief experimentation will often lead to the correct answer.

It is important that the employee change his posture during the rest pause. If he sits while working, he should be encouraged to stand or walk around. If he stands, he should be encouraged to sit down or, if possible, to lie down. The important thing in a rest pause is that the activity and posture of the body be different from the activity and posture required on the job. For example, basketball as it is played today is strenuous and active. Watching the leading college teams play provides a vivid illustration of the principle that the activity should change during the rest pause. However, various teams do different things during the rest period. Some just sit; others stand, and still others lie down. Some teams seem to be so disorganized that the men can do anything they want during the period. But regardless of posture or activity, all the teams seem to recover equally.

In some industrial organizations refreshments are served during the rest pause. The value of the coffee, soda, or ice cream that is provided lies in creating a favorable employee attitude toward the rest and in filling the time, rather than in supplying nourishment for the energy needed to continue work.

ABSENTEEISM

The incidence of absenteeism should not be overlooked in considering the relationship between production and nominal hours of work per week. An increase in the workweek often results in an increase in absenteeism, which ultimately means a reduction in the hours worked. Any serious investigation of the causes of absenteeism reveals that there are numerous reasons for it. Also, an employee often has more than one reason for being absent. Schenet (1945) conducted a careful survey of absenteeism in a certain factory. The study, which was based upon an analysis of the records of 280 men and 470 women over a four-month period, showed that absenteeism was three times greater among women than among men. For sickness absences, the rate for women was twice that for the men, and for personal absences the rate was three or four times greater for women than for men. There was no great difference in absentee rates on the basis of age groups, except for a very slight tendency for older employees to be absent more frequently. Schenet found vast differences in absenteeism from one department to another; in the final assembly department the absentee rate for both sickness and total absences was the highest of all. He regards this as extremely puzzling and says that no data were uncovered that shed any light upon it.

However, the work referred to earlier when the concept of morale was considered may present some light. Differences in supervision and/or differences in morale could have existed in the departments. Either could contribute to the differences in the absentee rates even though the final assembly department is assumed to have more interesting work (which it may not). The more recent work of Covner (1950) and Mann and Baumgartel (1952) shows that absenteeism rises as supervising quality decreases.

In a study based upon a sample of 550 employees, Jackson (1944) found that the causes of excessive absenteeism were as follows:

1.	Poor work habits	6%
2.	Personal adjustments	9%
3.	Dissatisfaction with work	16%
4.	Irresponsibility	17%
5	Outside difficulties	17%
6.	Sickness or fatigue	35%

Also, in a study reported earlier in Chapter 5, Naylor and Vincent (1959) found female absenteeism among 220 clerical workers to be highly related to the number of dependents claimed by these women on their withholding forms.

This work, together with other material in the literature, indicates that although sickness is the greatest single cause of absenteeism, it is only a minor cause when the total picture is considered.

Individual adjustment, emotional difficulties not related to the job, and other individual problems contribute to making absenteeism a psychological problem. It is advisable to ask an employee with a high absentee rate to come in for an interview in which the basic reasons for his absenteeism can be discussed. Often, during such an interview the groundwork can be laid for a better understanding with him, and attempts can be made to correct the irrelevant conditions which may be responsible for the absenteeism.

It is interesting to note that absenteeism is highest at the beginning and end of the week and lowest on payday regardless of what day in the week it falls. The distance traveled to and from work and household responsibilities, especially on the part of women employees, are also factors which contribute to excessive absenteeism.

The point to be remembered about absenteeism is that employees must have time to take care of their personal needs; if the hours of work prevent this, then absentee rates will be high. When an employee takes time off on his own, he often feels that since he is paying for it he does not have to inform management in advance. If management were to grant an employee a half-day off for shopping, it might result in an increase in production rather than a decrease, if for no other reason than that management would have someone else take his place on the production line.

Covner (1950) presents a delightfully refreshing and different view on absenteeism. He states, "At the time of inaugurating the study management voiced the hope that the study would provide a 'convenient means for getting rid of worst offenders, and preventing the hiring of potential offenders.' Somewhat to management's surprise, the findings pointed to management factors as main causes of absenteeism."

The major finding in this study is that two factors appear to be most closely associated with absenteeism—size of department and quality of supervision. Covner considers absenteeism as management centered or worker centered and believes that these two centers should not be thought unrelated. He also cogently sees the relativity of absenteeism to loose or tight labor market conditions, and inferentially suggests that, depending on such conditions, symptoms other than absenteeism—for example, grievances or uncooperativeness—may arise.

Mann and Baumgartel (1952) report a study typical of the University of Michigan group. They compared the attitudes of high and low absentee rate employees. The subjects were 163 office employees and 251 "blue-collar" men working for the Detroit Edison Company. The investigators found that a more favorable attitude toward supervisor and a more favorable feeling about work associates accompanied lower absentee rates for both blue- and white-collar workers.

The foregoing makes it clear that the incidence of absenteeism is low due primarily to excessive fatigue, and that, indeed, absenteeism is a very complex variable which is affected by a variety of factors.

REFERENCES

Accident-Prone Employee, The. Metropolitan Life Insurance Company, New York, 1945.

Angles, A. Unproductive working time. In C. S. Myers (ed.). *Industrial psychology,* Thornton Butterworth, London, 1930.

Arbous, A. G., and J. E. Kerrich. Accident statistics and the concept of accident-proneness. *Biometrics,* 1951, 7, 340–432.

Bedale, E. M. Comparison of the energy expenditure of a woman carrying loads in eight different positions. *Industrial Fatigue Research Board Report* No. 29, His Majesty's Stationery Office, 1924.

Bowled Over. New York *Post,* New York, February 4, 1954.

California Driver Study, 1964, Part 2: Accidents, traffic citations and negligent operator count

by sex. *Report No. 20*, State of California, Department of Motor Vehicles, Division of Administration, March 1965.

Cobb, P. W. The limit of usefulness of accident rate as a measure of accident proneness. *Journal of Applied Psychology*, 1940, 25, 154–159.

Covner, B. J. Management factors affecting absenteeism. *Harvard Business Review*, 1950, 28, 42–48.

Crowden, G. P. *Muscular work, fatigue and recovery*. Pitman, London, 1932.

De Silva, H. R., P. Robinson, and T. W. Forbes. Some psychological factors in accident repeater drivers. *Journal of Abnormal and Social Psychology*, 1939, 34, 124–128.

Dill, D. B. The nature of fatigue. *Personnel*, 1933, 9, 113–116

Drake, C. A. Testing for accident proneness Paper read before the American Psychological Association at the University of Minnesota, 1937.

Fatigue and hours of service of interstate truck drivers. Public Health Bulletin No. 265, Government Printing Office, Washington, D.C , 1941.

Flanagan, John C *Research on near-accidents* American Institute for Research, Pittsburgh, 1953.

Forbes, T. W. The normal automobile driver as a traffic problem. *Journal of General Psychology*, 1939, 20, 471–474.

Geldreich, E. W. Some physiological concomitants of mental work *Psychological Monographs*, 1953, 67(8).

Goldstein, L. G. Human variables in traffic accidents National Academy of Science, *National Research Council Highway Research Board Publication* No. 1054, Washington, D.C., 1962.

Greenwood, M , and H. M. Woods. The incidence of industrial accidents upon individuals with special reference to multiple accidents. *Industrial Fatigue Research Board Report No. 4*, 1919.

Griffith, J. W., W. A. Kerr, T B. Mayo, Jr , and J. R. Topal. Changes in subjective fatigue and readiness for work during the eight hour shift *Journal of Applied Psychology*, 1950, 34, 163–166.

Haner, C. F. Use of psychological inventory in writing insurance for youthful male drivers. *Traffic Safety*, 1963, 7, 5–14.

Harris, F. J. Can personality tests identify accident prone employees? *Personnel Psychology*, 1950, 3, 455–459.

Hours of work and output. Bulletin No. 917, U.S. Department of Labor, Government Printing Office, Washington, D.C., 1947

Hours of work and production. *Labor Information Bulletin*, 1944, 2, 4–7.

Huxtable, Z. L., M. H. White, and M. A. McCarton. A re-performance and re-interpretation of the Arai experiment in mental fatigue with three subjects. *Psychological Monographs*, 1946, 55(5).

Jackson, J. J. .Factors involved in absenteeism. *Personnel Journal*, 1944, 22, 289–295.

Johnson, H. M. The detection and treatment of accident-prone drivers. *Psychological Bulletin*, 1946, 43, 489–532.

Kerr, W. A. Accident proneness of factory departments. *Journal of Applied Psychology*, 1950, 34, 162–170.

Knight, R. Work and rest. In C. S. Myers (ed.). *Industrial psychology*. Thornton Butterworth, London, 1930.

Lauer, A. R., *et al*. Aptitude tests for Army motor vehicle operators. P.R.S. Report 981, Contract Research Report to Personnel Research Section, Washington, D.C., Department of the Army, 1952.

McFarland, R. A., and A. L. Moseley. *Human factors in highway transport safety*. Harvard School of Public Health, Boston, 1954.

Mann, F., and H. Baumgartel. Absences and employee attitudes in an electric power company. University of Michigan Human Relations Program Series 1, Report 2, University of Michigan, Ann Arbor, 1952.

Mintz, A., and M. Blum. Accident proneness is overrated. Paper read before the Eastern Psychological Association at Temple University, Philadelphia, 1948.

Mintz, A., and M. Blum. A re-examination of the accident-proneness concept. *Journal of Applied Psychology*, 1949, 33, 195–211.

Mosso, A. *Fatigue*. Putnam, New York, 1915.

Muscio, B. Is a fatigue test possible? *British Journal of Psychology*, 1921, 12, 31–46.

Nader, Ralph. *Unsafe at any speed*. Grossman, New York, 1965.

National Academy of Sciences, National Research Council. *The Field of Highway Safety Research*. National Research Council, Washington, D.C., 1952.

Naylor, J. C., and N. L. Vincent. Predicting female absenteeism. *Personnel Psychology*, 1959, 12, 81–84.

Optimum hours of work in war production. *New York Times Magazine*, April 5, 1942.

Ryan, T. A. *Work and effort*. Ronald, New York, 1947.

Scheidt, V. Paper delivered before the American Association of Applied Psychologists, University of Minnesota, Minneapolis, 1937.

Schenet, N. G. An analysis of absenteeism in one war plant. *Journal of Applied Psychology*, 1945, 29, 27–39.

Uhlaner, J. E., L. G. Goldstein, and N. J. Van Steenberg. *Road-user characteristics: Development of criteria of safe motor-vehicle operation*. Highway Research Board Bulletin 60, National Academy of Sciences, National Research Council, Washington, D.C., 1952.

U.S. Department of Labor. Guide to industrial accident prevention through a joint labor-management safety committee. *Labor Information Bulletin*, February 1947.

United States Women's Bureau Bulletin No. 105, U.S. Department of Labor, Washington, D.C., 1933.

Vernon, A. M. The influence of rest pauses and changes in posture on the capacity of muscular work. *Industrial Fatigue Research Board Report No. 29*, 1924.

Wilson, C. E. Speech to American Society for Metals. Reported in the *New York Herald Tribune*, October 27, 1950.

19 WORK ENVIRONMENT

The problem of increasing production and also making work more pleasant has been approached through the introduction of changes in the working environment. There is a fundamental difference between this approach and that in which increased efficiency is attained as a result of time and motion studies. Although such studies sometimes lead to environmental changes, the changes are usually related to the job, such as changing the height of a stool or the size of a work area. In other words, time and motion studies most often result not in environmental changes but in changes in an integral part of the job or task being performed.

There are still other approaches, such as increasing efficiency through refined selection techniques. Under these circumstances no direct attention is given to the environmental factors related to the job.

The list of the various environmental changes that could conceivably be introduced in industry is lengthy. Changes related to noise as it affects work, as well as changes in connection with the illumination, ventilation, and temperature of the work environment, have been introduced with varying claims of success. A popular environmental change is the introduction of music into the office or factory. Many claims in connection with a change in production have been based on the use of various color schemes, primarily on factory walls, but also on benches and machines, and in rest rooms.

Another category of environmental changes for improving production includes such miscellaneous items as eating facilities, fresh drinking water, and even the physical distance between two coworkers.

Unfavorable environmental conditions supposedly contribute to a slowdown of the employee's activities and production. They allegedly increase turnover, promote high absenteeism, and generally contribute to inefficiency.

There is no doubt that people generally prefer pleasant to unpleasant surroundings and that when attention is paid to creating a favorable working environment, as well as to actual job performance methods, overall pleasantness prevails. However, one must be somewhat cautious in accepting all the claims made as to the results of creating a favorable work environment. Much of the work that has been done in this field has been guilty of serious error in its experimental methodology, and naive assumptions have too often been made with reference to the work. In an

ideal experiment one factor is varied and all other factors are either eliminated, neutralized, or held constant.

It is not safe to assume that a change in production can be attributed to a specific environmental change. When an environmental change is made, at least two things happen. First, there is the changed environment, and second, and equally important, *there is the response to change in general.* This response may be due only partially to the specific change; it may also be due to change of a more generalized nature. Suppose that sound control is introduced into a plant and that work proceeds with less noise. Let us assume that production increases 5 percent. According to sound scientific procedure, we are not justified in attributing this 5 percent increase solely to sound control. Although a certain amount of the increase *is* the result of the sound control, a certain amount of the increase is due to the fact that *a* change has taken place. The attitude of the employee must be reckoned with in this connection, for the same 5 percent increase in production might have occurred when *more* noise was introduced. For instance, in a hypothetical situation production might conceivably be low because the employees spend their time talking to one another. If someone introduced a new infernal machine which created so much noise that it made conversation impossible, we would be more cautious in coming to the conclusion that the introduction of the noisy machine, of itself, increased production.

In other words, the major problem in the introduction of environmental change is whether the resulting increase in production is to be attributed to the factor that *is* changed or to the factor that is incidental to *a* change.

In addition to this error in methodology, there is the fact that the employee and his attitude toward change have not been fully taken into consideration. An altruistic employer may take steps to improve the work environment, but whether such steps will result in increased production—as is always claimed—will depend upon how the employees as a group interpret this change. If they believe that he has spent a lot of money in order to take advantage of them, they will resist the change, even though the environment is made more pleasant. If, on the other hand, the change raises morale, it will have the result he desired.

An excellent illustration of this is the case of the employer who had his factory and sales room in the same building. Air conditioning was installed in the sales room, but not in the factory. The employer succeeded, primarily, in creating greater dissatisfaction. The salesmen disliked the air conditioning because they believed that they would have more colds on account of it. The factory employees, who recognized that they held jobs of less prestige value, interpreted the change as just another instance of their being held in lesser esteem. Both groups felt that the installation of air conditioning in the sales room was motivated not by their employer's concern for his employees' welfare but solely as a method of increasing business.

The Hawthorne Studies, reported rather fully in Chapter 10, reviewed the effects of a change in illumination on production. It will be remembered that no direct relationship was shown to exist between the physical change and production. The contribution of the Hawthorne Studies in relation to all changes involving environmental conditions is significant.

Too often variables are not controlled. For example, an increase in temperature and the resulting production figures will be reported without taking the precaution of determining whether no temperature change or a decrease in temperature would result in a change in production.

The essential point of connection with studying changes in work environment is the use of adequate experimental controls before conclusions are drawn. The authors consider that changes in work environment, especially when they are conducive to more pleasant surroundings, are desirable. However, a change in production figures often has little or nothing to do with such environmental changes. This does not mean that one should forsake the idea of introducing music, decreasing noise, or controlling the temperature of the workroom. It *does* mean that miracles cannot be expected as a result of these changes, and that each positive claim must be substantiated.

MUSIC IN INDUSTRY

A popular environmental change is the introduction of music during working hours. Although plant broadcasting is little more than thirty-five years old, most of the "sound and fury" has been relatively recent. Possibly one of the reasons for the popularity of music is the variety of ways in which it can be provided. On an extremely informal basis, music can be brought into a plant by merely plugging in a radio and allowing it to blare from the beginning to the end of the workday. Some employers have introduced music in an even more informal way by encouraging their employees to sing. With modern transistor radios and their ear-plug attachments, it is possible for a worker to have his own individual music source—often without the knowledge of his coworkers or supervisor!

In this connection, a sad yet humorous experience was called to the writers' attention by an employer who had a small factory. This man had serious doubts about the advisability of music in industry; he said that there had been bedlam in his plant ever since he allowed his employees to sing. His story was so pathetic and so amusing that a visit was made to the factory. The girls worked at a series of long work tables in relative silence; suddenly one girl would start to hum, and the others would join her shortly afterward. Actually, this man's employees comprised two minority groups; singing seemed to make the girls suddenly conscious and proud of their minority status. When one girl began singing a native song and was joined by others in her group, this was a signal for the other group to sing some of its native songs. From then on, challenge was hurled upon challenge, volume increased steadily, and chaos ensued. Needless to say, this is not the way to use music in industry.

Music broadcast by radio requires a minimum capital investment—the purchase of a radio—but may introduce other difficulties. Just what station is to be tuned in and who shall have the right of decision often becomes a problem of management. Years ago a factory that manufactured radio tubes had to prohibit playing the radio during lunch hour because of the excessive arguing among the employees as to what type of program they should listen to. The installation of additional radios was no solution. This company temporarily solved the problem of music during the lunch hour by permitting phonograph records to be played. Since the employees brought in the records and they were played in sequence, there was less chance for bickering.

Another disadvantage of the radio for industrial use is the commercials, although FM stations generally have fewer of these. Furthermore, the type of music cannot

be controlled. It is a fact that some young women workers will stop work and listen with rapt attention when some star is broadcasting, and the star is on the air more often than the boss would like. The radio may not be heard with equal intensity and clarity in various parts of the plant, but a loud-speaker system can overcome this.

A more desirable method of providing music is to use either piped-in music or a standard system of industrial broadcasting controlled by the firm itself or by an outside source.

EFFECT OF MUSIC

The generalization is possible that music during repetitive factory work results in a slight increase in production. Some of the best experimental work in this field has been directed by Williard A. Kerr (1954). Recognizing the need for basic research in order to clear up the maze of anecdotes and armchair speculations on the subject, he carried out four specific experiments on the introduction of music in actual industrial conditions. In all four experiments, records were played over the company broadcasting system.

The first experiment was made in the Paper Capacitator Department of a modern factory and lasted a little less than two months. On "music days" a varied musical program was broadcast to all 197 employees at the following times: 9 to 9:15, 10:15 to 10:35, 11:30 to 12:00, 12:00 to 12:30, 1:30 to 1:45, 2:45 to 3:10, and also at the beginning and the end of the shift. The subjects in this experiment were 64 women operators, 90 percent of whom had been accustomed for at least five months to music in the department. This latter control is important because the fact that the employees were accustomed to music made it possible to test the effect of *music* rather than the effect of a *change*. The subjects were not told that they were taking part in an experiment; this is also important, because production might vary as a result of the suggestion of the experiment rather than of the experiment itself.

The experiment involved no change in working habits or conditions except that there was no music two days out of every four. On "no-music days" the employees were told that work was being done on the plant's sound system. Saturday was excluded from the experiment because the company recognized that the level of operation that day was irregular and subject to irrelevant outside influences. Although these influences are irrelevant from the company's point of view, it is ridiculous to presume that they are irrelevant from the employees' point of view. An important date Saturday night or the anticipation of a day off on Sunday can interfere with an experiment on music in industry. Furthermore, if these so-called irrelevant influences are not taken into consideration, they will interfere with an entire efficiency system, regardless of whether it is based on industrial psychology or something else.

Kerr collected data for each of forty days on three operations—roll assembly, winding, and can assembly. Quantity, quality, and net good yield—that is, production with both quantity and quality taken into consideration—were the measures of performance obtained. Although none of the differences were statistically significant, *quantity* of production was higher in all three operations when there was music, but *quality* was worse in two out of the three operations where such a measure was available.

In the roll assembly operation, quantity was + 0.75 percent better and the net good yield was + 0.57 percent better; but scrappage, which is a measure of the

quality of production, was 9.89 percent greater with music. The fact that production increased but so did spoilage raises a serious question as to what is the ultimate measure of success. In this instance, the answer is provided by applying the criterion of "net good yield," which is a combination of quality and quantity; in these terms overall production was bettered by approximately 0.5 of 1 percent with music. For the winding operation, quantity was 1 percent better, but scrappage was 14 percent greater. Kerr found it impossible to obtain a net good yield value for this operation. In the can assembly operation, quantity increased by +0.43 percent. This first experiment of Kerr's indicates that there is a slight increase in production, but there is also an increase in spoilage.

The second experiment conducted by Kerr took place in a factory whose employees worked on quartz crystals. There were 53 subjects, all members of a union and working on a straight hourly rate; the experiment lasted 107 workdays. Again,

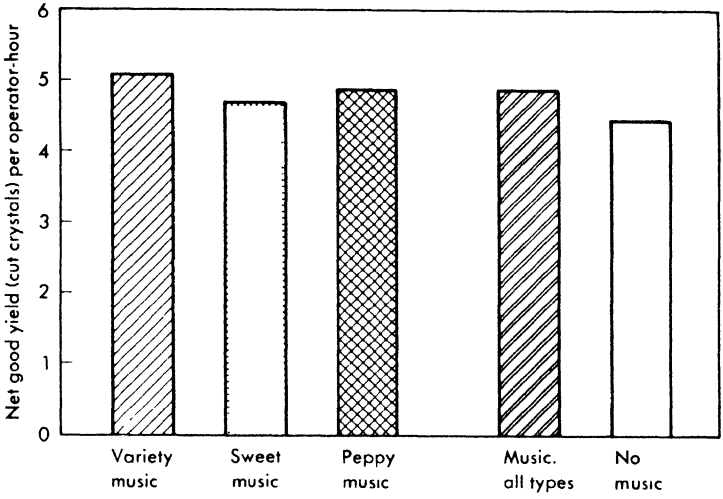


Figure 19.1. Effect of various types of music production. (Adapted from W A. Kerr Experiments on the effects of music on factory production. *Applied Psychology Monographs No.* 5, 1945, 24. With the permission of the author and of the publishers, Stanford University Press, Stanford, Calif.)

music was played at the same periods of the day as in the first experiment. There were three days of music, then three days without music. Three kinds of music were broadcast, but on no two days in any three-day cycle was the same type used. The first kind consisted of a variety program; the second was "sweet" music (not jazz and not too heavily accented); and the third type was "peppy." Because of the length of time covered by the experiment, it was assumed that the effects of outside influences, such as weather, payday, personal joys and sorrows, or a bad run of quartz, would be eliminated or held constant.

Six measures of production were obtained for "no music" and the various types of music. Figure 19.1 presents the results, with net good yield as the criterion. Once again Kerr finds none of the obtained differences approaching statistical significance. The quantity of crystals finished is greatest with peppy music, but the quality is poorer than with variety or sweet music.

There are no consistent differences on the basis of type of music, but a production increase is found when the average for 56 music days is compared with that for 51 no-music days. Scrappage is lower on music days than on no-music days. Whereas in the first experiment the average increase in production was about 0.5 of 1 percent, in this experiment the average increase is much greater. For example, the forty employees doing the finishing job showed an increase in quantity of 4.82 percent with music and 8.3 percent less scrappage; the net good yield is 9.07 percent higher. The net good yield among the 53 employees is + 7.64 percent. Kerr finds that all differences in the six measures of production favor music as opposed to no music.

In his third experiment, Kerr used a group of 520 women operators in a glass radio-tube factory. This factory had broadcast to its employees a radio music program known as the "Make-Believe Ballroom" over a long period of time. In addition, a birthday, anniversary, and soldier request program was broadcast between 1:00 and 1:30 P.M. According to management, this program was very popular and it was deemed unwise to discontinue it during the experiment; accordingly it was continued.

A music-preference questionnaire survey was conducted among the employees, sixteen types of music being ranked from most to least popular, as follows:

1. Hit parade
2. Patriotic
3. Waltzes
4. Marches
5. Hawaiian
6. Semiclassical
7. Polkas
8. Sacred
9. Hillbilly
10. Classical
11. Blues
12. Western
13. Fast dance
14. Spirituals
15. Humorous-novelty
16. Square dances

All sixteen types were liked to some extent; the average of each of these categories was always greater than the indifference point. On the "more-music" days one of three types of music—Hit Parade, waltz-Hawaiian, and march-polka—was played at the following times: 8 to 10:30 A.M., 11:00 to 11:45 A.M., and 1:30 to 3:15 P.M. In terms of the questionnaire rankings, the first type was the most popular; the second type included the third and fifth most popular, and the third group included the fourth and seventh most popular.

Before the experiment was begun, the group was told that it would hear a specific type of music on a given day. On one day it would be Hit Parade music, on another only waltzes and Hawaiian music, on the third polkas and marches. It was also told on some days there would be no music, except the two programs which it had been accustomed to hearing. The group was informed that after some weeks it would be asked to vote on the type of program it most preferred.

Three jobs, all requiring a high degree of hand-eye coordination, were performed by these 520 operators. One interesting finding was that the amount of music played on march-polka days had to be reduced because of the complaints from the employees. Some of these workers threatened to quit unless "something is done about those marches and polkas." However, when this music was included in the Hit Parade programs, there were no such complaints.

Again, as in the preceding experiments, Kerr obtained production averages for

the various types of music days. We should mention that these employees were on an incentive basis. The findings are presented in Table 19.1.

Both output and quality were lower on waltz-Hawaiian days than on Hit Parade or march-polka days, even though complaints and comments indicated that the employees liked waltz-Hawaiian better than march-polka music. The complaint against the waltz-Hawaiian programs was, "They put us to sleep." Production on march-polka days was slightly greater than on Hit Parade days, but the quality was better on Hit Parade than on march-polka days in two of the three cases.

TABLE 19.1 *Average Quantity and Quality in Radio-Tube Manufacture with More Music, Less Music, and Various Kinds of Music*

Output	*Hit Parade (14 days)*	*Waltz-Hawaiian (14 days)*	*March-Polka (13 days)*	*Less Music (13 days)*	*More Music (41 days)*	*Percent Better with More Music*
Quantity						
Miniature tubes	13.65	12.89	15.34	13.70	13.92	+1.61
Glass and power tubes	13.12	13.09	12 92	12.92	13.05	+1.01
Quality						
Miniature tubes	95.94	94.30	95.54	99 40	95.41	−4.18
Glass tubes	79.35	76.29	76.65	85.64	77.48	−10.53
Power tubes	97.77	100.35	104.83	101 97	100.90	−1 06

SOURCE: W. A. Kerr. Experiments on the effects of music on factory production. *Applied Psychological Monographs, No. 5,* 1945, 29.

Kerr's fourth experiment concerned the differences between orthacoustic recordings and ordinary records. Slight but statistically nonsignificant differences appeared in favor of orthacoustic records. However, since the playing time for the ordinary record was three minutes and for the orthacoustic two and a half minutes, it may be that this, rather than any difference in the quality of recording, was responsible for the results.

The primary reason for reporting this series of experiments in such detail is to illustrate careful industrial experimentation. Although the findings in general show that production increased slightly, they make it clear that the introduction of music does not increase production to any considerable extent. The big problem in relation to music in industry—the attitude of employees toward the music—is touched upon only slightly by Kerr. His third experiment showed that there is no definite relationship between the order of preference as to type of music and production. In fact, music of an unpopular variety, the polka-march category, does not affect production at all.

In another study in this field, Smith (1947) distributed a questionnaire to approxi-

mately 1000 employees in a factory manufacturing a small radio part. The results showed that 98 percent of the employees thought that music during working hours would be "mildly pleasant" or "extremely pleasant." In studying a day shift and a night shift of 21 employees each, Smith found that when music was introduced the average production increase for the day shift was 7 percent, and for the night shift it was 17 percent. However, this increased production may be due only in part to the music; some of it may be due to the change.

Smith found that maximum production was obtained when music was played 12 percent of the time on the day shift and 50 percent of the time on the night shift. The more an employee wanted music, the more the music tended to increase his production; and the more the employee's job permitted conversation while working, the more the music tended to increase his output. Smith concludes, "Music probably produces its major direct effect when the individual's capacity for attention is not absorbed by his work; in this circumstance, music appears to direct unused attention from brooding, talking, or off-the-job activities."

McGehee and Gardner (1949) conducted a study to determine the effects of music on production in a relatively complex industrial job known as "setting" in rug manufacturing. Their results are unlike the findings of Kerr and Smith. They found that music had neither a favorable nor an unfavorable effect on production. In response to a questionnaire, 59 percent of the workers said they got more work done with music as compared with a negative response of 7 percent. This is most interesting since it clearly shows that attitude cannot be accepted as a measure of behavior. Although the workers may have had a more favorable attitude and believed that they produced more, they did not produce more.

McGehee and Gardner find that the assumption held by many that more favorable attitudes and a reduction in monotony increase production is not substantiated in fact. With reference to the specific investigation, the results show that for a complex industrial job and under stable conditions music does not increase production.

A factor that is generally ignored in studies of the effect of music on production is rhythm. It is entirely possible that music influences production most when its rhythm is compatible with the employee's rhythm of work or when it tends to pace him.

An unpublished study by a student in a class in industrial psychology clearly established that more people skate on a rink when waltzes are played than with other types of music or when there is no music. Furthermore, when the loudspeaker is switched to waltz music, the speed of the skaters increases.

While some work has been done in industrial situations, little, if any, worthwhile experimental work has been done with office work situations. However, music in industry has been extended to offices.

As McGehee and Gardner point out (1949), "Too often the effect of music on production, absenteeism, turnover, accident rates, and workers' attitudes is 'measured' in terms of the optimistic beliefs concerning its effectiveness held by those responsible for its installation and programming."

An organization known as Muzak, which is probably the largest single supplier of industrial music systems, has done considerable research on the effect of music. According to the research director of this company, waltzes, ballet music, light concert selections, and concert arrangements of popular dance music are suitable for office workers, but an inspection of procedure, controls, and experimental method

forces the conclusion that this work is not as rigorous as that of Kerr, Smith, and McGehee and Gardner.

Among the concerns regularly using planned musical programs for their office employees and being programmed by Muzak are Metropolitan Life Insurance Company; Northern Mutual Insurance Company; Research Institute of America; Reader's Digest; Erwin, Wasey & Company, Inc. (advertising agency); Bank for Savings (all branches); American Cyanamid Company; American Tobacco Company (general and executive offices); Chemical Construction Company (architectural and drafting rooms); Prudential Insurance Company of America; Emigrant Industrial Savings Bank; Reuben H. Donnelley Corporation; Liberty National Bank; McGraw-Hill Publishing Company, Inc. (direct-mail offices); McClellan Stores Company (general and executive offices); National City Bank of New York (personal loan division); Detroit Edison Company; Washington Gas Light Company (accounting offices); Sears, Roebuck & Company of Philadelphia (mail-order department); and Thomas Publishing Company.

Production records suitable for office workers are more difficult to obtain than those for factory workers; consequently, the results indicating the desirability of music are based on employee responses to a questionnaire. All the surveys conducted by Muzak show that employees are in favor of planned music during work.

There is need for an experiment among office workers of the type that Kerr conducted among factory workers. Apparently employee effort and attentiveness can counteract or vary the influence of periods of music. Although there is not much evidence to support this assumption, it is partly supported by the conclusions in a somewhat related field, noise in work.

THE "AROUSAL" HYPOTHESIS

Very little has been accomplished toward developing any type of theory relating human performance to either noise or music. The one notable exception is the application of the "arousal hypothesis" as a model for understanding the effects of auditory stimulation upon the worker. Duffy (1951) was one of the earliest to point out the possible importance of the overall *level of activation* of a person in determining his task performance. Activation level may be defined as "the degree of excitation of the brain stem reticular formation" (Scott, 1966). The determinants of an individual's activation level are postulated to be such things as (1) *stimulus intensity*, (2) *stimulus variation*, (3) *stimulus complexity*, (4) *stimulus uncertainty*, and (5) *stimulus meaningfulness*.

Thus, intense stimulation, complex stimulation, etc. are postulated to be able to "arouse" the individual. Since noise and music are external stimuli which can be controlled in terms of intensity, variation, and meaningfulness, they qualify as potential "arousers." Human performance, according to Duffy (1962), tends to vary in accordance with the general activation level of the individual. However, the relationship is not a direct linear one—rather, it is described better by an inverted U-shaped function, as shown in Figure 19.2.

As Scott (1966, p. 13) explains: "At low activation levels, performance is handicapped by lack of alertness, a decrease in sensory activity, and lack of muscular

coordination (all of which are due to insufficient cortical stimulation from the BSRF).[1] At intermediate levels of activation, performance is optimal, and at high levels performance is again handicapped by hypertensiveness, loss of muscular control, 'impulse to action,' and in the extreme, total disorganization of responses."

Direct tests of the arousal hypothesis have not been frequent. Scott has reviewed the research and found it supportive of the hypothesis but still quite meager. Two recent examples of studies which have tried to tie music and noise effects on job performance to the arousal hypothesis are those of McBain (1961) and Smith and Curnow (1966).

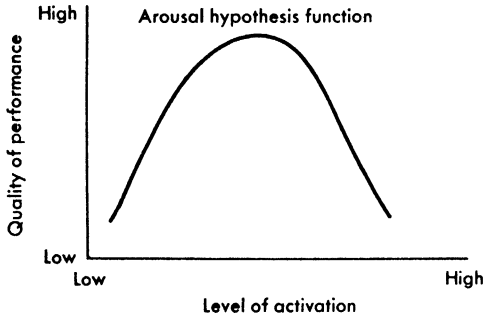


Figure 19.2. Hypothesized relationship between level of activation and human performance.

MCBAIN STUDY

In this study members of the Royal Canadian Air Force were made to perform a monotonous work task under both quiet (Q) and noise (N) conditions. The (N) condition consisted of playing in reverse a tape recording of speech (note: the decibel level was not given in the study). The results were ambiguous, in that while there was evidence with some performance measures that the N condition facilitated performance, the effect of the order in which a subject received the condition also seemed to have an effect as did a measure of the "degree of hypnotizability" of a person.

SMITH-CURNOW STUDY

In a rather intriguing test of the arousal hypothesis, Smith and Curnow varied the intensity of the background music in two large supermarkets to see what effect it would have on purchasing behavior. They found that shoppers spent significantly *less* time in the store during the loud music time periods, but that there were *no* significant changes in total sales or in customers' reported satisfaction as a function of the music intensity.

As a hypothesis for explaining the influence of noise, the "arousal" notion has a good deal of intuitive appeal. Certainly it could be used to explain a great deal of the findings discussed earlier in the research carried out by Kerr. In general, he found a tendency for performance to increase due to music as far as quantity was

[1] BSRF stands for Brain Stem Reticular Formation.

concerned, but at the same time a decrease in quality took place—thus people may have been "overly" aroused if quality is taken as the criterion of importance.

NOISE

Noise is generally regarded as a distractor and therefore as interfering with efficiency. In an experiment to determine the effects of noise, Vernon and Warner (1932) had a group of subjects do arithmetic problems and also read material in a book on psychology during alternating periods of noise and quiet. It was found that noise had no appreciable effect on the speed or accuracy of doing arithmetic problems, but that there was a slight increase in the expenditure of energy as measured by oxygen consumption. The factor primarily determining whether noise is a distracter is its character—whether it is steady or intermittent. When the noise is steady, the person adapts himself to it; but when it is intermittent, he must make a greater effort to maintain efficiency.

Park and Payne (1963) found that *average* performance was not affected by intense noise. What they did find was that *variability* of performance became much greater! Also important in their study was the fact that they only found this effect for a boring, easy task. With a task of high difficulty they observed no noise effect at all. Culbert and Posner (1960) examined the degree to which individuals can successfully adapt to noise. They found that after a period of several weeks individuals even adapted to a noise as intense as jet aircraft. However, no performance change was observed due to the adaptation to the noise.

The effect of noise on a job is apparently determined by whether the noise is a necessary accompaniment to the job or not. For example, a typist becomes used to the clatter of the machine because it is a necessary accompaniment to her work, whereas a person working next to her finds this noise an interference. Office workers in close proximity to machines in a factory are disturbed by the noisy machinery to a greater extent than the workers operating the machines.

Some years ago, Morgan conducted an experiment (1916) to discover how a person reacts to irrelevant noises. This experiment indicated that at first such noise generally retarded the speed of work, but that this was often followed by an increase in speed. The resulting speed was frequently greater than that achieved prior to the introduction of the irrelevant noises, because the subjects made an extra effort to overcome the effect of the noises. In Morgan's experiment the subjects exerted greater pressure on the keys and evidenced an increase in articulatory reactions. Morgan found that although noise does not necessarily interfere with efficiency as measured by production, it results in inefficiency as measured by energy cost. Ford reports similar findings (1929). The only additional evidence indicates that quiet as well as noise can constitute a distraction; thus the quiet that follows the cessation of noise acts as a distracting influence. This was confirmed both by the objective data and by the introspective reports of the subjects. In other words, an employee who has adjusted himself to a noisy work situation may be distracted by a sudden silence.

This does not mean that bigger and better noises should be sought. The evidence indicates that, although production is not curtailed by noise, more energy is ex-

pended in the form of increased effort. Also, although quiet working conditions are desirable, it does not follow that silence is necessarily golden.

Of course, noise of extreme intensity not only can be irritating but also can lead to actual physical pain and permanent ear damage. Generally, noise levels above 100 decibles (db) should be so considered, and employees should be protected from such noise by ear plugs or ear muffs.

Sleight and Tiffin (1948) reviewed the literature on industrial noise and point out that the complete condemnation of noise in industry may be unwarranted and that the harmful effects of noise have been overemphasized. It does appear, however, that hearing is impaired by industrial noise and that those subjected to the loudest noises are most affected. This would mean that although very noisy work conditions may be detrimental to hearing, ordinary work conditions are not deafening. It also appears that acoustical treatment to lower noise has more of an effect on attitude than on lowering noise.

Berrien (1940) also did a careful review on the effects of noise in work and finds that the popular literature abounds with emotional outbursts on the deleterious effects of noise, but the scientific literature rarely supports such views. Apparently, adaptation takes place—but seldom completely. Under high noise levels, hearing defects are frequently produced after long exposure. Still precisely unanswered, however, is the question of what levels at what exposure durations result in physical damage.

Summarizing the work reported on noise leads again to an important point. The "common sense" assumption that what is annoying is harmful and that noise is annoying and therefore harmful should be factually and experimentally determined and not emotionally proved. Noise may lead to production curtailment and deafness in some situations but surely not in all. Noise abatement campaigns should make less noise and stick to the point of being against noise because it is noisy.

ILLUMINATION

Ferree and Rand of the Research Laboratory of Physiological Optics have conducted much research on illumination and its relation to work. Their findings (1940a, 1940b) summarize much of the worthwhile information in this field and make possible the following generalizations:

Daylight provides the best illumination for work. Artificial lighting which closely approximates daylight in color and composition is next best.

Important characteristics of lighting which should be taken into consideration are the distribution and location of lights, the intensity of the light, glare, and the combination of artificial light and daylight. The best artificial light is the Mazda or yellow light; this has been found to be superior to the blue glass light, for maximum visibility is obtained in daylight or in lighting that approximates daylight as closely as possible. Artificial light should be as free from color as possible; light that is unbalanced toward any color is a detriment and not a help. Of the colored lights when equalized for brightness and saturation, yellow causes the least discomfort.

Possibly the most important difference between daylight and artificial light is diffuse-

ness. Sufficiently diffuse light tends to produce less glare. Ferree and Rand are somewhat disturbed at the tendency to sacrifice diffuseness of light for high intensities. Excessive intensity and poor diffusion can result in considerable eye damage.

One of the most common causes of visual discomfort and fatigue is brightness in the field of vision. Excessive brightness is often due to the source of light or the light fixture. Attempts to solve this problem have been made by devising lamp shades or in some other way shielding the eyes from the glare. A reflector that is turned down gives what is known as direct lighting; the light is directed toward the work area, the walls and ceilings being left dark or very poorly illuminated. In indirect lighting, the light is directed to the ceiling; from there it is reflected to the other parts of the room, especially the work area. This usually results in a disproportionately high brightness for the ceiling and a correspondingly low intensity in the work area. Translucent bowls which reflect part of the light to the ceiling and transmit part of the light downward have overcome the disadvantages of both direct and indirect lighting.

The problem of unevenness of lighting is clearly shown in the ordinary lamp. Although the desk lamp may make the work area sufficiently bright and at the same time save electricity costs, it results in a very unevenly lighted room. Most people, while working at a desk, continually look up or away from the work area; this requires continual pupillary adjustment, with resulting fatiguing effects. A makeshift arrangement would provide a lamp in another part of the room, thus reducing the difference in the light at the desk area and in the balance of the room and lessening eyestrain.

Ferree and Rand conducted tests on 550 people; 100 were in each 10-year age range from 10 to 60 years, and 50 subjects were over 60 years of age. Approximately 70 percent of these people preferred less than 15 foot-candles for reading 10-point type (average book text type); 50 percent preferred less than 11.3 foot-candles. In dividual differences are evident in this type of experiment, as in every other in psychology. These authors found a wide variation in the preference expressed in each age group. The people above 35 years of age tend generally to prefer more light for reading than those below 35.

A very important factor in lighting is glare. Glare may emanate from the work area or from the light source. All glare from the light source can be eliminated by the Glare-Baffles devised by Ferree and Rand.

These authors have also conducted useful research to dispel the peculiar notion that a mixture of daylight and artificial light is detrimental (1932, 1939). Even at present it is generally believed that such a mixture is unfavorable to vision; hence when artificial lighting is needed, some people take elaborate precautions to curtail daylight. Actually, there is no reason for this, because a mixture of the two gives a better and more comfortable light than an equal amount of artificial light alone.

A possible source of this popular misconception may be the difficulty experienced at twilight, when it is neither dark nor light. Automobile drivers are especially aware of this. The difficulty is not due to the combination of artificial and natural light; it results from the process of visual adaptation. The eye has been accustomed to bright light and the change in light requires a change in adaptation. As the darkness increases and adaptation to the new light intensity becomes more perfect, one sees more clearly.

In their experiments on the effects of the color of paper and ink on visibility, Ferree and Rand find that black ink on white paper that is free from gloss is best. White is the best color; saturated colors—saturation is the amount of color in the color—are inferior to unsaturated colors. Darker shades are inferior to lighter shades.

When colors are equalized in saturation and brightness, yellow is found to give the best results and an orange-yellow is next; but all colors are inferior to white.

Illumination and color are also considered in the combinations used for automobile license plates. From what has been said, black on white should be best, but road conditions tend to make the white look black—and black on black is not outstanding. Actually, the experiments that have been conducted in this field show that black on yellow is conducive to greatest visibility. The greens, blues, and other colors used on other license plates probably are a matter of local pride rather than an aid to visibility.

COLOR

As far as preposterous claims and unsubstantiated allegations are concerned, the so-called "color specialists" are eligible for the "grand prize." An article which appeared in *Popular Science Monthly* in 1947 dealt vividly with the use of color in industry. With reckless abandon it claimed that less fatigue, increased production, and greater safety result from the "scientific use of color in the factory." Some of the stories cited in this article are amazing. For example, as a result of "color conditioning" (whatever that is) one factory is reported to have had an increase of 15 percent in production and 40 percent in accuracy, along with a 60 percent decrease in absenteeism; in addition, "workers take more pride in caring for plant and equipment." This latter may provide a clue to what really happened. If the factory was originally dirty and unpainted and then the painters went to work—providing they were prevented from painting in stripes in hideous hues—it may well be that the employees liked the new conditions of work. But many of a number of color combinations might have had a similar effect. The essential point is how badly the factory needed the paint job in the first place. Any homeowner knows the marvels wrought by a coat of paint inside or outside the house.

This is not to imply that all work involving the color of walls is nonsense. The ability of the surface to reflect light and the contrast between the color of the work area and the wall color may in certain circumstances reduce eyestrain. For example, according to the above article, girls who inspected blue denim in a textile mill reported that they saw a peach color when they looked at the nonpeach wall. It is a fact that positive and negative afterimages occur when the eye is overestimated by one color. If a person looks steadily at a red square for approximately one minute and then looks immediately at a neutral background, he will see a green square on that background. This is a negative afterimage. But concerning the textile inspectors who saw "peach," there is no reason to believe that peach was the afterimage, because the negative afterimage of blue is yellow. However, if the blue had green in it, the afterimage may have been a poorly saturated red which might be called peach. The article goes on to say that a "color engineer greatly increased the time the girls could work without strain at this job by providing what their eyes demanded: peach-colored walls."

There is some basis for the idea that blue is a cool color and red a warm color, and interior decorators as well as "color experts" recognize this. Depending upon the illusion to be created, feelings of warmth or cold may be encouraged by the use of

these colors. But it is unlikely that the introduction of these colors will compensate for even a five-degree change in temperature.

Berry (1961) tried to determine to what extent color could indeed influence a person's subjective evaluation of temperature. The experimental design was quite complicated and elaborate. He studied five colors (white, yellow, amber, green, and blue). In his major test of the effect he found no significant differences due to color.

An example of work that is definitely the cause of skepticism is a report of an interview with Faber Birren on the occasion of the publication of his book *Color Psychology and Color Therapy*. According to the report, Birren maintains that "the right illumination and right color are worth $139.25 annually [to] an average employee in American industry." "Sounds like a bargain" appears to be the only appropriate comment to make to their claim. According to Birren, yellow is the color of intellectuals, blue is the favorite of introverts. Again a comment seems appropriate: "Want to bet?" In addition, what happens when a person is equal parts intellectual and introvert?

The evidence on the relation between color in industry and increased production is primarily based upon data that have not been subjected to rigid experimental tests. Consequently, this field must be considered a greater unknown than music, noise, or illumination.

VIBRATION

Many work environments involve a substantial amount of vibration. For example, crewmen attempting to operate complex electronic equipment in helicopters often complained during the Korean War that their task was made quite difficult due to the vibration caused by the rotors of the helicopter. Tank crewmen are another example of individuals who have to perform under conditions of extreme vibration —particularly when traveling over rough terrain.

One of the major problems under high vibration conditions is that the visual and motor processes of man are affected (McCormick, 1964). For example, the eyeball has a critical resonance frequency which when approached, appears to cause large performance decrements (Dennis, 1965).

MISCELLANEOUS FACTORS

A cafeteria is often installed in a plant because of necessity. A factory at some distance from restaurants or other eating places will have to have a cafeteria in order to attract and hold employees. However, a cafeteria is likely to become a thorn in the side of management. An insurance company in New York City supplies hot lunches to its employees; the meals are wholesome and provide a thoroughly balanced diet. But to induce employees to eat in the cafeteria this company has a rule forbidding them to get their coats until the end of the day. Consequently, even on the coldest days employees can be seen dashing from the building without coats to go

to the nearby soda fountain for a sandwich, a cigarette, and an ice-cream soda. Complaints about food, especially when it is supplied by the company, are very common. Hence many industrial cafeterias are a source of employee dissatisfaction as well as overhead cost. Even so, they are sometimes necessary as an environmental change.

A novel solution to the cafeteria problem has been proposed by Douglas Aircraft Company. It has 12 mobile cafeteria trains (see Figure 19.3). These facilities can feed 6000 employees during a single 30-minute lunch period. Each food train travels to a dining area and becomes a two-line cafeteria.



Figure 19.3. Mobile lunch carts solve the cafeteria problem. (Courtesy Douglas Aircraft Company, Inc.)

The snack bar or its equivalent in a series of automatic drink and sandwich dispensers is now a common sight in most factories. Employees like a chance to have a soft drink or coffee and munch a candy bar from an automatic dispenser during a rest pause. Here again, however, it must be remembered that the employees' attitude toward environmental change will determine whether such a change will be accepted or rejected. A rest pause spent in a canteen is apparently more favored than the regular rest pause. It promotes social activity.

The drinking fountain with its cool bubbling water sometimes provides the employees with an excusable rest period that, especially during warm weather, is relaxing and refreshing. Employees may resent the fact that drinking fountains may not be located near enough to the work area, but small capital investment often solves this problem.

The greatest bone of contention among many employees is the condition of the rest rooms. The amount of space devoted to the rest rooms, as well as the sanitary conditions, are important contributors to employee attitudes. When such facilities are inadequate, this may have a serious effect on job performance.

There are many other environmental aspects of a job that are related to the ultimate attitude of the employee and the morale of the group. Although few if any experiments have been made in this field, no claims as wild and weird have been made as those made in connection with color in industry.

In closing this chapter the reader is reminded of the Hawthorne Studies reported

in Chapter 10. The experiments on illumination suggested the danger of overlooking employee attitudes and morale in attempting to relate changes in production to different environmental conditions. In addition, these studies showed the relationship between production and the interpersonal relations of employees. The foreman, the supervisor, the boss, the "expert" all play an important role as individuals and help to determine the employees' perception of the change in the environmental situation. These people may contribute to an increase in production as much as a new environmental factor does. Change in work environment should be considered not as a separate entity but rather in relation to interpersonal relationships of employee and employer.

REFERENCES

Berrien, F. K. The effects of noise. *Psychological Bulletin*, 1946, 43, 141–161.

Berry, P. C. Effect of colored illumination upon perceived temperatures. *Journal of Applied Psychology*, 1961, 45, 248–250.

Culbert, S. S., and M. I. Posner. Human habituation to an acoustical energy distribution spectrum. *Journal of Applied Psychology*, 1960, 44, 263–266.

Dennis, J. P. Some effects of vibration on visual performance. *Journal of Applied Psychology*, 1965, 49, 245–252.

Duffy, Elizabeth. The concept of energy mobilization. *Psychological Review*, 1951, 58, 30–40.

Duffy, Elizabeth. *Activation and behavior*. Wiley, New York, 1962.

Ferree, C. E., and G. Rand. The effect of mixing artificial light with daylight on important functions of the eye. *Transactions of the Illuminating Engineering Society*, 1939, 21, 588–609.

Ferree, C. E., and G. Rand. The problems of late afternoon lighting. *Archives of Ophthalmology*, 1932, 7, 558–575.

Ferree, C. E., and G. Rand. Work and its illumination. *Personnel Journal*, 1940(a), 19, 55–64.

Ferree, C. E., and G. Rand. Work and its illumination. *Personnel Journal*, 1940(b), 19, 93–98.

Ford, A. Attention—auto-matization, an investigation of the transitional nature of mind. *American Journal of Psychology*, 1929, 41, 1–32.

Kerr, W. A. *Experiments on the effects of music on factory production*. Applied Psychological Monographs No. 5, 1945.

McBain, W. N. Noise, the "arousal hypothesis" and monotonous work. *Journal of Applied Psychology*, 1961, 45, 309–317.

McCormick, E. J. *Human factors engineering*. McGraw-Hill, New York, 1964.

McGehee, W., and J. E. Gardner. Music in a complex industrial job. *Personnel Psychology*, 1949, 2, 405–417.

Morgan, J. J. B. The overcoming of distraction and other resistances. *Archives of Psychology*, No. 35, 1916.

Park, J. F., Jr., and M. C. Payne, Jr. Effects of noise level and difficulty of task in performing division. *Journal of Applied Psychology*, 1963, 47, 367–368.

Scott, W. E., Jr. Activation theory and task design. *Organizational Behavior and Human Performance*, 1966, 1, 3–30.

Sleight, R. B., and J. Tiffin. Industrial noise and hearing. *Journal of Applied Psychology*, 1948, 32, 476–489.

Smith, H. C. *Music in relation to employee attitudes, piece work production and industrial accidents.* Applied Psychological Monographs No. 14, 1947.

Smith, Patricia C., and R. Curnow. "Arousal hypothesis" and the effects of music on purchasing behavior. *Journal of Applied Psychology,* 1966, *50,* 255–256.

Vernon, H. M., and C. G. Warner. Objective and subjective test for noise. *Personnel Journal,* 1932, *11,* 141–149.

HUMAN PERFORMANCE 20

In Chapters 18 and 19 we examined various aspects of job-related behavior such as fatigue, accidents, and the general influence of the work environment. Psychologists have historically been interested in specific aspects of task-oriented behavior such as these in the hope that a better understanding would result in a more "satisfied" worker and/or in increased work productivity. In this chapter we shall continue to examine topics of the same general nature, such as general principles of human performance, engineering psychology and time and motion study. These topics represent what might be classed as the nonmotivational aspects of human work behavior. We started to make the transition from motivational aspects of work in the preceding chapters as we discussed such environmental work variables as noise, vibration, etc., although music, which was also discussed in length, clearly has motivational aspects associated with it.

One topic in this chapter may seem somewhat misplaced to the reader—time and motion study. This topic has much in common with job analysis, and indeed it may be viewed as a particular method or philosophy of job analysis. The decision to include it here was that time and motion study, particularly during its heydey, represented something more than just a job analysis procedure; it represented a particular philosophy of how to study man at work with a clear-cut objective of *improving work proficiency* utilizing time and motion methodology. Growing primarily out of an industrial engineering background, time and motion study has been developed over the years as a particular approach designed to accomplish basically the same set of objectives as human engineering (or engineering psychology, a title currently preferred by many). We decided, therefore, to place together the topics of time and motion study, engineering psychology, and principles of human performance in order to give a more integrated view of the different approaches to studying the nonmotivational aspects of man's interaction with his job and how the area has steadily progressed in its sophistication of approach over the years.

TIME AND MOTION STUDY METHODOLOGY

Time and motion studies have as their objective the elimination of waste and inefficiency. This is achieved through reduction of costs, improvement of work

methods, and minimizing of work fatigue. Establishment of a basis for effective training and the determination of wage rates are also results of these studies. In time studies, the emphasis is on determining the standard time required to complete a task. In motion studies, the methods, motions, and movements of the worker are analyzed. Whereas these two studies were once treated separately, they now are often combined, for they are intimately related; rarely is there justification for performing one study without the other.

The stock-in-trade of the "scientific manager" or the "efficiency expert" is usually a variation on the theme of time and motion studies. This subject has for some years been a controversial one, with management glorifying it and labor condemning it. Both sides have been right and wrong at the same time. The main difficulty has resulted from the misapplication of time and motion studies. If, as a result of these studies, more efficient production were achieved through a rearrangement of tools, fatigue were decreased, and costs were reduced, along with a fair increase in the earning capacity of the employees, there could be no valid objection. To object to time and motion studies on the ground that they result in more efficient work is not only illogical but often purely emotional. Nevertheless, the objection has been heard that, too often, they are not concerned with a general increase in efficiency but are used as an excuse for a speed-up or for the dismissal of employees. This objection is valid, but it is not so much a criticism of the studies as of their misapplication. It is important to emphasize that time and motion studies need not be unfair. They can be useful in promoting the aims of both management and labor, provided they are carried out correctly and are not misapplied. Furthermore, when they lead to increased efficiency, they must result in some gains to the worker as well as to management.

POTENTIAL FAILURE OF TIME AND MOTION STUDIES

Despite the objectives of a time and motion study, and the "scientific" attempts to measure the time and motions involved in the performance of a task, the average study in industry is doomed to failure and is likely to be scrapped. This is a radical statement, but it is made in spite of the facts uncovered and the claims made; it also contradicts much of the literature in the field which has appeared since Frederick W. Taylor's work in the 1890s. Time and motion studies in and of themselves are not nearly the formidable bulwarks of efficiency that they appear to be on cursory examination.

The psychological components of time and motion studies are very important. Whenever these are overlooked—and they often are—these studies become ineffective tools in the hands of fools. Under these circumstances they do not produce greater efficiency; they merely produce greater dissatisfaction. The idea here is not that such studies are morally wrong, but that unless the problems that result from them are recognized, the studies can never really have a chance to accomplish anything of value.

The story of Johnny will illustrate the point. Johnny was the informal leader of a group of employees in a certain department of one of the Western Electric plants. He and his group were on a piece-rate basis and were earning more than the average

hourly rate for similar departments. This resulted in a series of visits from a man who made time and motion studies. Before going into the reasons why his studies failed in this instance, some background material is necessary. A description of Johnny and his coworkers is desirable for an understanding of the total situation.

In many respects they were just like any other group of factory workers, and in some respects "even more so." A college professor might consider them a rough and tough bunch. But they really were not. They were young, aggressive, and totally uninterested in education. They had two hobbies. All of them were amateur prize fighters, and they liked to train "on the job." Whenever the foreman was away they were likely to do a little sparring. Their enthusiasm was real, and every now and then someone would be "knocked out." Naturally, the presence of a body in a prone position caused some dismay, but the problem was usually solved by rolling the fellow under a worktable. None of them ever really got angry over this because the next bout might result in a reversal of the decision.

Their second hobby was trying to put something over on the special policeman stationed at the entrance of the factory. They regarded him as a person who distrusted them, and they accepted this challenge by seeing how successfully they could slip pieces of equipment past him. At first it was only parts, but later it was telephones. That they had no earthly use for the equipment was proved by their practice of bringing it back into the factory undetected.

Their work methods were also highly individualistic and aggressive. The telephones which they were to disassemble arrived on rather large hand trucks. Whenever such a truck left the elevator someone in the group would give a blood-curdling yell, "Breaks." This was the signal for all of them to stop work immediately and rush to the truck. In their language a "break" was a telephone partly disassembled at the time it reached them, which meant less work for them to do. All of them pushed and fought for these "breaks," cramming them under one arm and grabbing for the others with their free hand. The one who collected the greatest number of the telephones was the winner, and he strutted accordingly.

As indicated, this group consisted of a number of highly aggressive individuals who delighted in physical violence and rather violent horseplay. With this knowledge of the group, we can return to the man who made the time and motion studies.

In the company of Johnny and his cohorts, this man's identity was immediately recognizable. He wore a white starched shirt without a coat and carried a board with a clock inserted at the top. He would stand behind the worker with one foot on a stool and observe the employee on the job.

In the case of Johnny's group the time and motion study man made a series of trips. Before long he observed only Johnny, who incidentally was the best worker. Johnny, however, always slowed up when he appeared. When the man accused him of this, Johnny always answered, "If you can do the job better, why don't you do it?" Since the time and motion study man could not do the job, he had to be content with merely observing and making notes. After a period of time, the rates were changed. Johnny's group objected to this, but at the same time saw to it that the change did not affect their total earnings. Their goal was to earn about $10 a week more than the prevailing hourly rate, and they did. At first they did not fool around so much, but the more stringent control of rates led them to change their entire work method. (This was unknown to the company, or at least to the time and motion study man.) Within three months they became a highly cooperative group. When the telephones arrived on the floor, there was no rushing to the truck.

One man rolled the truck to his station and began to remove the base of the phone. Then he passed the phone to the man who did the next step in disassembling, and then it went to the third and fourth men until, in their language, it was stripped. The fifth man served as a substitute and was also responsible for watching for an unexpected visit from a supervisor or a time and motion study man.

This story illustrates how these workers, by their own efforts, overcame the wage rate set by the time and motion study man. They had to increase their production but they maintained their earnings. Dissatisfaction also increased in the group, as evidenced by their vocal disapproval of the company and its policy.

Another illustration along the same line is a story told by Mr. Ruttenberg, formerly of the CIO. After stating that when standards are set exclusively by management without the men's participation the company does not obtain anywhere near top production, the president of one of the largest companies in Pittsburgh challenged him to prove his point. In accepting this challenge, Ruttenberg requested the company engineer to tell him which job he thought most ideal for the test, i.e., the job with the greatest production. He thereupon offered to double this production in a month. Ruttenberg talked with the men on the job and assured them that a production rise would not affect their rates. In a month they had achieved 210 percent production.

Most time and motion studies, however, do not constitute a challenge to employees, as in these two instances; rather, they result in what might be interpreted as a concerted effort on the part of the employees to prove that the new method is no good or that the new gadget is a waste of time. And they resist the change. The resistance is psychological in nature, but it must be reckoned with if time and motion studies are to lead to greater efficiency without a decrease in employee satisfaction.

RESISTANCE TO TIME AND MOTION STUDIES

There are three major forms of resistance to time and motion studies: (1) objection to change, (2) the fact that the change is initiated by an outsider, and (3) the worker's increased feeling of insecurity. People are not capable of adjusting to rapid change. This applies to changes in doing a job as well as to changes in social codes.

We must recognize the serious implications of this resistance to change and newness. Time and motion studies often point to the fact that something new is needed in the job. The new thing may lead to improvement, but when it is introduced no improvement will be apparent because the worker will resist it, not because it is a better way of doing the job but because it is a "new" way.

The second objection to a time and motion study is the fact that an outsider initiates the change. Practically every worker does his job in a certain way and has his reasons for doing it that way. He resents being told by someone that he is not doing it efficiently. He may suspect that the outsider is there to "show him up," and he naturally will do all that is humanly possible to prevent it. The best way to defeat the purpose of a change is to have an "outsider" propose the change. Whether it is a worker on his job or a community that has the problem, the importation of experts or "outsiders" either retards or halts progress in the direction of the desired change. Thus Johnny's group decided at all costs to show up the time and motion study man, regardless of the change in their rate.

The "efficiency expert" who comes into an office and sets up a new filing system or installs a centralized stenographic pool may have the right idea, but the workers will not believe it and they will manage to adopt as little of the change as is possible. Furthermore, when supervision is relaxed they are likely to go back to their old method, not only because they dislike the new ways but also because they still resent the intrusion of an outsider.

Increased feelings of insecurity, the third resistance to time and motion study, is in part psychological and in part economic. Too often in the history of time and motion studies the results have been used to speed up production and lower wage rates. Even though a specific employee may not have been involved, he has heard about such cases and regards these results as an end product of such study. Insecurity increases, and he often fears that he will be dismissed from his job. Hence he does all he can to avoid having his pay lowered or being fired. His usual defense is an attempt to show that the new method will not work.

MEN IN MOTION

The father of "scientific" management was Frederick Winslow Taylor, a mechanical engineer who performed management studies in 1890. According to him (1947) "scientific" management (the quotes are ours) is based upon the firm conviction that the interests of the employer and the employee are the same, and that this system makes it possible to give the workman what he wants most—high wages—and the employer what he wants—a low labor cost. Taylor proposed scientific man-

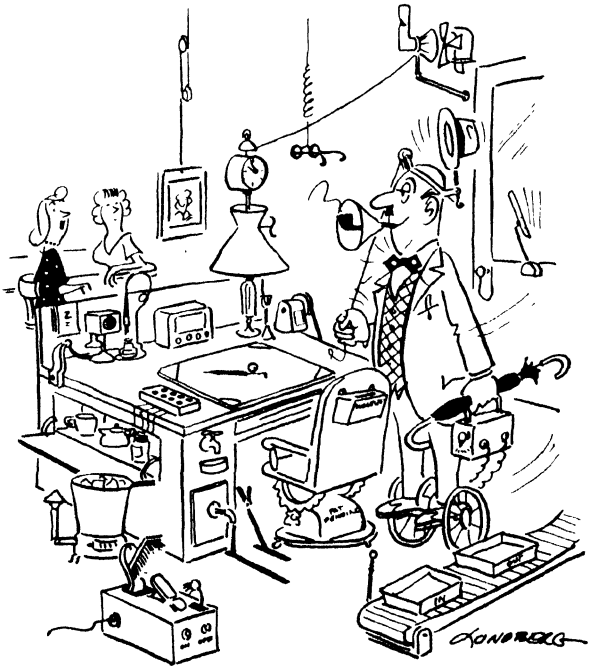


Figure 20.1. "He almost saved a minute this morning." (From *Pic*, December 1948. By permission of Street & Smith Publications, Inc., and the cartoonist, Gustav Lundberg.)

agement as opposed to "management of initiative and incentive." He believed that his system was scientific because it gathered together the knowledge formerly possessed only by workmen and classified, tabulated, and reduced it to laws, rules, and formulas. He proposed that management take on new duties, such as (1) developing a science for each element of a man's work, (2) scientifically selecting and training workmen, (3) cooperating heartily with the men, and (4) taking over all duties and work for which it is better fitted than the workmen. Taylor states in summary, "Under the management of initiative and incentive practically the whole problem is up to the workmen, while under scientific management fully one-half the problem is under management." To be perfectly blunt, either Taylor must have been a supersalesman or industry in 1890 was very primitive and its executives were very naive. Analyzing Taylor's system and the manner in which it was put to use allows no justification for the term "scientific."

The Taylor study which has achieved historical importance is the "Pig-Iron Handler," a study of a gang of men loading pigs weighing 92 pounds and loading an average of 12½ tons per day. Taylor's lack of concern and sympathy with workmen is revealed in his statement, "This work is so crude and elementary in its nature that the writer firmly believes that it would be possible to train an intelligent gorilla, so as to become a more efficient pig-iron handler than any man can be." His lack of sympathy and respect is further reflected in the way he reports his handling of a worker named Schmidt (1947, pp. 44 ff.)[1]:

"Schmidt, are you a high-priced man?"

"Vell, I don't know vat you mean."

"Oh yes, you do. What I want to know is whether you are a high-priced man or not."

"Vell, I don't know vat you mean."

"Oh, come now, you answer my questions. What I want to find out is whether you are a high-priced man or one of these cheap fellows here. What I want to find out is whether you want to earn $1.85 a day or whether you are satisfied with $1.15, just the same as those cheap fellows are getting."

"Did I vant $1.85 a day? Vas dot a high-priced man? Vell, yes I vas a high-priced man."

"Oh, you're aggravating me. Of course you want $1.85 a day—every one wants it! You know perfectly well that has very little to do with your being a high-priced man. For goodness' sake, answer my questions, and don't waste any more of my time. Now come over here. You see that pile of pig iron?"

"Yes."

"You see that car?"

"Yes."

"Well, if you are a high-priced man, you will load that pig iron on that car tomorrow for $1.85. Now, do wake up and answer my question. Tell me whether you are a high-priced man or not."

"Vell—did I got $1.85 for loading dot pig iron on dot car tomorrow?"

"Yes, of course you do, and you get $1.85 for loading a pile like that every day right through the year. That is what a high-priced man does, and you know it just as well as I do."

"Vell, dot's all right. I could load dot pig iron on the car tomorrow for $1.85, and I get it every day, don't I?"

"Certainly you do—certainly you do."

"Vell den, I vas a high-priced man."

"Now, hold on, hold on. You know just as well as I do that a high-priced man has to do exactly as he's told from morning till night. You have seen this man here before, haven't you?"

[1] Copyright 1947 by Harper & Row Publishers, Incorporated. Reprinted by permission of the publishers.

"No, I never saw him."

"Well, if you are a high-priced man, you will do exactly as this man tells you tomorrow, from morning till night. When he tells you to pick up a pig and walk, you pick it up and you walk, and when he tells you to sit down and rest, you sit down. You do that right straight through the day. And what's more, no back talk. Now a high-priced man does just what he's told to do, and no back talk. Do you understand that? When this man tells you to walk, you walk; when he tells you to sit down, you sit down, and you don't talk back at him. Now you come on to work here tomorrow morning and I'll know before night whether you are really a high-priced man or not."

By introducing rest periods and closely supervising the job, Taylor increased Schmidt's production (provided it was average) from 12½ to 47½ tons per day, an increase of 280 percent. Schmidt's salary was increased from $1.15 per day to $1.85, per day, or 61 percent. The present writer makes no claim that Schmidt should have received a 280 percent salary increase, but it is apparent that the salary increase was disproportionate to the increase in production. It is entirely possible that a problem which still confronts America's mass-production industry originated with the Schmidt episode.

Another aspect of Taylor's work involves scientific selection of workmen. He glibly stated that only one man in eight can produce 47½ tons per day. This meant job changes for seven out of every eight men on a gang. There is a great likelihood that his norm for production was too high and that his selection was not scientific.

To understand Taylor, it is necessary to refer to his early work at Midvale Steel (1947, pp. 49–50).[2]

As soon as the writer was made gang-boss, one after another of the men came to him and talked somewhat as follows:

"Now, Fred, we're very glad to see that you've been made gang-boss. You know the game all right, and we're sure that you're not likely to be a piece-work hog. You come along with us, and everything will be all right, but if you try breaking any of these rates you can be mighty sure that we'll throw you over the fence."

The writer told them plainly that he was now working on the side of the management and that he proposed to do whatever he could to get a fair day's work out of the lathes. This immediately started a war; in most cases a friendly war, because the men who were under him were his personal friends, but none the less a war, which as time went on grew more and more bitter. The writer used every expedient to make them do a fair day's work, such as discharging or lowering the wages of the more stubborn men who refused to make any improvement, and such as lowering the piece-work price, hiring green men, and personally teaching them how to do the work, with the promise from them that when they had learned how, they would then do a fair day's work. While the men constantly brought such pressure to bear (both inside and outside the works) upon all those who started to increase their output that they were finally compelled to do about as the rest did, or else quit. No one who has not had this experience can have an idea of the bitterness which is gradually developed in such a struggle. In a war of this kind the workmen have one expedient which is usually effective. They use their ingenuity to contrive various ways in which the machines which they are running are broken or damaged—apparently by accident, or in the regular course of work—and this they always lay at the door of the foreman, who has forced them to drive the machine so hard that it is overstrained and is being ruined. And there are few foremen indeed who are able to stand up against the combined pressure of all of the men in the shop. In this case the problem was complicated by the fact that the shop ran both day and night.

The writer had two advantages, however, which are not possessed by the ordinary

[2] Copyright 1947 by Harper & Row Publishers, Incorporated. Reprinted by permission of the publishers.

foreman, and these came, curiously enough, from the fact that he was not the son of a working man.

Referring to the third item in Taylor's system, cooperation with the men, cooperation for him cannot be interpreted in the sense that cooperation was used in the chapter on morale. Taylor attributed his success as a gang boss to two facts: (1) he was not the son of a workingman and management would therefore believe him sooner than an ordinary worker, and (2) he was different and lived apart from the other workmen. From this point of view he cannot possibly be talking about true cooperation.

Taylor also believed that increases of more than 60 percent in the uniform wages usually paid would make the men shiftless, extravagant, and dissipated. He reported this after referring to a long series of experiments and observations but he did not mention the type of experiment; it does not appear that his observations were unbiased. Since his time, wages have gone up considerably more than 60 percent and workmen have not become "shiftless, extravagant, and dissipated." Taylor represents those who believe that the average workman is dull and has no interests except earning more money, but that he is so stupid that earning too much is bad for him; therefore management is protecting him by keeping his salary increases small.

A world of difference exists between Taylor's views and the present-day views of many industrial psychologists and industrial engineers. Some of these differences can be attributed to the work which has been done by Frank B. Gilbreth, an engineer, and his wife, Lillian, who was trained in psychology.

The first work done by Frank Gilbreth in this field was in connection with bricklaying (1911). His first study involved much more than the introduction of rest periods and constant supervision of the worker on the job. It was a minute study of the motions involved in bricklaying and it enabled him to reduce the motions from 18 to 5, thereby increasing production from 120 to 350 bricks per man-hour.

The two Gilbreths worked together very closely and the results of their work demonstrate the advantages in having an engineer and a psychologist work in cooperation. From their training, the engineer and the psychologist acquire a great deal of complementary information. Alone they can make serious mistakes. Together they can make real progress. They can act as a check on each other so that the human element does not cause neglect of the mechanical, and vice versa. It is indeed too bad that such teams are not more frequent.

The Gilbreths devised a system of efficiency that included nine principles: (1) individualization, (2) functionalization, (3) measurement, (4) analysis and synthesis, (5) standardization, (6) records and programs, (7) teaching, (8) incentives, (9) welfare.

Their system was all-inclusive and in many respects valid, not only in the production increases obtained but also in the recognition that the individual workman was the unit to be measured. The Gilbreths believed that the details of the work situation should be adjusted to the individual rather than the individual being forced into the pattern of the job. Their writings (1916, 1917) show much more insight and understanding of the individual than Taylor's work, and there is little evidence of the contempt for the workman which Taylor seems to have had.

One of the most interesting of the Gilbreths' contributions was the analysis and breakdown of a task into its basic elements of motion, which they called "therbligs." This word, invented by Frank Gilbreth as the result of a suggestion from his wife,

can in some respects be considered as a forerunner of Serutan. Reading therblig backward gives the name Gilbreth except that the *t* and *h* are reversed. The therbligs are very useful; each has a name and a symbol, as shown in Figure 20.2.

Every job involves a number of these therbligs. By way of illustration, consider signing one's name. A man usually keeps his fountain pen in his inside coat pocket. In therblig terms he must search, find, select (assuming he has more than one pen),

Symbol	Name of symbol	Symbol	Name of symbol
	Search	#	Disassemble
	Find		Inspect
	Select		Pre-position
	Grasp		Release load
	Transport loaded		Transport empty
	Position		Rest for overcoming fatigue
#	Assemble		Unavoidable delay
	Use		Avoidable delay
			Plan

Figure 20.2. Gilbreth's therbligs. (From L. P. Alford, ed. *Management's handbook*. Ronald, New York, 1924.)

transport loaded, position, disassemble, position, use, assemble, transfer, search, find, position. A lot of work for such a simple task that is done in so short a time.

After all the therbligs involved in the task have been identified, it is then necessary to ask six questions:

1. Is each therblig necessary?
2. Can the task be made simpler by having fewer motions?
3. Can there be less motion in performance or degree?
4. Can the steps be combined?
5. Can the sequence be changed?
6. Can more than one be done at the same time?

To go back to signing one's name, it is immediately obvious that a more efficient system would use a desk-set pen in a fixed position; this requires no assembling and disassembling and no selecting, and many of the other therbligs are not required.

Another simple illustration is the number of steps involved in inserting letters in envelopes and addressing and stamping them. A good exercise at this point would be to plan an efficient system. Remember that the introduction of equipment and machinery is not always desirable. An automatic stamping machine and addressograph might cost too much in relation to the average amount of mailing that is done by the typical small organization.

A good perspective on what has happened in the field of time and motion study since the early 1900s will be gained from Schumard's A *Primer of Time Study*

(1940). He writes, "It is believed that the make up of a good time-study man is comprised of a percentage of 80%–20%. The 80% can be called contact, the other 20% can be called education and common sense." He then lists 24 inherent and acquired qualities of such men. Although Schumard is serious, he shows the nonsense that an industrial engineer can indulge in when he tries to play the role of a psychologist. So, by way of illustrating what *not* to do, his list is presented in Table 20.1.

This list is open to criticism because these 24 traits do not define a time study man any more than they define a doctor, a lawyer, an Indian chief—or even a psychologist. Probably any self-respecting person in any self-respecting job believes that these 24 traits are necessary in his specific job, too, and they probably are. Open-mindedness, power of sympathy, personality, courage, etc., are all desirable traits.

TABLE 20.1 *Twenty-Four Inherent and Acquired Qualities of Time-Study Men*

1. Honesty	13. Self-confidence
2. Personality	14. Cooperation
3. Imagination	15. Sense of responsibility
4. Sense of fairness	16. Observation
5. Open-mindedness	17. Analysis
6. Power of sympathy	18. Judgment
7. Tact	19. Accuracy
8. Resourcefulness	20. Planning ability
9. Reliability	21. Power of instruction
10. Self-control	22. Optimism
11. Energy	23. Salesmanship
12. Proper conduct	24. Leadership

SOURCE: F. W. Schumard. *A primer of time study*. McGraw-Hill, New York, copyright 1940. Reprinted by permission.

But it is safe to assume that Schumard did not reach his conclusions as the result of tests and measures. The remainder of his book is a sound exposition of the slide rule, stopwatch, wage scale, rest factor, and other items. Although it does give tips on how to make a time study, application of the method would run into the same difficulties that doom time and motion studies to failure. It does not take into consideration the most important item, the resistance factor, which is invariably encountered.

Mogensen (1932) has a much more practical and valuable approach—the education of foremen and employees along lines that will encourage their becoming "motion-minded." He does not believe that the average worker is dull witted, but he admits that he sometimes lacks enthusiasm. Mogensen encourages motion-mindedness in a plant by showing "before and after" motion pictures of work done with time and motion studies in other plants. He then trains the foreman and executives in some of the specifics of time and motion study; it is the executives' job to convince the workers that they will not be subjected to speed-ups or dismissal. In this manner the resentment that a time and motion study engineer usually encounters is avoided. It is to be emphasized that Mogensen's approach tends to overcome the three serious difficulties in time and motion studies. It avoids both the resistance to an outsider's initiating change and the insecurity-producing factors of

time and motion studies, and it encourages an acceptance of change, since the workers themselves participate in and contribute to the change. The foreman and employees become involved in many of the intricacies of time and motion study. They construct process charts and make suggestions as to how to achieve greater efficiency.

As a consultant in this field, Mogensen has had considerable success with his method. It is exceedingly interesting to find William Gomberg of the International Ladies Garment Workers Union (ILGWU) writing of the "evangelical activities of Allen Mogensen on behalf of the work-simplification movement."

It is a mistake to believe that time and motion study can be a tool of efficiency only for management. Through its unions, labor can, and in rare cases does, conduct time studies which result in benefits to its members. An outstanding illustration of this is the work that has been done by the ILGWU under Gomberg's direction. The management engineering department of the ILGWU was created to pursue two objectives (Gomberg, 1943):

1. To assist in improving the manufacturing techniques and operating methods of all branches of the industry with which our workers' earnings are intimately bound. This will be done through plant inspections by department representatives, followed by specific recommendations.
2. To serve as a central information agency:
   a. To determine the level of "fair piece rates."
   b. To record the production system and manufacturing techniques under which these rates are paid.
   c. To assist in training shop members and committees in distinguishing bad time-study practices from good time-study practices in the determination of rates.

The union's primary concern, as reflected in its practices and behavior, is to use time studies for obtaining data so that it may bargain collectively with employers in establishing salary standards and piece rates. Its attitude toward these studies is generally thoughtful and critical. Gomberg regards them as, at best, an empirical guide to setting up a range within which collective bargaining on production rates can take place. He is very critical of the stopwatch as a measuring device when one is striving for a reasonable range in accuracy of measurement, and he has data to prove the lack of reliability of such measurements. He questions that there is a single standardized method of performing a job and considers modification of method from worker to worker both desirable and efficient. Gomberg also recognizes a third limitation of time studies (1946). "It is at once apparent that nothing has been developed in industrial time study practice that can be considered an objective measure of normality or an objective method for comparing operator performance with any normal standard."

The work of the ILGWU management engineering department was primarily concerned with time rather than motion study. The department was interested mostly in the setting of piece rates and considers most motion studies to be primarily the concern of management. However, on invitation from management, it did attempt to increase production in specific plants, and in these cases it was concerned not only with the setting of a fair piece rate but also with motion study.

Gomberg does not deny the usefulness of time studies, provided their limitations in accuracy, both as a measuring device and as a means of setting standards, are recognized. He believes that no alternative method of measuring or estimating a

reasonable day's work has as yet appeared and that therefore time study methods and procedures must be called upon until a better technique is found.

One additional feature of this union's approach is the training of its members on the various shop committees in time study techniques. Sometimes these trained employees are accused by their fellow union members of not representing the workers' best interests, and yet they are backed by the union. This means that the technique, when applied fairly, allows for a reasonable settlement of claims rather than resulting in a one-sided and emotional view. Mogensen and this union both have faith in the training of workers in time and motion study techniques. Regardless

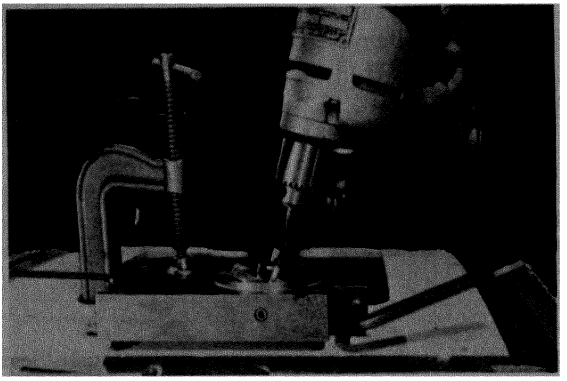


Figure 20.3. Ball-and-socket drill block developed by an employee. It permits more accurate use of portable electric drills at different angles. (Courtesy Temco Aircraft Corporation.)

of one's opinion as to whether management alone or management in collaboration with labor should set wage standards, there is no doubt but that employees can contribute to their own efficiency at least as much as the expert can, and in some cases more. They need only training, an opportunity for expression, and confidence that their improvements will be adopted.

SOME SIMPLE PRINCIPLES

Time and motion studies are not extremely technical or intricate; they are relatively simple to conduct provided one recognizes a few basic principles.

A very useful principle is that used in determining the most favored work area, sometimes known as the "semicircular work pit." Regardless of whether the employee is doing office work at a desk or assembling at a workbench, this principle can be applied. By circumscribing an arc with your right hand when your elbow is at the hip, and then with your left hand with elbow at hip, you will define a semicircle in which you can work with a minimum of moving and reaching. Then, keeping your back erect, swing your right arm to the left, and your left arm to the right, and you will define two other semicircles. These two overlapping areas determine a second, and larger, semicircle. This is the area wherein you can reach things with the minimum of stretching and body movement. Most work can be done in the first semicircle; but if necessary, it may be extended to the second one. This semicircular work pit is the most favored work area from the point of view of economy of motion and minimum effort.

The prepositioning of tools, another simple principle, is a useful technique in time and motion studies. It avoids much unnecessary delay in looking for tools which have been misplaced. Prepositioning operates effectively for a bank teller, a shipping clerk, a watch assembler, and many other workers.

The prolonged holding of an object is poor motion practice. The introduction of a jig or a vise to clamp the work will free the holding hand. Most efficiency concepts encourage the use of two hands. Motion of the two hands is most efficient when they are used in opposite and symmetrical direction. Make slight circular movements with both hands, moving them both to the right as if you were tightening a light bulb in its socket. Then do it again, first moving both hands inward and then both outward. It makes a difference.

The introduction of rhythm in work is also an important principle. Experienced carpenters, typists, and painters have rhythm in their work. In fact, a simple but crude manner of judging work performance and ability is to observe the worker on the job and watch for regular, smooth, and even movements. These encourage rhythm and indicate efficiency. The unsure, erratic, staccato movements of the novice can be detected immediately.

A drop delivery will prevent the cluttering of the work space and the waste of time involved in transferring the finished product. A hole at the work place leading to a chute which carries the finished product away makes for efficiency and economy. Labor savings of 5 to 10 percent have often been accomplished by introducing a drop delivery system.

The redesigning of tools and equipment has been an accepted principle ever since Taylor redesigned the shovel. The amateur gardener or farmer would find the new rake shown in Figure 20.4 a welcome improvement.

The process chart is a desirable and not very technical device. A process chart records exactly what goes on. It may be a *flow chart*, which records the flow of work



"And with this new, improved handle, you can step on it without injury"

Figure 20.4. Redesign of a rake. (By permission of *American Magazine*.)

during the process. Or it may be a *man-and-machine chart,* which records the work of the machine in relation to the work of the man. A third type is the *operator chart,* which is mainly concerned with the separate work done by the left hand and the right hand. The *cost chart,* which establishes the cost of the operation, is a process chart that brings new happiness to the accountants in our society.

Another principle in time and motion study is the reduction of motion. Motions involving only the fingers are considered most efficient because they require least muscular effort. The least efficient motion is one which requires the use of the entire body. It is sometimes claimed that an attempt to reduce motions to such simple levels introduces monotony and related phenomena. Although in some instances this may be true from the point of view of efficiency, it must be conceded that the motion that involves the fewest muscles is the least fatiguing to the individual.

NEED FOR ALLOWANCES IN TIME AND MOTION STUDY

After a time and motion study has been made, but before it is put into use as a standard of production or a basis for setting a wage rate, certain time allowances must be made for it. The four most commonly recognized allowances are (1) differences between the skill and effort of the worker on whom the study was based and the skill and effort of the group performing the task, (2) personal needs, (3) fatigue, (4) unavoidable delays. The first three are related to the human factor, the worker. The fourth is related to production and is largely independent of the worker. Because all machines occasionally break down and need repair, or material supplies may not be available, an allowance is made in connection with work which is determined solely by the motions and speed of a machine. The other three allowances must be made because of individual variability.

A psychologist recognizes the wide variations in human behavior, and when he develops norms he tries first to obtain data on large numbers of subjects. Most standards established as a result of time and motion studies are based upon the production of one, two, or possibly three people, but rarely more. Since the standard thus established will be applied to large numbers of people, the danger in drawing conclusions on the basis of such a limited sample is great. In fact, the weakness of the entire structure of time and motion study may be due to the fact that the sampling is so limited as to give invalid results. The time and motion study people partially recognize this difficulty and attempt to cope with it by making allowances part of their standard procedure. They realize than an individual's skill on the job will vary, as will his effort to produce. People with great skill may put forth little effort, and people with very little skill may exert great effort. The two other combinations—high skill and great effort, and low skill and little effort—are also possible.

In setting a standard, the time and motion study engineer basically and ultimately must use his own judgment as to the degree of skill and effort of the particular worker on whom he is basing the production standard in comparison with that of the other workers. This is a considerable responsibility and demands both skill and effort on his part. His judgment will be reasonable only to the extent to which he is concerned with this problem. By the same token, the extent to which he blindly follows the fashionable inefficient mathematical formula of the moment

will determine the extent to which errors are possible. The Bedaux or speed-rating system attempts to solve the skill and effort problem. The writers are not fully convinced of the validity of the claims made for the Bedaux system despite its wide use in the past.

Factors which should be included in the allowances for personal needs are the employee's visits to the water cooler, locker room, rest room or lavatory, etc.

The allowance for fatigue, the third human factor, rests ultimately on sheer estimate. The problem of fatigue was discussed in Chapter 18; here we shall say only that susceptibility to fatigue is relative to the individual and that, in many instances, what is called fatigue is not fatigue at all.

The unavoidable-delay allowance is more easily reckoned. A time study expert can take up his post near a machine and record, over a day, a week, or several months, the time consumed by machine repair or by shortages of parts which prevent the employee from doing his work on the machine.

In a survey on the problem of time allowances in 360 plants in the Pittsburgh area, Blair (1941) found that 106 of these plants used a stop watch or the time study method to set production standards. The remaining 254 set their standards on the basis of the judgment and experience of either a foreman, a pace-setter, or a worker who was assumed to be an average performer. In some instances the standard set was based on the speed recommended by the manufacturer of the machine.

Most of the plants studied by Blair did not use a really objective basis for determining allowances, regardless of how they set the production standard. Extra time allowances were determined essentially by an overall estimate of all four factors combined: skill-effort, personal needs, fatigue, and unavoidable delay. The general practice in the other plants was usually to overlook "allowances" completely.

A little-known book called *How to Run a Bassoon Factory* (Spade, 1936), a brilliant and delightfully humorous satire on American efficiency systems, presents the problem of allowances as follows:

> He [the time study man] proceeds something like this: (1) He sets a man to work on making a bassoon and times him to one hundredth of a second. This gives him a figure to work on; (2) He then takes his figure and applies the following reasoning: (a) It took old Sam two days to make that bassoon, but he was not working as hard as he might, therefore, divide by two equals, one day; (b) Anyhow old Sam is old and not representative, therefore divide by two again equals one half day; (c) on the other hand, old Sam has had a lot of experience, therefore multiply by two equals one day; (d) but there were seventy-nine periods of two minutes where he did not do anything, therefore subtract one hundred fifty-eight minutes, equals five and a half hours; (e) But he will get tired in the day so we'll allow him one half hour, therefore, add one half hour equals six hours; (f) On the other hand if the rate is to be tight old Sam will kick up a row therefore multiply by two equals one and a half days; (g) Anyhow, we always leave a bit of margin, so let's call it two days. This is now made into a formula. The numerator is old Sam's time, times two, times two. The denominator is two times two equals (158—30—33⅓) and by taking old Sam's time and applying this formula we get the scientific determined time of two days. The time study man then works out what it usually pays old Sam, for work, and there you are. This, however, is piece rate in its simplest form. If you particularly wish to acknowledge the worker, it is usual to say that you will pay him seven pence a bassoon for the first ninety bassoons and with each subsequent bassoon 6/90 of a penny. Thus the more he does the more he gets paid but not so much, if you see what I mean, then

> he strikes and you arbitrate and begin again. The majority of strikes about wages are the result of the worker being unable to appreciate an ordinary logical piece of reasoning like this.

Of course the pseudonym-endowed author, Mark Spade, is spoofing; but to anyone with a sense of humor some of the rationalizations made in connection with extra time allowances will appear to be little better. These allowances are a serious pitfall in time and motion studies; they should not be overlooked or speedily estimated. They should be based on actual data from a large sample of workers and not from just one or two employees. Estimates of the average worker made by foremen, pace-setters, or experts do not provide a sound or scientific basis for determining extra time allowances in time and motion studies.

Lifson (1953) conducted an experiment which raises further questions on the accuracy of judgments in time study. He had six expert time study men make ratings of the filmed performance of five workers doing each of four jobs at each of five previously established paces. He found that the experts had different concepts of normal pace. Further, there was a marked trend to underrate fast paces and overrate slow paces. In fact, lack of agreement was found among experts' ratings on both job performance and the job.

The objective evidence that experts differ in their judgments is not new to psychologists. The fact that it applies to time study experts means that the general practice of having a single rater determine standards may lead to gross errors in "standards" and pay rates. Having more than one time study expert do the rating and then combining their judgments might well lead to greater objectivity.

Pace rating or leveling involves considerable error and experts differ in the extent to which they err. The total error is a result of the time study man's concept of normal work pace as well as his rating of each worker (and this varies) and each job (this also varies).

The time study should be regarded as a technique which may not be objective and factual. It should be rigidly inspected as to methodology and statistics before its conclusions warrant acceptance as standard. Perhaps Gomberg is reasonably correct: It provides a basis for bargaining.

TIME AND MOTION STUDY IN THE HOME

It must be emphasized that time and motion studies are not merely a concern of big business and big unions. The application of these studies has also resulted in increased efficiency in the home. Marvin E. Mundel and Janet Armstrong (1946) have done much work in applying the findings of these studies in the home. They recommend six rules of efficiency:

1. Eliminate all unnecessary parts of a job. For example, in dishwashing, use very hot water and let the dishes drain dry; do not wipe them.
2. Keep everything within easy reach; for example, the coffee pot, coffee can, and measuring spoon should be stored close together in a cabinet.
3. Use the best tool for the job. For instance, use a step-on can with a removable inner sack for garbage.

4. Use both hands for such work as putting away dishes or stacking fresh linens.
5. Combine two jobs into one; for example, fold pillowcases with the free hand as they are ironed.
6. Whenever possible, sit down to work—when preparing vegetables or feeding the children, etc.

These six rules for efficiency in the home are similar to the rules regarding efficiency in industry. Housewives and their husbands are not likely to object to new methods which will make housework less of a burden. By the same token, employees are not likely to object to efficiency methods which make their jobs less of a burden, providing they are not subjected to a speed-up or made to feel their position is insecure. Any small office, shop, or factory can improve its efficiency tremendously by applying the principles recommended here.

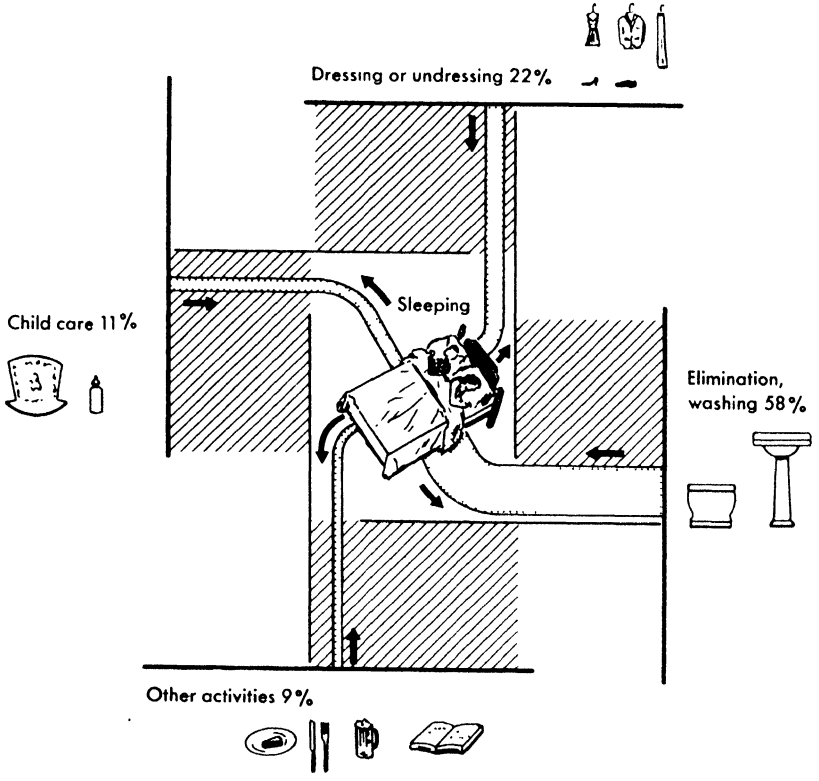


Figure 20.5. Activities related in a time sequence to sleeping. Such study can result in more effective design of space surrounding the bedroom. (From Milton L. Blum and Beatrice Candee. *Family behavior, attitudes and possessions.* The John B. Pierce Foundation, New York, January 1944, vol. 4, 203.)

EXAMPLES OF CONTINUED INEFFICIENCIES

Two popular and widely used machines in the present era are notoriously inefficient—the typewriter and the sewing machine. The typewriter now in use makes it

necessary for the left hand to do 57 percent of the work. August Dvorak has rearranged the keyboard so that the right hand does 56 percent of the work. This keyboard makes it possible for trained typists to do between 165 and 180 words a minute. Its arrangement is as follows:

7 5 3 1 9 0 2 4 6 8
? , . P Y F G C R L /
A O E U I D H T N S –
, Q J K X B M W V Z

Another widely used inefficient machine is the sewing machine. The present electric machine requires that the left hand, which is ordinarily less agile, do more work than the right. This is the case not because the machine was designed by a left-handed person, but because originally the right hand did the important work of starting and stopping the machine as well as regulating the wheel. In the electric sewing machines the right hand is relieved of this task. The present position of the needle head at the left does not allow the right hand to feed conveniently the material to be sewed. Turning the machine around would probably result in greater accuracy and speed of operation. However, this will probably not be done, nor will the Dvorak typewriter keyboard gain immediate popular acceptance. This can be understood, primarily, in terms of resistance to change. It must, however, be added that although the use of the present typewriter and sewing machine may be irrational in terms of ultimate efficiency, it is rational to the individual employer who has a capital investment in these machines. It is also rational to individual operators who can use the present machines but would have to learn to operate a new, even though more efficient, machine.

BREAKING RESISTANCE

One of the simplest aids in breaking a person's resistance to change is to ask him to write his full name, leaving out every other letter. After the signal "Go," time him accurately with a stop watch; then without any comment, ask him to write his full name in his usual manner, and again time him. Writing the name using only half the letters takes just as long as writing the full name, and sometimes longer. Tell him the length of time it took to do both tasks. Of course, he will object. Usually people say, "But I am used to writing my full name, and writing just every other letter required effort!" You then counter with, "Yes, you are used to doing the job in one way, but won't you admit that a short amount of practice will probably result in your signing your name in approximately half the time?" Most people will accept this, and you continue, "Then you may admit that doing your job this new way will, with practice, eventually require less effort and time."

Another simple means which can be used to break resistance to change is to ask a male employee to describe in detail how he puts on his suit coat. Then ask him to put on his jacket and call his attention to the discrepancy (there usually is one) between his description of an often-performed task and the actual performance. Time him while he is putting the coat on. Then ask him to take it off and put it

on again, but this time have him start with his right hand rather than his left. Many people experience considerable difficulty with the left-hand right-hand reversal, but it serves to drive home the point that new tasks or operations are sometimes judged more difficult solely because they are new, and not because they are more difficult.

HUMAN ENGINEERING

The increase in mechanization has led to a shift in emphasis from studying the time and motions involved in the performance of a job to an awareness that man and machine must function in relation to each other. Engineers design machines and are justifiably concerned with effectiveness and efficiency. However, principles of physics and engineering are not enough as long as man is a necessary accompaniment to the workings of a machine. A machine, no matter how complex and automatic, requires a man to make judgments in connection with its continued performance (Figure 20.6). Whether he serves as an operator or as a maintainer of a machine, he reacts to signals from the machine in accordance with his abilities, characteristics, and limitations. Because of this, the psychological aspects of man-machine systems cannot be avoided. Further, the transmitting of signals from the

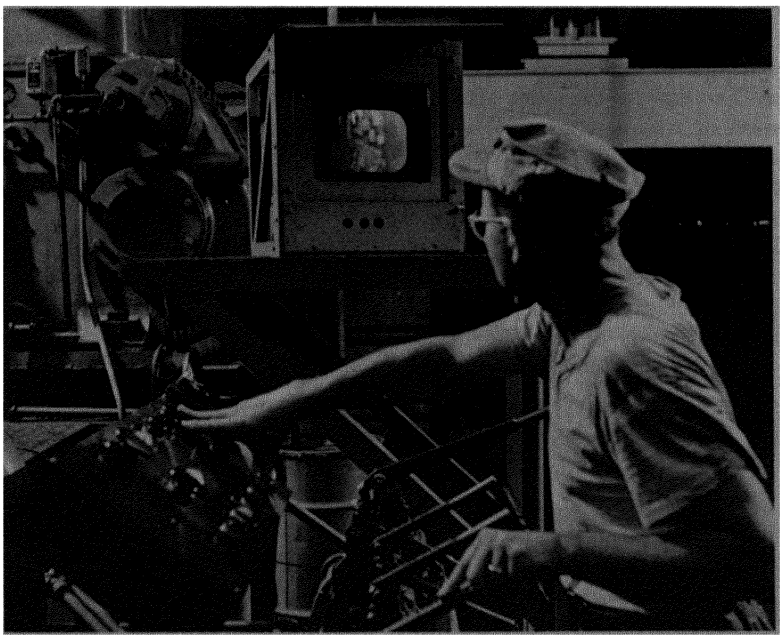


Figure 20.6. This employee watches industrial TV. His job is to operate scrap balers and to see that bales fall into a gondola car spotted outside the building. He watches the gondola via television. When the car is full, he can, by remote control, move it out of the way and move in an empty car. (From *Fortune*, September 1951. Photo by Jerry Cooke.)

machine is through dials, scales, pointers, or sounds. Sometimes speed is important; sometimes accuracy is important. Usually both speed and accuracy are desirable. Accordingly the size, shape, position, pointer, lettering, or illumination of the signal becomes a factor in man's ability to respond more accurately or more rapidly.

The sum total means that problems in machine design are definitely and intimately related to human ability and limitation. This field of relating man-machine systems is known as human engineering or engineering psychology.

Its growth has proceeded at an accelerated pace since World War II, when its importance was clearly recognized. Planes, guns, and other military equipment became more complicated, and as electronics became more usable, the increase in necessary dials, levers, knobs, and controls created problems. The interrelatedness of machine performance and operator performance must continually be recognized.

Psychological research on equipment design has stemmed mainly from a series of practical problems. In solving these problems a system has arisen which emphasizes the importance of psychological knowledge, facts, and methodology.

Fitts (1951) distinguishes between display and control problems. A display is any device that can be used to present information to individuals by visual, auditory, or the other senses. A good display should be rapidly and accurately detectable. Control design is concerned with the utilization of human effort in directing a machine. The dial on the radio set is a display indicating station whereas the knob is an on-off or volume control. The problems of accuracy of tuning require vision, touch, and hearing, and one or two hands may be used. The automobile is a more complicated machine to operate. The gas pedal is a control, but speed is determined by seeing a marker on a dial—or a motorcycle policeman.

Dunlap and Associates have specialized in the extension of the field of human engineering with reference to industry. A good illustration is their work in the remodeling of the cab of a dragline used by the International Minerals and Chemical Corporation. A study of the dragline (see Figure 20.7), a million-dollar machine,

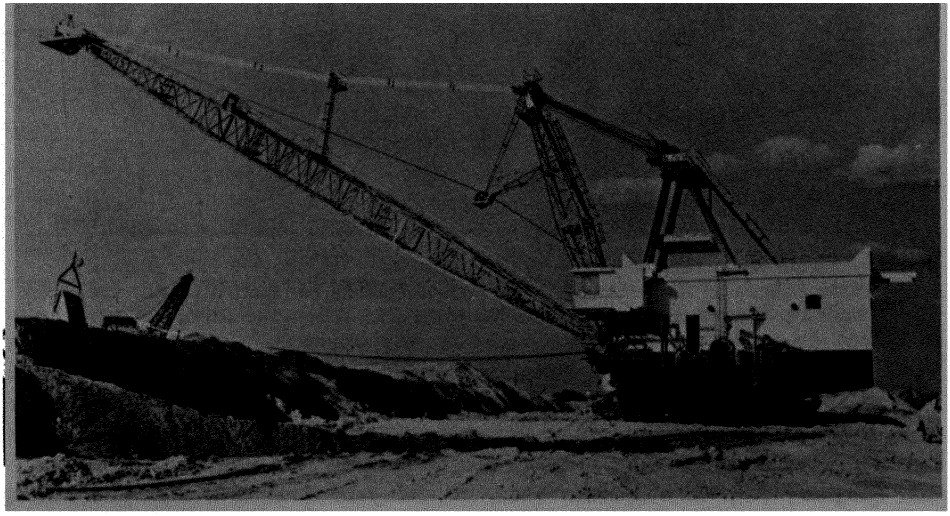


Figure 20.7. "Tillie the Toiler." Real control over one-million-dollar dragline is possible with operator's position giving good visibility, precise information, comfortable working place, and accurate, easily operated controls. "Tillie" is a 17½-yard bucket, 185-foot boom. (From *Engineering and Mining Journal*, April 1954.)

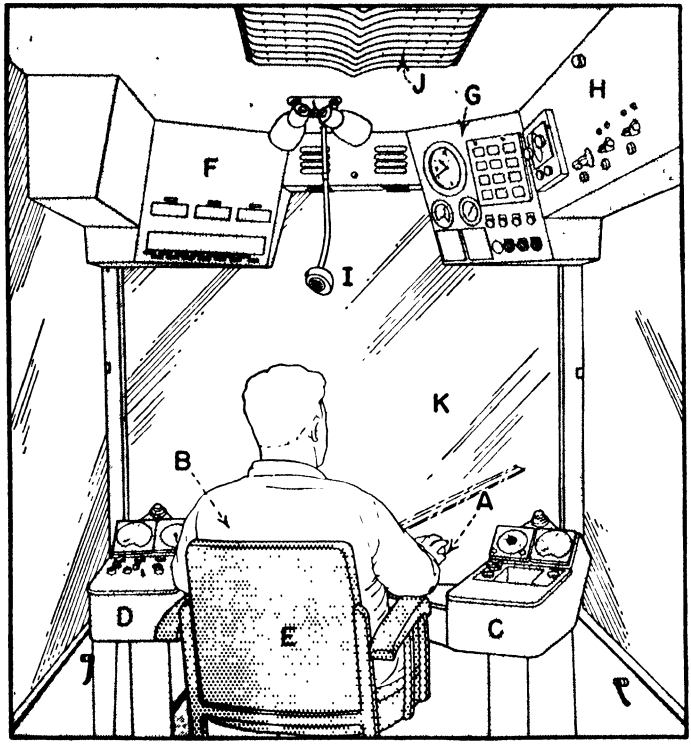


Figure 20.8. Key to cab International Minerals & Chemical Corporation builds for its draglines: *A*, hoist-swing control lever; *B*, drag control (operator's left hand); *C*, digging control indicators; *D*, switches, drag indicators; *E*, movable chair; *F*, tonnage totalizers; *G*, "trouble-spot" indicators; *H*, radio cabinet. *I*, microphone; *J*, fluorescent lighting; *K*, plate glass windshield with wipers. (From *Engineering and Mining Journal*, April 1954.)

revealed that it was operated by one man who was seated uncomfortably, exposed to the elements, and unable to see what he was supposed to be doing for much of the time. In addition, his job required awkward movements of both hands and feet.

Figures 20.7 and 20.8 show the main features of the redesign which resulted in improved vision, increased comfort and easier controls. In addition a communication system was introduced which enables the operator to talk to his oiler, pit man, pump man or the head office.

Another redesign job of Dunlap and Associates involved the electric stove. Figures 20.9 and 20.10 show the before and after and illustrate the simple principles involved. Unfortunately, the design of many consumer products is more a function of attempting to be "attractive" rather than to be functional.

Once redesign takes place, the usual reaction is that in the first place this is not psychology and in the second place this is easy and requires no special knowledge. Actually such arguments are ridiculous. Human engineering has come up the hard way. It is a bold recognition of attempting to fit the machine to meet man's abilities and limitations. When human engineering is successfully applied to the automobile

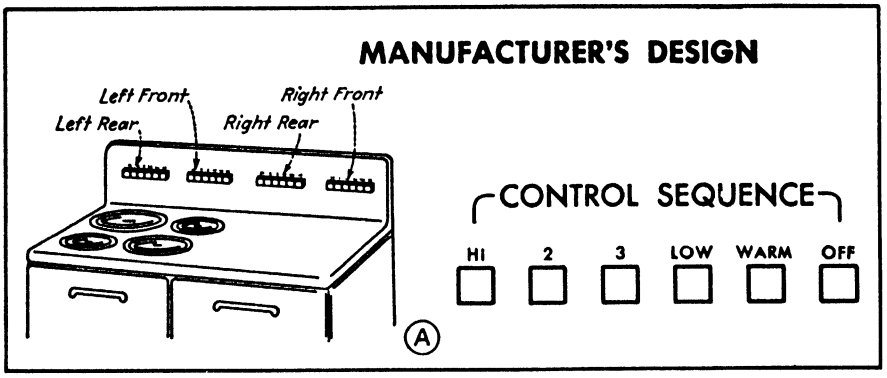


Figure 20.9. Pushbutton control system for electric range, case history. A. *Left:* Manufacturer's original version of an electric range with the controls positioned as shown on back panel. *Right:* Sequence of cooking-speed pushbutton controls violates good human engineering since it does not increase in the normal order of magnitude from left to right. (From Jack W. Dunlap. Introduction to human engineering in product design. *Electrical Manufacturing,* March 1952.)

we will have a better distribution of instruments on the panel, the visibility will be improved, glare will be eliminated, and auditory signals may successfully curb relatively innocent speeding when ordinary visual cues are lacking as is the case on superhighways.

CODING DISPLAYS AND CONTROLS

The collection of studies in the field of human engineering has led to principles related to the design of displays and controls. Some of them are as follows: A fixed

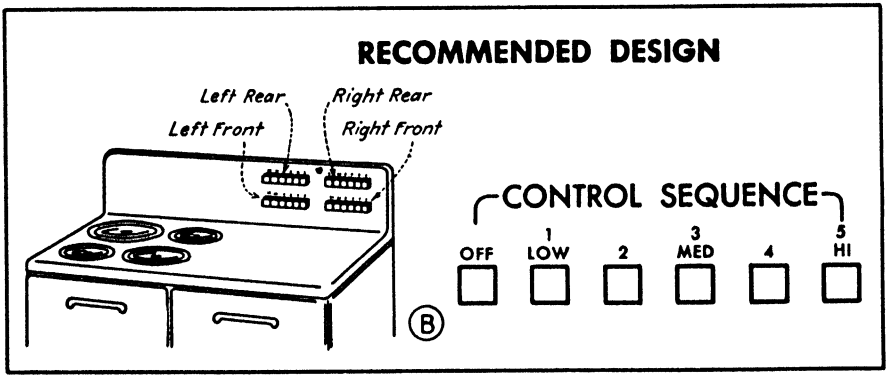


Figure 20.10. Pushbutton control system for electric range, case history B. *Left:* Recommended redesign of pushbutton control panel provides better relationship to heating units and easier operation. *Right:* The recommended redesign of pushbutton sequence showing proper order of increasing magnitude from left to right, and also showing improved clarifying labels. (From Jack W. Dunlap. Introduction to human engineering in product design. *Electrical Manufacturing,* March 1952.)

dial with a moving pointer is better than a moving dial with a fixed pointer. For speed in dial or counter reading, the finer the markings, the better. An open-window dial (direct reading counter) is best for rapid reading. All dials indicating increasing magnitudes should rotate in the same direction, preferably upward or clockwise. Whenever possible, displays should be at eye height. Spacing between markings on dials should be consistent and the distance should be about one-half inch. Shapes, sizes, and colors of controls should be designed to reduce or eliminate confusion errors.

A study by Dashevesky (1964) has demonstrated how dial reading can be tremendously improved by human engineering research. Using the Gestalt notion of figural continuity, he hypothesized that since prior research had shown that pointer alignment aided dial reading, extending the line formed by the pointers across the entire display should be even more effective. He developed six different types of dial displays (Figure 20.11) for experimental evaluation of his hypothesis.

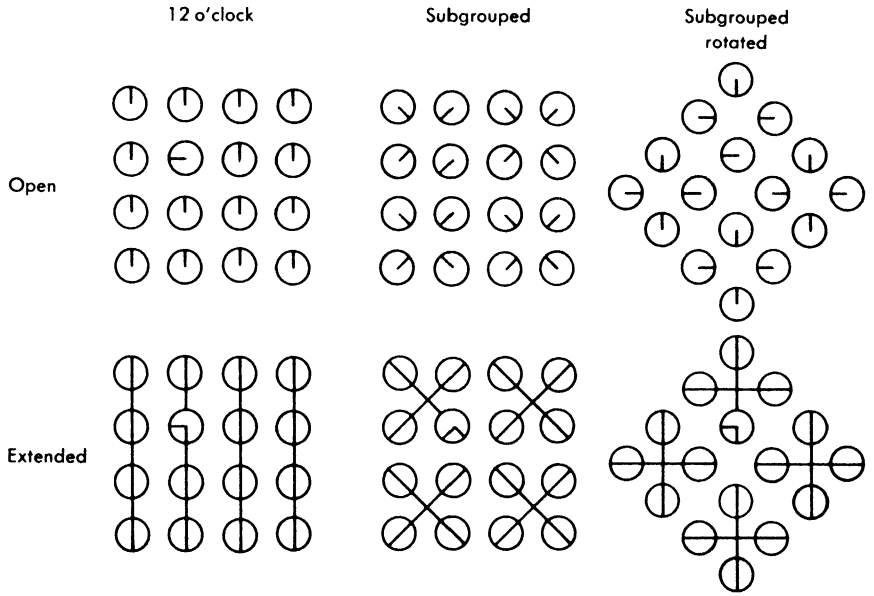


Figure 20.11. Dial displays used by Dashevsky. (From S. G. Dashevsky. Check-reading accuracy as a function of pointer alignment, patterning, and viewing angle. *Journal of Applied Psychology*, 1964, *48*, 344–347.)

He found that using the extended displays resulted in dial reading performance which was 85 percent more efficient than with the open displays, even though the latter were displays in which the pointers were all aligned according to some type of system.

Knobs in close proximity might best be designed according to different and readily discernible shapes. Jenkins (1947) found the 11 shapes shown in Figure 20.12 to be readily identifiable by touch, even when gloves are worn. Even though this research was related to aircraft, it is entirely possible that such designs of knobs would be appropriate for automobile dashboards and other types of machines.

Smith and Thomas (1964) studied the relative effectiveness of color coding dis-

plays and shape coding displays in an information processing task which required persons to count objects of a specified class presented to them on a visual display. The four different coding systems studied are shown in Figure 20.13.

They found that color coding was clearly the most effective scheme for minimizing the number of errors made by a person. This is shown quite dramatically by Figure 20.14. Colors were most efficient, military symbols were next most easily discriminated, followed by geometric forms, then aircraft shapes.

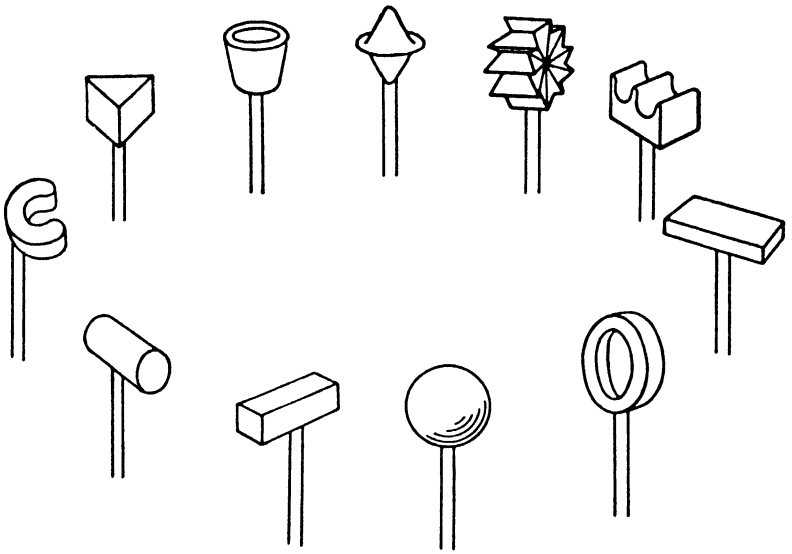


Figure 20.12. Eleven knob shapes that are readily identified by touch. (From W. O. Jenkins. Tactual discrimination of shapes for coding aircraft-type controls. In P. M. Fitts. *Psychological research on equipment design.* Government Printing Office, Washington, D.C., 1947.)

One very interesting aspect of the study was the finding that the effectiveness of the three different form coding schemes increased noticeably if color was held constant in the display; on the other hand, color coding did not show a very large improvement when form was held constant. This would seem further to indicate the "primacy" or powerfulness of color as an attention-getting device for coding. Even if it is not relevant, it can cause confusion.

In a follow-up study by Smith, Farquhar, and Thomas (1965) much the same sort of finding was obtained, except that in the second study the relative advantage of color coding over other coding systems became increasingly more dramatic as the display density (number of targets) was increased. For displays of small density, color was only moderately more effective, while with highly dense displays color became tremendously more efficient.

A good illustration of the recognition of the problem of man-machine systems is McFarland's (1953a, 1953b) work on automobiles and other kinds of vehicular equipment. The principles of human engineering have been applied in the evaluation of current vehicles with the hope of achieving in future models a more effective integration of drivers and their equipment.

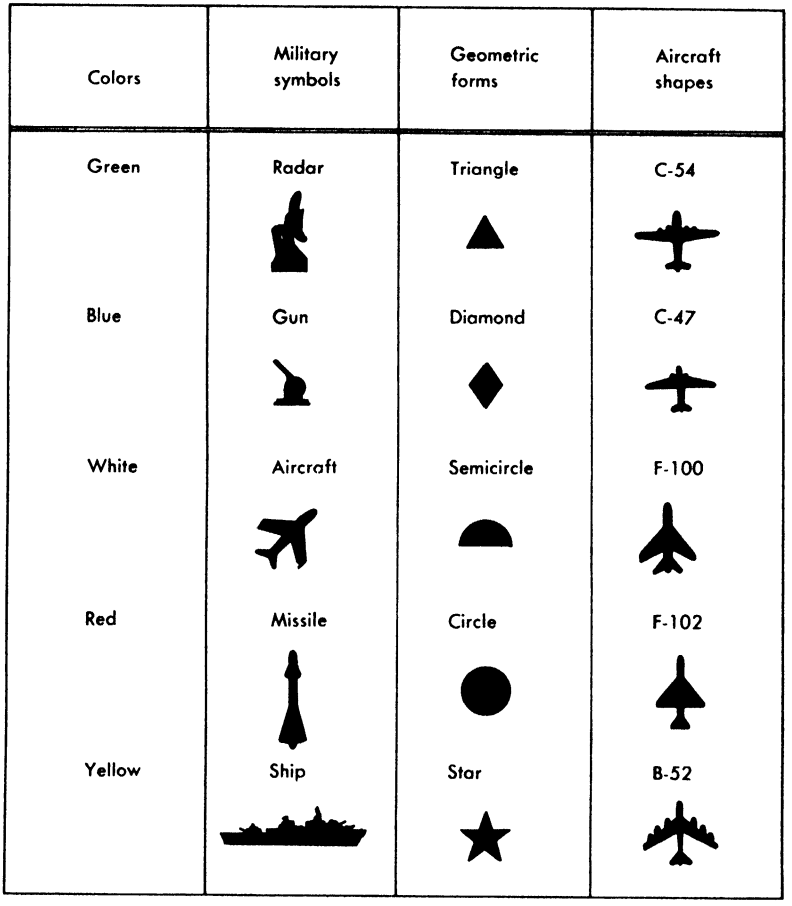


Figure 20.13. Different display coding systems investigated by Smith and Thomas. (From S. L. Smith and D. W. Thomas. Color versus shape coding in information displays. *Journal of Applied Psychology*, 1964, 48, 137–146.)

In one study an attempt was made to evaluate the cabs of twelve vehicles. Its purpose was to determine the optimum arrangements of controls, displays, seating, and window areas for the most comfortable, efficient, and safe vehicle operation. The thesis of the study is essentially based on the fact that since man cannot be redesigned it is necessary to start with man and design the machine around him. Essentially this is the clear distinction between human engineering and engineering. In engineering the machine is designed first. In human engineering the recommendation is to design the machine to conform with human needs.

The study by McFarland *et al.* found many defects in design of truck cabs. For example, it appears that satisfactory design of the instrument panel has been sacrificed for aesthetic appeal. By observation it appears that this is even more true when applied to passenger cars. In trucks the dials are placed too far to the driver's right, possibly as a concession to symmetry but surely as a contribution to ineffec-

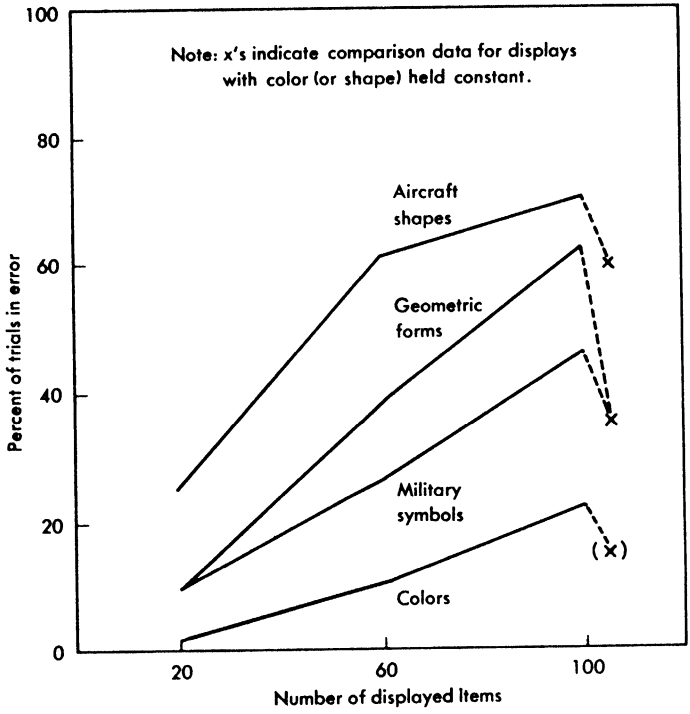


Figure 20.14. Counting errors as a function of display density, comparing color coding with the three shape codes. (From S. L. Smith and D. W. Thomas. Color versus shape coding in information displays. *Journal of Applied Psychology*, 1964, 48, 137–146.)

tiveness. Brake pedals are often placed too close to the accelerator, and the emergency brake sometimes is not conveniently accessible.

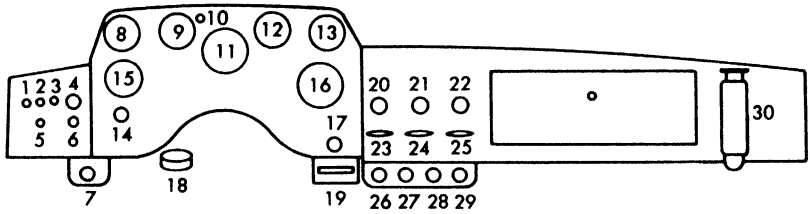
Figures 20.15 and 20.16 reveal the difference in location and design of two dashboards. The question that may realistically be asked is, How did they get that way and why?

This study made it apparent that more information was needed in regard to human body size and capabilities, and so a series of anthropometric measures of man in relation to driving requirements were made. An operator of a vehicle must not have unreasonable demands placed upon him if he is to operate controls and react to displays in an effective manner.

The complete study of man-machine systems requires at least three phases: (1) operational job analysis, (2) the study of man's limitations, both physical and psychological, and (3) the environmental conditions of work.

PRINCIPLES OF HUMAN PERFORMANCE

There is a growing tendency among applied psychologists to become interested in the *basic principles* underlying human performance in realistic task settings. This interest has been motivated by two primary factors: (1) a dissatisfaction with



Instrument	Type of Control	Instrument	Type of Control
1. Marker lights	Toggle	16. Speedometer	Dial
2. Panel light dimmer	Toggle	17. Engine primer control	Push-pull
3. Dome light	Toggle	18. Front brake control	Knob
4. Lights	Push-pull	19. Compression release	Hand pull
5. Fuel tank switch	Toggle	20. Left side vent control	Push-pull
6. Starter	Button	21. Top vent control	Push-pull
7. Wipers	Knob	22. Right side vent control	Push-pull
8. Fuel	Dial	23. Throttle	Push-pull
9. Temperature	Dial	24. Emergency engine stop	Hand pull
10. High beam	Light	25. Idle control	Push-pull
11. RPM indicator	Dial	26. Heater	Knob
12. Oil pressure	Dial	27. Defroster	Knob
13. Amperes	Dial	28. Water valve	Knob
14. Ignition	Key	29. Air control	Knob
15. Air pressure	Gauge	30. Ether bomb	

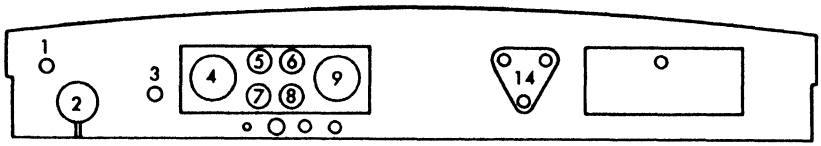
Figure 20.15. Dashboard, code no. 240. (From R. A. McFarland, *et al. Human factors in the design of highway equipment, a summary report of vehicle evaluation.* Harvard School of Public Health, Boston, 1953.)

the results obtained by simply looking continually at specific situations instead of trying to develop a broad base of knowledge about human performance, and (2) an increased willingness on the part of industry and the military to support research aimed at gathering general knowledge rather than directed at answering an immediate problem.

Current interest in human performance centers around a number of topic areas. These areas represent different aspects of skilled behavior that define what the psychologist is really talking about when he speaks of human performance. These are: perceptual-motor skills behavior, information processing behavior, and monitoring or vigilance behavior.

PERCEPTUAL-MOTOR SKILLS

The topic of motor skills extends from the study of behavior in simple manipulative-dexterity tasks (for example, pegboards) to the highly complex type of skill and coordination involved in the control of vehicles such as airplanes and space capsules. Motor activity, or perceptual-motor behavior as it is often designated, plays



Instrument	Type of Control	Instrument	Type of Control
1. Wipers	Knob	8. Temperature	Dial
2. Hi-low shift range	Knob	9. Speedometer	Dial
3. Lights	Push-pull	10. High beam	Light
4. Air pressure	Gauge	11. Ignition	Key
5. Fuel	Dial	12. Throttle	Push-pull
6. Amperes	Dial	13. Choke	Push-pull
7. Oil pressure	Dial	14. Heater controls	Knobs

Figure 20.16. Dashboard, code no. 220. (From R. A. McFarland, *et al. Human factors in the design of highway equipment, a summary report of vehicle evaluation.* Harvard School of Public Health, Boston, 1953.)

such a prominent role in our daily existence that we often take it for granted. Only when we find ourselves in certain kinds of situations do we become aware that physical coordination and dexterity is not something that is natural to all people. A few hours spent watching golfers go off the first tee on a Saturday afternoon can be a dramatic demonstration of the varieties of motor behavior possessed by different individuals.

TYPES OF MOTOR MOVEMENTS

For purposes of convenience it is helpful to classify motor movements into various categories. McCormick (1964) lists five such groups:

Positioning Movements. These are movements from one specific location to another specific location. Examples would be turning the page of a book or moving a lever from one discrete position to another discrete position.

Repetitive Movements. The essential characteristic here is that the same movement is repeated continuously, such as bouncing a ball, winding a watch, etc.

Continuous Movements. These movements involve a continuous *control* process, usually in response to some external stimulus. For example, driving a car involves continuous steering in response to the turns and bends of the road.

Serial Movements. These are a series of relatively separate movements in a sequence. They may be known in advance or may result partly as a function of previous performance. Playing a musical instrument illustrates a case of serial movements in which all of the movements are essentially the same type; preparing a peanut butter sandwich illustrates a case of serial movement where the movements are rather different throughout the series, that is, picking up knife, screwing open jar, digging out spread, spreading peanut butter, etc.

Static Movement. Maintaining a constant position over a period of time is called static movement. Although no actual movement is required, muscular exertion is necessary for static movement.

A large body of data is currently available concerning the ability of humans to perform the different types of movements listed above. For an excellent summary

of this material, examine the book *Human Factors Engineering* by E. J. McCormick (1964).

FACTOR ANALYTIC STUDIES OF MOTOR SKILL

One approach to determining the basic dimensions of motor ability is through the use of factor analysis. Typically, these studies involve giving each of a large number of people (usually several hundred) a variety of different kinds of motor tasks to perform. The tasks will range from paper-and-pencil tasks such as putting dots in circles to actual dexterity tests such as the Purdue Pegboard (see Chapter 4). By clustering these tasks into similar groups according to their intercorrelations, one can gain considerable insight into what different kinds of special skills make up the more gross skill we call motor ability.

Fleishman and his coworkers have done the greatest amount of research using this methodology. A good sample of their research is the 1962 study of Fleishman and Ellison. They administered twelve apparatus tests and nine paper-and-pencil type tests of a manipulative nature to 760 Air Force trainees, and then factor analyzed the intercorrelations among the 21 tests with the results shown in Table 20.2. They were able to identify (make sense out of) the first five factors, and they named them as follows:

Factor I:	Wrist-finger speed
Factor II.	Finger dexterity
Factor III:	Speed of arm movement
Factor IV:	Manual dexterity
Factor V:	Aiming

This kind of study is what might best be called a "static" correlational study of motor ability: it catches people at one moment in time and examines their basic skill dimensions. Fleishman has also done some factor analytic research using the procedure of testing individuals at *several* times during training to see if, as people become more skilled, the basic skill dimensions shift in their importance and emphasis. Fleishman's work on this is discussed in Chapter 8 on training.

TRACKING PERFORMANCE

Tracking skill is a very important subarea within the broader topic of motor behavior. Tracking is rather hard to describe in a specific way—it may be thought of as any motor behavior which involves aiming or guiding or pointing either a part of one's own body or some object at a target of some sort. A hunter aiming his shotgun at a pheasant in flight is "tracking" the pheasant. A driver steering his car down a highway is "tracking" the highway. A center fielder preparing to catch a fly is "tracking" the ball in flight. A great deal of man's daily activity may be viewed as tracking behavior (take, for example, raising a spoonful of carrots to one's mouth at dinner), but it is such second nature to us that we rarely think of it in these terms, except perhaps when we watch a very young child who is just learning these habitual skills.

MAN AS A SYSTEM CONTROLLER Current skill research has tended to view the perceptual-motor performance of man in terms of the task of serving as a controller of

TABLE 20.2 *Rotated Factor Loadings*

Variable	Factors						
	I	II	III	IV	V	VI	VII
1. Medium tapping	77	10	28	22	14	10	21
2. Large tapping	75	10	31	25	15	06	19
3. Aiming	52	30	20	01	57	13	−01
4. Pursuit aiming I	52	27	19	−01	63	11	−02
5. Pursuit aiming II	54	28	18	−01	63	05	−04
6. Square marking	46	26	10	20	31	00	−11
7. Tracing	12	05	11	13	15	25	19
8. Steadiness	−03	06	08	01	04	27	13
9. Discrimination reaction time—printed	30	07	03	34	00	05	−02
10. Precision steadiness	05	17	−00	10	13	34	09
11. Ten target aiming—errors	03	08	−70	09	−04	27	08
12. Ten target aiming—corrects	−04	14	72	13	31	07	06
13. Hand precision aiming—errors	−02	−04	−51	15	13	00	64
14. Hand precision aiming—corrects	15	14	56	−00	05	40	−50
15. Minnesota rate of manipulation—placing test	10	37	24	53	34	14	09
16. Minnesota rate of manipulation—turning test	09	34	23	52	30	23	08
17. Pin stick	08	34	07	25	24	20	03
18. Purdue pegboard—right hand	01	60	18	17	12	16	14
19. Purdue pegboard—left hand	05	55	18	10	05	26	11
20. Purdue pegboard—both hands	08	66	16	20	06	10	13
21. Purdue pegboard—assembly	09	59	14	32	00	12	−01
22. O'Connor finger dexterity	13	59	20	19	11	14	07

SOURCE: E. A. Fleishman and G. D. Ellison. A factor analysis of fine manipulative tests. *Journal of Applied Psychology*, 1962, *46*, 96–105.

some "system." Figure 20.17 is a diagram of the basic man-machine system concept. In almost any skilled task man may be conceived as an integral element of some larger dynamic system. That is, he (1) receives some type of information from the environment (usually some sort of display) to which he must respond, and (2) makes some type of response to this input, using whatever controls are provided by the system.

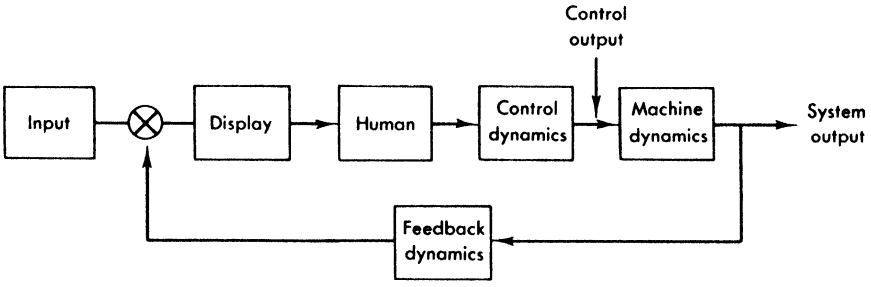


Figure 20.17. Diagram of man-machine control system.

The response is then "transmitted" through the remaining system components into actual system output. This output is then "fed back" into the display so that the operator can view his performance according to how much "error" was present in his response. For example, in the case of driving an automobile, it is the steering linkage and the tire dynamics that intervene between the human response (steering action) and the system output (the position on the road). The feedback in this system is, of course, provided via the windshield display, through which the car operator can compare his actual position on the road with the internal standard of where he knows he "should be."

SYSTEM ORDER Tracking tasks may be classified in terms of the dynamics of the control system manipulated by the operator. Generally speaking, the higher the system control order, the more complex is the operator's task.

Zero Order Control A zero order control system is often referred to as positional control. The control system simply requires a person to make a response that is proportional to the desired system output. Since the desired system output is typically a matching of the input signal, then the control system is essentially asking the operator to make responses proportional to the input signal. The control system translates (e.g., through a gear box having some specified ratio) a positioning response of the operator into a new output position for the system.

First Order Control Commonly called a *rate* or *velocity* control system, a first order control device gives the operator control over the *rate* (velocity) of movement

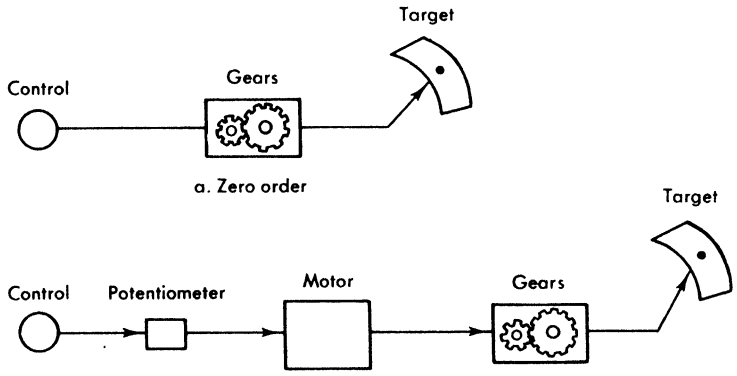


Figure 20.18. Illustrations of zero and first order control systems. In the zero order system, the position of the arrow is directly proportional to the position of the control. In the first order system, the control device regulates the speed (rate of motor and thus the *rate* of *movement* of the arrow).

of system output. An example would be having the control device linked to a motor so that a position movement by the operator changes the speed of the motor which in turn is hooked up to the gear box (see Figure 20.18). Thus, the operator controls the *rate of change* of system position, rather than its position.

Second Order Control In some control systems the operator has control over the rate of system change. By this we mean that a positional response by the operator results in a change of *acceleration* on the part of the system output.

PURSUIT AND COMPENSATORY TRACKING Tracking tasks may also be categorized in terms of whether they are *pursuit* or *compensatory* in nature. The essential difference between these two forms of tracking lies in the way the two critical elements of the task—the location of the "target" and the location of the system being controlled—are displayed to the system operator. In pursuit tracking the relative locations of both the target and the system being controlled are displayed, and both move on the display. The task of the operator is, by appropriately controlling the system, to bring the representation of the system on the display so that it coincides with the target element (i.e., he is "on target") even though the target may be in motion.

Compensatory tracking, on the other hand, provides a display in which the target is represented by a stationary element and the target alone moves. Its movement is such as to provide information to the operator as to how far and in what direction he is "off target." Any difference between the target and the system element represents degree of error at that moment in time.

SYSTEM CONTROL THEORY

A major advantage of the use of the system approach in studying the perceptual-motor aspect of human performance is that it permits the use of mathematical models in both describing and understanding such behavior. Mathematical models of human performance are always highly desirable since they permit both quantification *and* specificity. Strictly verbal models tend to be more general and of somewhat less usefulness.

System control theory is primarily based upon the notion of servomechanisms taken from the physical sciences. A servomechanism is a device which establishes a certain relationship between the input signal and the output signal. System control theory treats the human controller as a servomechanism, in that the human is described as a system element which provides a systematic relationship between the stimulus input and the response output. If the input can be described in a quantitative manner, and if the output can be similarly defined, then the relationship between output (Y) and input (X) can be expressed mathematically as some function, that is,

$$Y = f(x)$$

The function $f(x)$ is referred to as the "human transfer function" and represents mathematically the transformations the human controller applies to the input signal in the process of producing his control response.[3] Thus, the transfer function is in a very real sense a mathematical expression of human performance in a complex perceptual motor task. Research on the human transfer function in the past decade has indicated that one can "fit" mathematical equations to human controller performance which are surprisingly stable and accurate, given that the complexity of the system is not too great to over tax human capabilities. Briggs (1964) has recently shown how this approach to human performance has considerable implication to general psychological behavioral theory.

[3] For a more complete description of this area, consult Birmingham and Taylor (1954) or McCormick (1964).

INFORMATION PROCESSING BEHAVIOR

Another dimension of human performance is the ability to assimilate, process, and utilize information which is available to the individual. The study of man as an information processor thus tends to focus on these three aspects of the processing problem, *viz.* (1) the ability to *assimilate* information, (2) the ability to store and retrieve information, and (3) the ability to *utilize* information.

INFORMATION ASSIMILATION

Man assimilates information via his various sense organs; therefore, the study of his ability to obtain information from his surroundings is intimately related to the study of the limitations and advantages of the major sensory channels typically utilized for this purpose, such as vision, audition, and kinesthetic (that is, tactile). For each sensory channel, the interest has traditionally been with man's ability to *identify, discriminate,* and *classify* stimuli which are presented to him under a variety of environmental and task conditions. What characteristics of stimuli tend to make them (1) more discriminable, (2) easier to identify, and (3) easier to classify? What channels (senses) are best for presenting various kinds of information to a human? How should information be most effectively coded for presenting to the information processor? These are the kinds of questions being asked today by those studying the way in which information is assimilated.

INFORMATION STORAGE AND RETRIEVAL

The middle phase of the information processing process is generally referred to as the storage and retrieval phase. Here the interest centers on the ability of man to (1) accumulate information, (2) hold that information for use, and (3) bring it forth for use at such time as it is needed. Basically, then, the interest centers around the memory capacity of man. How much information can you reasonably give a person and expect him to still be able to use it efficiently? How can he most effectively store information? What are his storage capacity limits? Do these limits differ as a function of the kind of information, the sense channel used, the time duration they are in storage, etc.? Can the recall of information be facilitated by various coding techniques used during the storage process? Answers to date seem to indicate that man's memory capacity, even on a short-term basis, has distinct limitations (Melton, 1963), although his memory or ability to retain complex motor skills over long time periods without practice seems quite good (Naylor and Briggs, 1961).

INFORMATION UTILIZATION

The final phase of information processing concerns what man actually *does* with the information he has at his disposal at the time an action or a response is required. If the act is basically a motor skill act, then the question simply becomes one of how skillful is his motor response; if the act is more cognitive than motor, it then reduces to the problem of decision making or choice behavior. How good or how accurate a decision does he make? This topic was covered in greater detail in Chap-

ter 15, and the reader should examine that chapter for a more complete discussion of decision making as one critical dimension of human performance.

MONITORING BEHAVIOR

A dimension of human performance that has become of increasing interest in our modern technology is the ability of people to monitor displays. Since the dawn of time, men have found themselves in a position of having to keep "vigil" for events of various kinds. The best example of this can be found in military service where guard duty or "watchkeeping" duties require a man to spend long periods trying to keep alert for certain critical events or stimuli. The lookout in the "crow's nest" of the old sailing ship is a perfect example of traditional vigilance behavior. During World War II psychologists became involved with the problem of human vigilance because they wanted to learn more about man's capacity and tolerance for this kind of task. Only very recently, however, has this kind of skill become important to modern industry. The major reason for it becoming important to industry has been the tremendous shift to automated production systems by many industrial firms. This automation has typically resulted in a sharp reduction in the extent to which the worker actually manipulates the production equipment *directly*, and a sharp increase in the extent to which his task has become one of simply *monitoring* the production process—keeping watch that the machines are doing their job in the prescribed manner.

Technically speaking, vigilance or monitoring tasks are those which require the individual to detect *or* discriminate critical changes occurring in his environment. Typically, these are "rare" events, that is, the changes occur infrequently and on no predictable schedule. However, such need not be the case. What *is* critical to the process is that vigilance tasks involve rather long periods of such signal monitoring.

VIGILANCE DECREMENT

The outstanding characteristic of most vigilance performance is that there typically occurs what is known as a performance *decrement*, that is, the longer a person maintains his vigil, the poorer his performance becomes, as is shown in Figure 20.19.

It indeed was this decrement that led to much of the original interest in vigilance —what caused it and how could it best be overcome (a rather important question where poor vigilance can cost dollars and/or lives). The decrement does not occur *universally*, it should be pointed out. There is some evidence that vigilance tasks of high complexity do not result in lowered performance over time (Jerison, 1963; Adams, 1963).

SOME THEORIES OF VIGILANCE

A number of theories have been proposed to explain the performance of individuals in a vigilance-type task. Frankmann and Adams (1962) have investigated these theories and have found that none seem to explain more than a part of the present vigilance data adequately.[4]

[4] For a comprehensive review of the theories of vigilance, consult Frankmann and Adams (1962).

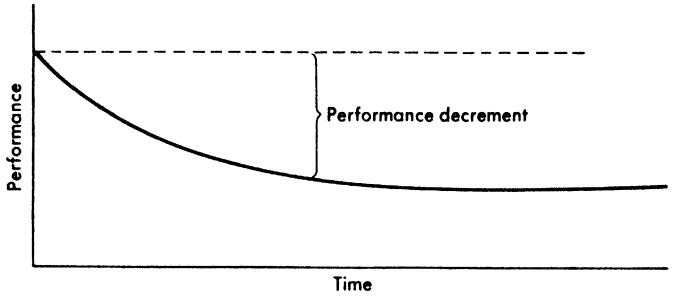


Figure 20.19. Typical performance in a vigilance-type task as a function of time.

INHIBITION THEORY Mackworth (1950) proposed that the Pavlovian classical conditioning model be used to explain vigilance performance. He argued that the vigilance decrement was due to extinction of a previously conditioned response. The major problem with this theory, however, is that complete *extinction* is never obtained in vigilance—that is, although individuals do suffer large performance decrements, they still perform at 50–75 percent efficiency even after very long time periods.

ATTENTION THEORY Broadbent (1953) has suggested that vigilance behavior can best be explained through the use of basic principles of attention. He argues that individuals "select" those stimuli to which they attend, and that stimuli of *high intensity, great biological importance,* and *high novelty* are those most apt to be selected by a person. The vigilance decrement is attributed to stimuli losing their novelty with repeated occurrences.

EXPECTANCY THEORY Deese (1955) has offered a third explanation for vigilance behavior, saying that it is the individual's "expectancy" level which determines his vigilance level, and in turn the likelihood that an individual will detect the event which occurs.

OTHER THEORIES Other theories which have been used to explain vigilance are Hebb's arousal theory (Hebb, 1955) and operant conditioning (Jerison and Pickett, 1964). However, as Howell, Johnston, and Goldstein (1966) point out, none of these models have been overly successful, and perhaps an entire new look at the area may be necessary in order to develop a behavioral model which *can* explain all the aspects of performance in both simple and complex monitoring tasks.

REFERENCES

Adams, J. A. Experimental studies of human vigilance. *USAF ESD Technical Document Report,* 1963, No. 63-320.

Birmingham, H. P., and F. V. Taylor. A human engineering approach to the design of man-

operated continuous control systems. *USN Research Laboratory Report*, April 1954, No. 4333.

Blair, J. J. Report to committee on work in industry of the National Research Council. In National Research Council. *Fatigue of Workers*. Reinhold, New York, 1941.

Briggs, G. E. The generality of research on transfer functions: Comments on Professor Fitts' paper. In A. W. Melton (ed.). *Categories of human learning*. Academic Press, New York, 1964.

Broadbent, D. E. Classical conditioning and human watch-keeping. *Psychological Review*, 1953, *60*, 331–339.

Dashevesky, S. G. Check-reading accuracy as a function of pointer alignment, patterning, and viewing angle. *Journal of Applied Psychology*, 1964, *48*, 344–347.

Deese, J. Some problems in the theory of vigilance. *Psychological Review*, 1955, *62*, 359–368.

Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (ed.). *Handbook of experimental psychology*. Wiley, New York, 1951.

Fleishman, E. A., and G. D. Ellison. A factor analysis of fine manipulative tests. *Journal of Applied Psychology*, 1962, *46*, 96–105.

Frankmann, J. P., and J. A. Adams. Theories of vigilance. *Psychological Bulletin*, 1962, *59*, 257–272.

Gilbreth, F. B. *Brick laying system*. Clark Publishing Co., New York, 1911.

Gilbreth, F. B., and L. M. Gilbreth. *Applied motion study*. Macmillan, New York, 1917.

Gilbreth, F. B., and L. M. Gilbreth. *Fatigue study*. Sturges and Walton, New York, 1916.

Gomberg, W. The relationship between unions and engineers. *Mechanical Engineering*, 1943, *65*, 425–430.

Gomberg, W. The validity of time study techniques. Doctoral dissertation submitted to Faculty of Industrial Engineering, Columbia University, New York, 1946.

Hebb, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, 1955, *62*, 243–254.

Howell, W. C., W. A. Johnston, and I. L. Goldstein. Complex monitoring and its relation to the classical problem of vigilance. *Organizational Behavior and Human Performance*, 1966, *1*, 129–157.

Jenkins, W. O. Tactual discrimination of shapes for coding aircraft-type controls. In P. M. Fitts. *Psychological research on equipment design*. Government Printing Office, Washington, D.C., 1947.

Jerison, H. J. On the decrement function in human vigilance. In D. N. Buckner, and J. J. McGrath (eds.). *Vigilance: A symposium*. McGraw-Hill, New York, 1963, 199–212.

Jerison, H. J., and R. M. Pickett. Vigilance: The importance of the elicited observing rate. *Science*, 1964, *143*, 970–971.

Lifson, K. A. Errors in time-study judgments of industrial work pace. *Psychological Monographs*, 1953, *67*(5).

McCormick, E. J. *Human factors engineering*. McGraw-Hill, New York, 1964.

McFarland, R. A., *et al.* Human body size and capabilities in the design and operation of vehicular equipment. Harvard School of Public Health, Boston, 1953.

McFarland, R. A., *et al.* Human factors in the design of highway transport equipment. Harvard School of Public Health, Boston, 1953.

Mackworth, N. H. Researches on the measurement of human performance. *Medical Research Council Special Report Service*, 1950, No. 268.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 1–21.

Mogensen, A. H. *Common sense applied to time and motion study*. McGraw-Hill, New York, 1932.

Mundel, M. E., and J. Armstrong. *Life*, Sept. 9, 1946.

Naylor, J. C., and G. E. Briggs. Long-term retention of learned skills: A review of the literature. *USAF ASD Technical Report*, No. 61-390, August 1961.

Schumard, F. W. *A primer of time study*. McGraw-Hill, New York, 1940.

Smith, S. L., Barbara Farquhar, and D. W. Thomas. Color coding in formatted displays. *Journal of Applied Psychology*, 1965, 49, 393–398.
Smith, S. L., and D. W. Thomas. Color versus shape coding in information displays. *Journal of Applied Psychology*, 1964, 48, 137–146.
Spade, M. (pseudonym). *How to run a bassoon factory*. Houghton Mifflin, Boston, 1936.
Taylor, F. W. *Principles of scientific management*. Harper & Row, New York, 1947.

APPENDIX

Naylor-Shine Table for Determining the Increase in Mean Criterion Score Obtained by Using a Selection Device

In Chapter 2 several methods were discussed for evaluating the efficiency of a selection instrument. Given the assumption of a normal bivariate distribution, the use of a mean criterion score is generally the most appropriate measure to use. Kelley (1923) provided a table for this purpose which has been extended and modernized by Naylor and Shine (1965).

While there are tables available which are even more detailed and which, for example, consider the effect of unequal group sizes and variances (most notably the work of Rorer, Hoffman, LaForge and Hsieh, 1966; and Darlington and Stauffer, 1966), these latter tables seemed too elaborate for our purpose. Thus the Naylor-Shine tables were selected for inclusion here.

USING THE TABLE

In order to use the table the following definitions must be understood:

r_{xy} = Validity coefficient
Z_{xi} = Cutoff point (score) on the predictor in standard score units
ϕ_i = Selection ratio
$\overline{Z}_{yi}$ = Mean criterion score (in standard score units) of all cases above cutoff
λ_i = Ordinate of normal distribution at Z_{xi}

and the table is based upon the following equation:

$$\overline{Z}_{yi} = r_{xy}\frac{\lambda_i}{\phi_i}$$

Generally speaking the table can be used to answer several basic personnel questions: (1) Given a particular cutoff score or selection ratio, what sort of mean criterion score increase can one expect to achieve using a predictor or series of predictors having a known relationship to the criterion?; (2) Given that a particular mean criterion score of the selected group is desired, what ratio should one use to achieve this objective?

Examples of each of these selection problems are given in the following section. However, it is important to point out that the use of the table may differ slightly in the case where r_{xy} is really a multiple regression coefficient. The major difference occurs in the Z_{xi} column. With a single predictor there is no difficulty in expressing a cutoff score in terms of a particular value of X, the predictor variable (thus we use Z_{xi}). However, in the case of a multiple predictor situation it is no longer feasible to do so, since we have many different X variables. The easiest procedure, therefore, is to conceptually reduce the multivariate case to the bivariate case by treating the multiple correlation coefficient as the correlation coefficient between the observed criterion scores (Z_y) and the predicted criterion scores (Z'_y). When viewed in this fashion it now becomes possible to talk about cutoff values for the multiple predictor case, but these cutoff scores are expressed in terms of Z'_{yi} values, rather than Z_{xi} values. The only difficulty this creates is that $s^2_{z'_y} \neq 1$, but will always be equal to R^2_{xy}, the squared multiple correlation coefficient. Thus, in order to use the tables when r_{xy} is actually a multiple correlation coefficient it is necessary to transform Z'_{yi} values by

$$Z_{xi} = \frac{Z'_{yi}}{R_{xy}}$$

In the following paragraphs examples of how to use the table are given for various kinds of problems which the personnel technician is apt to encounter. Throughout these examples it is assumed that both criterion and predictor scores are in standard form unless otherwise stated and that it is desired to *increase* the mean criterion score of the selected group over that of the present group.

SINGLE PREDICTOR CASE

In each of the following examples assume that r_{xy}, the validity of our predictor, is positive and equal to 0.50.

(a) Given a specified selection ratio (e.g., $\phi_i = 0.33$), what will be the average performance level of those selected?

Solution: Enter Table at $\phi_i = 0.33$ and read $\frac{\lambda_i}{\phi_i} = 1.10$. $\overline{Z}_{yi} = r_{xy}\frac{\lambda_i}{\phi_i} = (0.50)(1.10) = 0.55$. Thus, the mean criterion score of those men selected, using a selection ratio of 0.33, is 0.55 Z-units better than the unselected sample.

(b) Given a desired cutoff score (e.g., $Z_{xi} = -1.14$) what will $\overline{Z}_{yi}$ be?

Solution: Enter Table at $Z_{xi} = -1.14$ and read $\frac{\lambda_i}{\phi_i} = 0.24$. $\overline{Z}_{yi} = r_{xy}\frac{\lambda_i}{\phi_i} =$

(0.50) (0.24) = 0.12. Thus, using this cutoff score on our predictor results in an improvement of 0.12 Z-units in the average criterion score of selected personnel.

(c) Given a desired improvement in the average criterion score of those selected (e.g., $\overline{Z}_{yi} = 0.70$), what selection ratio and/or predictor cutoff value should be used?

Solution: Since $\overline{Z}_{yi} = r_{xy}\frac{\lambda_i}{\phi_i}$, then $\frac{\lambda_i}{\phi_i} = \frac{\overline{Z}_{yi}}{r_{xy}} = \frac{0.70}{0.50} = 1.40$. Enter Table at $\frac{\lambda_i}{\phi_i} = 1.40$ and read $\phi_i = 0.2005$ and $Z_{xi} = 0.84$. Thus, to achieve the desired improvement in average criterion performance a selection ratio of 0.20 is necessary and, to achieve that particular ratio, one should employ a cutoff score on the predictor of 0.84 in Z-score units.

In each of the following examples assume that r_{xy}, the validity of our predictor, is negative and equal to -0.50. *The general rule for this case is to reverse the sign of* r_{xy} *and* Z_{xi} *everywhere in the calculations.*

(a) Given a specified selection ratio (e.g., $\phi_i = 0.33$), what will be the average performance level of those selected?

Solution: Enter Table at $\phi_i = 0.33$ and read $\frac{\lambda_i}{\phi_i}$ 1.10. Taking the sign reversal on r_{xy} into account, $\overline{Z}_{yi} = (-r_{xy})\frac{\lambda_i}{\phi_i} = [-(-0.50)\ (1.10)] = 0.55$. Thus, the mean criterion score of those men selected, using a selection ratio of 0.33, is 0.55 Z-units better than the unselected sample.

(b) Given a desired cutoff score (e.g., $Z_{xi} = -1.14$), what will $\overline{Z}_{yi}$ be?

Solution: It should first be noted that for the case of r_{xy} being negative ϕ_i is in the *left* tail instead of being in the right tail as it is for the case of r_{xy} being positive. Thus, a cutoff score of -1.14 for the negative case is equivalent to a cutoff score of 1.14 with r_{xy} positive. For the example at hand we therefore reverse the sign of Z_{xi}, enter Table at $Z_{xi} = 1.14$, and read $\frac{\lambda_i}{\phi_i} = 1.64$. Taking the sign reversal on r_{xy} into account, we have $\overline{Z}_{yi} = (-r_{xy})\frac{\lambda_i}{Z_{xi}} = [-(-0.50)\ (1.64)] = 0.82$. Thus, using a cutoff score of $Z_{xi} = -1.14$ on our predictor results in an improvement of 0.82 Z-units in the average criterion score of selected personnel.

(c) Given a desired improvement in the average criterion score of those selected (e.g., $\overline{Z}_{yi} = 0.70$), what selection ratio and/or predictor cutoff value should be used?

Solution: Reversing the sign on r_{xy}, gives $\overline{Z}_{yi} = (-r_{xy})\frac{\lambda_i}{\phi_i}$ or $\frac{\lambda_i}{\phi_i} = \frac{Z_{yi}}{(-r_{xy})}$

Therefore, $\frac{\lambda_i}{\phi_i} = \frac{0.70}{[-(-0.50)]} = 1.40$. Enter Table at $\frac{\lambda_i}{\phi_i} = 1.40$ and read $\phi_i = 0.2005$ and $Z_{xi} = -0.84$ (note that the sign on Z_{xi} has been reversed). Thus, to achieve the desired improvement in average criterion performance a

selection ratio of 0.20 is necessary and, to achieve that particular ratio, one should employ a cutoff score on the predictor of -0.84 in Z-score units.

MULTIPLE PREDICTORS

Let Z'_{yi} be the composite predicted score based upon a multiple regression equation in standard score form. To use the Table, it is only necessary to transform these Z'_{yi} values into a standard score, Z_{xi}, by dividing Z'_{yi}, by R_{xy}, the multiple correlation coefficient between the battery and the criterion. All the examples above for r_{xy} positive now apply with r_{xy} being replaced by R_{xy}. It must be remembered that Z_{xi} in the examples refers to a standardized score and *not* to the composite score Z'_{yi}; however, one can be obtained from the other by the simple equation: $Z'_{yi} = R_{xy} Z_{xi}$. The case of R_{xy} being negative does not arise since R_{xy} is always positive or zero.

A Table for Computing the Mean Criterion Score (Z_{yi}) for the Group Falling Above Some Cutoff Score (Z_{xi})[1]

ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i
0.9987	−3.00	0.0044	0.00	0.9970	−2.75	0.0091	0.01	0.9938	−2.50	0.0175	0.02	0.9878	−2.25	0.0317	0.03
0.9986	−2.99	0.0046	0.00	0.9969	−2.74	0.0093	0.01	0.9936	−2.49	0.0180	0.02	0.9875	−2.24	0.0325	0.03
0.9986	−2.98	0.0047	0.00	0.9968	−2.73	0.0096	0.01	0.9934	−2.48	0.0184	0.02	0.9871	−2.23	0.0332	0.03
0.9985	−2.97	0.0048	0.00	0.9967	−2.72	0.0099	0.01	0.9932	−2.47	0.0189	0.02	0.9868	−2.22	0.0339	0.03
0.9985	−2.96	0.0050	0.01	0.9966	−2.71	0.0101	0.01	0.9931	−2.46	0.0194	0.02	0.9864	−2.21	0.0347	0.04
0.9984	−2.95	0.0051	0.01	0.9965	−2.70	0.0104	0.01	0.9929	−2.45	0.0198	0.02	0.9861	−2.20	0.0355	0.04
0.9984	−2.94	0.0053	0.01	0.9964	−2.69	0.0107	0.01	0.9927	−2.44	0.0203	0.02	0.9857	−2.19	0.0363	0.04
0.9983	−2.93	0.0055	0.01	0.9963	−2.68	0.0110	0.01	0.9925	−2.43	0.0208	0.02	0.9854	−2.18	0.0371	0.04
0.9982	−2.92	0.0056	0.01	0.9962	−2.67	0.0113	0.01	0.9922	−2.42	0.0213	0.02	0.9850	−2.17	0.0379	0.04
0.9982	−2.91	0.0058	0.01	0.9961	−2.66	0.0116	0.01	0.9920	−2.41	0.0219	0.02	0.9846	−2.16	0.0387	0.04
0.9981	−2.90	0.0060	0.01	0.9960	−2.65	0.0119	0.01	0.9918	−2.40	0.0224	0.02	0.9842	−2.15	0.0396	0.04
0.9981	−2.89	0.0061	0.01	0.9959	−2.64	0.0122	0.01	0.9916	−2.39	0.0229	0.02	0.9838	−2.14	0.0404	0.04
0.9980	−2.88	0.0063	0.01	0.9957	−2.63	0.0126	0.01	0.9913	−2.38	0.0235	0.02	0.9834	−2.13	0.0413	0.04
0.9979	−2.87	0.0065	0.01	0.9956	−2.62	0.0129	0.01	0.9911	−2.37	0.0241	0.02	0.9830	−2.12	0.0422	0.04
0.9979	−2.86	0.0067	0.01	0.9955	−2.61	0.0132	0.01	0.9909	−2.36	0.0246	0.02	0.9826	−2.11	0.0431	0.04
0.9978	−2.85	0.0069	0.01	0.9953	−2.60	0.0136	0.01	0.9906	−2.35	0.0252	0.03	0.9821	−2.10	0.0440	0.04
0.9977	−2.84	0.0071	0.01	0.9952	−2.59	0.0139	0.01	0.9904	−2.34	0.0258	0.03	0.9817	−2.09	0.0449	0.05
0.9977	−2.83	0.0073	0.01	0.9951	−2.58	0.0143	0.01	0.9901	−2.33	0.0264	0.03	0.9812	−2.08	0.0459	0.05
0.9976	−2.82	0.0075	0.01	0.9949	−2.57	0.0147	0.01	0.9898	−2.32	0.0270	0.03	0.9808	−2.07	0.0468	0.05
0.9975	−2.81	0.0077	0.01	0.9948	−2.56	0.0151	0.02	0.9896	−2.31	0.0277	0.03	0.9803	−2.06	0.0478	0.05
0.9974	−2.80	0.0079	0.01	0.9946	−2.55	0.0154	0.02	0.9893	−2.30	0.0283	0.03	0.9798	−2.05	0.0488	0.05
0.9974	−2.79	0.0081	0.01	0.9945	−2.54	0.0158	0.02	0.9890	−2.29	0.0290	0.03	0.9793	−2.04	0.0498	0.05
0.9973	−2.78	0.0084	0.01	0.9943	−2.53	0.0163	0.02	0.9887	−2.28	0.0297	0.03	0.9788	−2.03	0.0508	0.05
0.9972	−2.77	0.0086	0.01	0.9941	−2.52	0.0167	0.02	0.9884	−2.27	0.0303	0.03	0.9783	−2.02	0.0519	0.05
0.9971	−2.76	0.0088	0.01	0.9940	−2.51	0.0171	0.02	0.9881	−2.26	0.0310	0.03	0.9778	−2.01	0.0529	0.05

[1] ϕ_i = proportion above cutoff (selection ratio)
Z_{xi} = predictor cutoff value in standard score form
λ_i = normal curve ordinate at Z_{xi}

A Table for Computing the Mean Criterion Score (Z_{yi}) for the Group Falling Above Some Cutoff Score (Z_{xi})[1]

ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i
0.9772	−2.00	0.0540	0.06	0.9554	−1.70	0.0940	0.10	0.9192	−1.40	0.1497	0.16	0.8643	−1.10	0.2179	0.25
0.9767	−1.99	0.0551	0.06	0.9545	−1.69	0.0957	0.10	0.9177	−1.39	0.1518	0.17	0.8621	−1.09	0.2203	0.26
0.9761	−1.98	0.0562	0.06	0.9535	−1.68	0.0973	0.10	0.9162	−1.38	0.1539	0.17	0.8599	−1.08	0.2227	0.26
0.9756	−1.97	0.0573	0.06	0.9525	−1.67	0.0989	0.10	0.9147	−1.37	0.1561	0.17	0.8577	−1.07	0.2251	0.26
0.9750	−1.96	0.0584	0.06	0.9515	−1.66	0.1006	0.11	0.9131	−1.36	0.1582	0.17	0.8554	−1.06	0.2275	0.27
0.9744	−1.95	0.0596	0.06	0.9505	−1.65	0.1023	0.11	0.9115	−1.35	0.1604	0.18	0.8531	−1.05	0.2299	0.27
0.9738	−1.94	0.0608	0.06	0.9495	−1.64	0.1040	0.11	0.9099	−1.34	0.1626	0.18	0.8508	−1.04	0.2323	0.27
0.9732	−1.93	0.0620	0.06	0.9484	−1.63	0.1057	0.11	0.9082	−1.33	0.1647	0.18	0.8485	−1.03	0.2347	0.28
0.9726	−1.92	0.0632	0.06	0.9474	−1.62	0.1074	0.11	0.9066	−1.32	0.1669	0 18	0.8461	−1.02	0.2371	0.28
0.9719	−1.91	0.0644	0.07	0 9463	−1.61	0.1092	0.12	0.9049	−1.31	0.1691	0.19	0.8438	−1.01	0.2396	0.28
0.9713	−1.90	0.0656	0.07	0.9452	−1.60	0.1109	0.12	0.9032	−1.30	0.1714	0.19	0.8413	−1.00	0.2420	0.29
0.9706	−1.89	0.0669	0.07	0.9441	−1.59	0.1127	0.12	0.9015	−1.29	0.1736	0.19	0.8389	−0.99	0.2444	0.29
0.9699	−1.88	0.0681	0.07	0.9429	−1.58	0.1145	0.12	0.8997	−1.28	0.1758	0.20	0.8365	−0.98	0.2468	0.30
0.9693	−1.87	0.0694	0.07	0.9418	−1.57	0.1163	0.12	0.8980	−1.27	0.1781	0.20	0.8340	−0.97	0.2492	0.30
0.9686	−1.86	0.0707	0.07	0.9406	−1.56	0.1182	0.13	0.8962	−1 26	0.1804	0.20	0.8315	−0.96	0.2516	0.30
0.9678	−1.85	0.0721	0.07	0.9394	−1.55	0.1200	0.13	0.8944	−1.25	0.1826	0.20	0.8289	−0.95	0.2541	0.31
0.9671	−1.84	0.0734	0.08	0.9382	−1.54	0.1219	0.13	0.8925	−1.24	0.1849	0.21	0.8264	−0.94	0.2565	0.31
0.9664	−1.83	0.0748	0.08	0.9370	−1.53	0.1238	0.13	0.8907	−1.23	0.1872	0.21	0.8238	−0.93	0.2589	0.31
0.9656	−1.82	0.0761	0.08	0.9357	−1.52	0.1257	0.13	0.8888	−1.22	0.1895	0.21	0.8212	−0.92	0.2613	0.32
0.9649	−1.81	0.0775	0.08	0.9345	−1.51	0.1276	0.14	0.8869	−1.21	0.1919	0.22	0.8186	−0.91	0.2637	0.32
0.9641	−1.80	0.0790	0.08	0.9332	−1.50	0.1295	0.14	0.8849	−1.20	0.1942	0.22	0.8159	−0.90	0.2661	0.33
0.9633	−1.79	0.0804	0.08	0.9319	−1.49	0.1315	0.14	0.8830	−1.19	0.1965	0.22	0.8133	−0.89	0.2685	0.33
0.9625	−1.78	0.0818	0.08	0.9306	−1.48	0.1334	0.14	0.8810	−1.18	0.1989	0.23	0.8106	−0.88	0.2709	0.33
0.9616	−1.77	0.0833	0.09	0.9292	−1.47	0.1354	0.15	0.8790	−1.17	0.2012	0.23	0.8078	−0.87	0.2732	0.34
0.9608	−1.76	0.0848	0.09	0.9279	−1.46	0.1374	0.15	0.8770	−1.16	0.2036	0.23	0.8051	−0.86	0.2756	0.34
0.9599	−1.75	0.0863	0.09	0.9265	−1.45	0.1394	0.15	0.8749	−1.15	0.2059	0.24	0.8023	−0.85	0.2780	0.35
0.9591	−1.74	0.0878	0.09	0.9251	−1.44	0.1415	0.15	0.8729	−1.14	0.2083	0.24	0.7995	−0.84	0.2803	0.35
0.9582	−1.73	0.0893	0.09	0.9236	−1.43	0.1435	0.16	0.8708	−1.13	0.2107	0.24	0.7967	−0.83	0.2827	0.35
0.9573	−1.72	0.0909	0.09	0.9222	−1.42	0.1456	0.16	0.8686	−1.12	0.2131	0.25	0.7939	−0.82	0.2850	0.36
0.9564	−1.71	0.0925	0.10	0.9207	−1.41	0.1476	0.16	0.8665	−1.11	0.2155	0.25	0.7910	−0.81	0.2874	0.36

ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i
0.7881	−0.80	0.2897	0.37	0.6915	−0.50	0.3521	0.51	0.5793	−0.20	0.3910	0.67	0.4602	0.10	0.3970	0.86
0.7852	−0.79	0.2920	0.37	0.6879	−0.49	0.3538	0.51	0.5753	−0.19	0.3918	0.68	0.4562	0.11	0.3965	0.87
0.7823	−0.78	0.2943	0.38	0.6844	−0.48	0.3555	0.52	0.5714	−0.18	0.3925	0.69	0.4522	0.12	0.3961	0.88
0.7794	−0.77	0.2966	0.38	0.6808	−0.47	0.3572	0.52	0.5675	−0.17	0.3932	0.69	0.4483	0.13	0.3956	0.88
0.7764	−0.76	0.2989	0.38	0.6772	−0.46	0.3589	0.53	0.5636	−0.16	0.3939	0.70	0.4443	0.14	0.3951	0.89
0.7734	−0.75	0.3011	0.39	0.6736	−0.45	0.3605	0.54	0.5596	−0.15	0.3945	0.70	0.4404	0.15	0.3945	0.90
0.7704	−0.74	0.3034	0.39	0.6700	−0.44	0.3621	0.54	0.5557	−0.14	0.3951	0.71	0.4364	0.16	0.3939	0.90
0.7673	−0.73	0.3056	0.40	0.6664	−0.43	0.3637	0.55	0.5517	−0.13	0.3956	0.72	0.4325	0.17	0.3932	0.91
0.7642	−0.72	0.3079	0.40	0.6628	−0.42	0.3653	0.55	0.5478	−0.12	0.3961	0.72	0.4286	0.18	0.3925	0.92
0.7611	−0.71	0.3101	0.41	0.6591	−0.41	0.3668	0.56	0.5438	−0.11	0.3965	0.73	0.4247	0.19	0.3918	0.92
0.7580	−0.70	0.3123	0.41	0.6554	−0.40	0.3683	0.56	0.5398	−0.10	0.3970	0.74	0.4207	0.20	0.3910	0.93
0.7549	−0.69	0.3144	0.42	0.6517	−0.39	0.3697	0.57	0.5359	−0.09	0.3973	0.74	0.4168	0.21	0.3902	0.94
0.7517	−0.68	0.3166	0.42	0.6480	−0 38	0.3712	0.57	0.5319	−0.08	0.3977	0.75	0.4129	0.22	0.3894	0.94
0.7486	−0.67	0.3187	0.43	0.6443	−0.37	0.3725	0.58	0.5279	−0.07	0.3980	0.75	0.4090	0.23	0.3885	0.95
0.7454	−0.66	0.3209	0.43	0.6406	−0.36	0.3739	0.58	0.5239	−0.06	0.3982	0.76	0.4052	0.24	0.3876	0.96
0.7422	−0.65	0.3230	0.44	0.6368	−0.35	0.3752	0.59	0.5199	−0.05	0.3984	0.77	0.4013	0.25	0.3867	0.96
0.7389	−0.64	0.3251	0.44	0.6331	−0.34	0.3765	0.59	0.5160	−0.04	0.3986	0.77	0.3974	0.26	0.3857	0.97
0.7357	−0.63	0.3271	0.44	0.6293	−0.33	0.3778	0.60	0.5120	−0.03	0.3988	0.78	0.3936	0.27	0.3847	0.98
0.7324	−0.62	0.3292	0.45	0.6255	−0.32	0.3790	0.61	0.5080	−0.02	0.3989	0.79	0.3897	0.28	0.3836	0.98
0.7291	−0.61	0.3312	0.45	0.6217	−0.31	0.3802	0.61	0.5040	−0.01	0.3989	0.79	0.3859	0.29	0.3825	0.99
0.7257	−0.60	0.3332	0.46	0.6179	−0.30	0.3814	0.62	0.5000	0.00	0.3989	0.80	0.3821	0.30	0.3814	1.00
0.7224	−0.59	0.3352	0.46	0.6141	−0.29	0.3825	0.62	0.4960	0.01	0.3989	0.80	0.3783	0.31	0.3802	1.01
0.7190	−0.58	0.3372	0.47	0.6103	−0.28	0.3836	0.63	0.4920	0.02	0.3989	0.81	0.3745	0.32	0.3790	1.01
0.7157	−0.57	0.3391	0.47	0.6064	−0.27	0.3847	0.64	0.4880	0.03	0.3988	0.82	0.3707	0.33	0.3778	1.02
0.7123	−0.56	0.3410	0.48	0.6026	−0.26	0.3857	0.64	0.4840	0.04	0.3986	0.82	0.3669	0.34	0.3765	1.03
0.7088	−0.55	0.3429	0.48	0.5987	−0.25	0.3867	0.65	0.4801	0.05	0.3984	0.83	0.3632	0.35	0.3752	1.03
0.7054	−0.54	0.3448	0.49	0.5948	−0.24	0.3876	0.65	0.4761	0.06	0.3982	0.84	0.3594	0.36	0.3739	1.04
0.7019	−0.53	0.3467	0.49	0.5910	−0.23	0.3885	0.66	0.4721	0.07	0.3980	0.84	0.3557	0.37	0.3725	1.05
0.6985	−0.52	0.3485	0.50	0.5871	−0.22	0.3894	0.66	0.4681	0.08	0.3977	0.85	0.3520	0.38	0.3712	1.05
0.6950	−0.51	0.3503	0.50	0.5832	−0.21	0.3902	0.67	0.4641	0.09	0.3973	0.86	0.3483	0.39	0.3697	1.06

A Table for Computing the Mean Criterion Score (Z_{yi}) for the Group Falling Above Some Cutoff Score (Z_{xi})[1]

ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i
0.3446	0.40	0.3683	1.07	0.2420	0.70	0.3123	1.29	0.1587	1.00	0.2420	1.52	0.0968	1.30	0.1714	1.77
0.3409	0.41	0.3668	1.08	0.2389	0.71	0.3101	1.30	0.1562	1.01	0.2396	1.53	0.0951	1.31	0.1691	1.78
0.3372	0.42	0.3653	1.08	0.2358	0.72	0.3079	1.31	0.1539	1.02	0.2371	1.54	0.0934	1.32	0.1669	1.79
0.3336	0.43	0.3637	1.09	0.2327	0.73	0.3056	1.31	0.1515	1.03	0.2347	1.55	0.0918	1.33	0.1647	1.79
0.3300	0.44	0.3621	1.10	0.2296	0.74	0.3034	1.32	0.1492	1.04	0.2323	1.56	0.0901	1.34	0.1626	1.80
0.3264	0.45	0.3605	1.10	0.2266	0.75	0.3011	1.33	0.1469	1.05	0.2299	1.57	0.0885	1.35	0.1604	1.81
0.3228	0.46	0.3589	1.11	0.2236	0.76	0.2989	1.34	0.1446	1.06	0.2275	1.57	0.0869	1.36	0.1582	1.82
0.3192	0.47	0.3572	1.12	0.2206	0.77	0.2966	1.34	0.1423	1.07	0.2251	1.58	0.0853	1.37	0.1561	1.83
0.3156	0.48	0.3555	1.13	0.2177	0.78	0.2943	1.35	0.1401	1.08	0.2227	1.59	0.0838	1.38	0.1539	1.84
0.3121	0.49	0.3538	1.13	0.2148	0.79	0.2920	1.36	0.1379	1.09	0.2203	1.60	0.0823	1.39	0.1518	1.84
0.3085	0.50	0.3521	1.14	0.2119	0.80	0.2897	1.37	0.1357	1.10	0.2179	1.61	0.0808	1.40	0.1497	1.85
0.3050	0.51	0.3503	1.15	0.2090	0.81	0.2874	1.38	0.1335	1.11	0.2155	1.61	0.0793	1.41	0.1476	1.86
0.3015	0.52	0.3485	1.16	0.2061	0.82	0.2850	1.38	0.1314	1.12	0.2131	1.62	0.0778	1.42	0.1456	1.87
0.2981	0.53	0.3467	1.16	0.2033	0.83	0.2827	1.39	0.1292	1.13	0.2107	1.63	0.0764	1.43	0.1435	1.88
0.2946	0.54	0.3448	1.17	0.2005	0.84	0.2803	1.40	0.1271	1.14	0.2083	1.64	0.0749	1.44	0.1415	1.89
0.2912	0.55	0.3429	1.18	0.1977	0.85	0.2780	1.41	0.1251	1.15	0.2059	1.65	0.0735	1.45	0.1394	1.90
0.2877	0.56	0.3410	1.19	0.1949	0.86	0.2756	1.41	0.1230	1.16	0.2036	1.66	0.0721	1.46	0.1374	1.91
0.2843	0.57	0.3391	1.19	0.1922	0.87	0.2732	1.42	0.1210	1.17	0.2012	1.66	0.0708	1.47	0.1354	1.91
0.2810	0.58	0.3372	1.20	0.1894	0.88	0.2709	1.43	0.1190	1.18	0.1989	1.67	0.0694	1.48	0.1334	1.92
0.2776	0.59	0.3352	1.21	0.1867	0.89	0.2685	1.44	0.1170	1.19	0.1965	1.68	0.0681	1.49	0.1315	1.93
0.2743	0.60	0.3332	1.21	0.1841	0.90	0.2661	1.45	0.1151	1.20	0.1942	1.69	0.0668	1.50	0.1295	1.94
0.2709	0.61	0.3312	1.22	0.1814	0.91	0.2637	1.45	0.1131	1.21	0.1919	1.70	0.0655	1.51	0.1276	1.95
0.2676	0.62	0.3292	1.23	0.1788	0.92	0.2613	1.46	0.1112	1.22	0.1895	1.70	0.0643	1.52	0.1257	1.95
0.2643	0.63	0.3271	1.24	0.1762	0.93	0.2589	1.47	0.1093	1.23	0.1872	1.71	0.0630	1.53	0.1238	1.97
0.2611	0.64	0.3251	1.25	0.1736	0.94	0.2565	1.48	0.1075	1.24	0.1849	1.72	0.0618	1.54	0.1219	1.97
0.2578	0.65	0.3230	1.25	0.1711	0.95	0.2541	1.49	0.1056	1.25	0.1826	1.73	0.0606	1.55	0.1200	1.98
0.2546	0.66	0.3209	1.26	0.1685	0.96	0.2516	1.49	0.1038	1.26	0.1804	1.74	0.0594	1.56	0.1182	1.99
0.2514	0.67	0.3187	1.27	0.1660	0.97	0.2492	1.50	0.1020	1.27	0.1781	1.75	0.0582	1.57	0.1163	2.00
0.2483	0.68	0.3166	1.28	0.1635	0.98	0.2468	1.51	0.1003	1.28	0.1758	1.75	0.0571	1.58	0.1145	2.01
0.2451	0.69	0.3144	1.28	0.1611	0.99	0.2444	1.52	0.0985	1.29	0.1736	1.76	0.0559	1.59	0.1127	2.02

ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i
0.0548	1.60	0.1109	2.02	0.0287	1.90	0 0656	2.29	0.0139	2.20	0.0355	2.55	0.0062	2.50	0 0175	2.82
0.0537	1.61	0.1092	2.03	0.0281	1.91	0.0644	2.29	0.0136	2.21	0.0347	2 55	0.0060	2.51	0.0171	2.83
0.0526	1.62	0.1074	2.04	0.0274	1.92	0.0632	2.31	0 0132	2.22	0.0339	2.57	0.0059	2.52	0.0167	2.84
0.0516	1.63	0.1057	2.05	0.0268	1.93	0 0620	2.31	0.0129	2.23	0.0332	2.57	0.0057	2.53	0.0163	2.85
0.0505	1.64	0.1040	2.06	0.0262	1.94	0.0608	2.32	0.0125	2.24	0.0325	2.60	0.0055	2.54	0.0158	2.86
0.0495	1.65	0.1023	2.07	0.0256	1.95	0 0596	2.33	0.0122	2.25	0.0317	2.60	0.0054	2.55	0.0154	2.87
0.0485	1.66	0.1006	2.07	0.0250	1.96	0.0584	2.34	0.0119	2.26	0.0310	2.61	0.0052	2.56	0.0151	2.88
0.0475	1.67	0.0989	2.08	0.0244	1.97	0.0573	2.35	0.0116	2.27	0.0303	2.61	0.0051	2.57	0.0147	2.89
0.0465	1.68	0.0973	2.09	0.0239	1.98	0.0562	2.35	0.0113	2.28	0 0297	2 63	0.0049	2.58	0.0143	2.90
0.0455	1.69	0.0957	2.10	0.0233	1.99	0.0551	2.36	0.0110	2.29	0.0290	2.64	0.0048	2.59	0.0139	2.90
0.0446	1.70	0.0940	2.11	0.0228	2.00	0.0540	2.37	0 0107	2.30	0.0283	2.64	0.0047	2.60	0.0136	2.91
0.0436	1.71	0.0925	2.12	0.0222	2.01	0.0529	2.38	0.0104	2.31	0 0277	2.65	0.0045	2.61	0.0132	2.92
0.0427	1.72	0.0909	2.13	0.0217	2.02	0.0519	2.39	0.0102	2.32	0.0270	2.66	0.0044	2.62	0.0129	2.93
0.0418	1.73	0.0893	2.14	0.0212	2.03	0.0508	2.40	0 0099	2.33	0.0264	2.67	0.0043	2.63	0.0126	2.94
0.0409	1.74	0.0878	2.15	0.0207	2.04	0.0498	2.41	0 0096	2.34	0.0258	2 68	0.0041	2.64	0.0122	2.95
0.0401	1.75	0.0863	2.15	0.0202	2.05	0.0488	2.42	0.0094	2.35	0.0252	2.68	0.0040	2.65	0.0119	2.96
0.0392	1.76	0.0848	2.16	0.0197	2.06	0.0478	2.43	0 0091	2.36	0.0246	2.69	0.0039	2.66	0.0116	2.97
0.0384	1.77	0.0833	2.17	0.0192	2.07	0.0468	2.44	0.0089	2.37	0.0241	2.71	0.0038	2.67	0.0113	2.98
0.0375	1.78	0.0818	2.18	0.0188	2.08	0.0459	2.44	0.0087	2.38	0.0235	2.71	0.0037	2 68	0.0110	2.99
0.0367	1.79	0.0804	2.19	0.0183	2.09	0.0449	2.45	0.0084	2.39	0.0229	2.72	0.0036	2.69	0.0107	3.00
0.0359	1.80	0.0790	2.20	0.0179	2.10	0.0440	2.46	0.0082	2.40	0.0224	2 73	0.0035	2.70	0.0104	3.01
0.0351	1.81	0.0775	2.21	0.0174	2.11	0.0431	2.48	0.0080	2.41	0.0219	2.74	0.0034	2.71	0.0101	3.01
0.0344	1.82	0.0761	2.21	0.0170	2.12	0.0422	2.48	0.0078	2.42	0.0213	2.74	0.0033	2.72	0.0099	3.02
0.0336	1.83	0.0748	2.23	0.0166	2.13	0.0413	2.49	0.0075	2.43	0.0208	2.76	0.0032	2.73	0.0096	3.03
0.0329	1.84	0.0734	2.23	0.0162	2.14	0.0404	2.49	0.0073	2.44	0.0203	2.76	0.0031	2.74	0.0093	3.04
0.0322	1.85	0.0721	2.24	0.0158	2.15	0.0396	2.51	0.0071	2.45	0.0198	2.77	0.0030	2.75	0.0091	3.05
0.0314	1.86	0.0707	2.25	0.0154	2.16	0.0387	2.51	0.0069	2.46	0.0194	2.79	0.0029	2.76	0.0088	3.06
0.0307	1.87	0.0694	2.26	0.0150	2.17	0.0379	2.53	0.0068	2.47	0.0189	2.80	0.0028	2.77	0.0086	3.07
0.0301	1.88	0.0681	2.26	0.0146	2.18	0.0371	2.54	0.0066	2.48	0.0184	2.80	0.0027	2.78	0.0084	3.08
0.0294	1.89	0.0669	2.28	0.0143	2.19	0.0363	2.54	0.0064	2.49	0.0180	2.82	0.0026	2.79	0.0081	3.09

ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i	ϕ_i	Z_{xi}	λ_i	λ_i/ϕ_i
0.0026	2.80	0.0079	3.10	0.0018	2.91	0.0058	3.20
0.0025	2.81	0.0077	3.11	0.0018	2.92	0.0056	3.21
0.0024	2.82	0 0075	3.12	0.0017	2.93	0.0055	3.22
0.0023	2.83	0.0073	3.13	0.0016	2.94	0.0053	3.23
0.0023	2 84	0.0071	3.13	0.0016	2.95	0.0051	3.24
0.0022	2 85	0.0069	3.14	0.0015	2.96	0.0050	3.25
0.0021	2 86	0.0067	3.15	0.0015	2.97	0.0048	3.26
0.0021	2.87	0 0065	3.16	0.0014	2.98	0 0047	3.26
0.0020	2.88	0 0063	3.17	0.0014	2.99	0.0046	3.27
0.0019	2.89	0.0061	3.18	0.0013	3.00	0.0044	3.28
0.0019	2.90	0.0060	3.19				

SOURCE: J. C. Naylor and L. C. Shine. A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 1965, *3*, 33–42.

REFERENCES

Darlington, R. B., and G. F. Stauffer. A method for choosing a cutting point on a test. *Journal of Applied Psychology*, 1966, *50*, 229–231.

Kelley, T. L. *Statistical method.* Macmillan, New York, 1923.

Naylor, J C., and L. C. Shine. A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 1965, 3, 33–42.

Rorer, L. G., P. J. Hoffman, G. E. LaForge, and K. C. Hsieh. Optimum cutting scores to discriminate groups of unequal size and variance. *Journal of Applied Psychology*, 1966, *50*, 153–164.

INDICES

INDEX OF NAMES

Abelson, R. P., 304, 378
Adams, J. A., 604, 605, 606
Ailsworth, K. A., 153, 155, 173
Albright, L. E., 164, 165, 171, 172
Alessi, S. L., 503, 504, 514
Alford, L. P., 579
Anastasi, Anne, 104, 108, 140
Anderson, C. W., 157, 171
Anderson, R. C., 153, 155, 171
Angles, A., 547, 550
Annett, J., 242, 271
Arbous, A. G., 521, 522, 550
Argyle, M., 197, 234
Argyris, C., 431, 448, 479
Armstrong, Janet, 586, 606
Ash, P., 370, 388, 390
Atkinson, J. W., 453, 454, 468

Baehr, M. E., 370, 388, 390
Bahrick, H. P., 238, 246, 271
Baker, K. E., 245, 271
Bakke, W. E., 471, 489
Ballachey, E. Z., 278, 305
Banas, P., 162, 171
Barnard, C. I., 414, 434, 448, 479
Barnes, L. B., 430, 433, 448
Barrett, R. S., 376, 389
Barron, F., 155, 173
Barthol, R. P., 117, 141
Bass, B. M., 153, 171, 188, 194, 427, 428, 432, 448, 467, 468, 469
Baumgarten, F., 301, 304
Baumgartle, H., 197, 234, 302, 304, 549, 550, 551
Bavelas, A., 271, 439, 448; 476, 489
Baxter, Brent, xi, 374
Bedale, E. M., 535, 550
Bellows, R. M., 105, 146, 171, 182, 194, 374
Bendig, A. W., 209
Bennett, G. K., 110, 116, 140
Bennis, W. G., 427, 430, 448, 479, 481, 482, 489
Bergen, H. B., 303, 304
Berkowitz, L., 429, 448
Berkshire, J. R., 213, 214, 234, 513
Bernberg, R. E., 300, 305, 374, 388, 401, 413
Berniger, J., 159, 160, 171
Bernen, F. K., 564, 569
Berry, P. C., 567, 569
Bills, M. A., 9, 21
Bilodeau, E. A., 248, 271
Bilodeau, Ina McD., 271
Bingham, W. V. D., 22, 148, 171, 490
Birmingham, H. P., 602, 605
Birren, F., 567
Bittner, R. H., 106
Black, H., 98, 140
Blackwell, R. H., 209
Blair, J. J., 585, 606
Blake, R. R., 427, 429, 430, 432, 433, 448, 479, 481
Blaskovics, T. L., 166, 171
Blum, M. L., 95, 101, 105, 117, 119, 120, 121, 140, 345, 346, 362, 382, 386, 388, 408, 413, 519, 521, 552, 587
Blumenfeld, S., 357
Bolster, B. I., 157, 171
Bonneau, L. R., 153, 155, 171
Borislow, B., 136, 140
Bottenberg, R. A., 192, 194
Boulding, K. E., 454, 468
Bourdon, R. D., 220, 235
Bradford, L. P., 415, 416, 448
Bradshaw, F. F., 302, 304
Brailey, L. G., 155, 171
Brainard, R., 387
Brayfield, A. H., 373, 374, 388
Bridgman, C. S., 136, 140

Briggs, G. E., 238, 241, 245, 246, 248, 249, 260, 261, 262, 271, 273, 602, 603, 606
Britt, S. H., 245, 271
Britten, R. H., 285, 286, 291, 297, 305
Broadbent, D. E., 605, 606
Brody, M., 374, 388
Brogden, H. E., 177, 179, 181, 187, 194
Brown, C. W., 96, 140
Brown, R., 278, 304
Brown, W. C., 171
Bruce, M. M., 437, 448
Brumback, G. B., 205, 235, 482, 490
Brunswik, E., 455, 468
Bryan, W. L., 238, 271
Buckner, D. N., 218, 219, 234, 606
Buel, W. D., 162, 171, 205, 234
Bunch, M. E., 245, 271, 272
Burns, Bob, 355
Buros, O. K., 104, 105, 116, 126, 140
Burtt, H. E., 374, 388

Calkins, R. D., 451, 468
Campbell, D. T., 194
Campbell, J. T., 155, 171
Candee, B., 95, 105, 121, 140, 587
Canter, R. R., 7, 8, 19, 21
Caplan, I., 547
Capwell, D. F., 371, 373, 389
Carli, A. R., 127, 141
Carroll, J. B., 78
Carter, L. F., 256, 272
Cattell, R. B., 86
Chaney, F. B., 166, 171
Cherington, P. T., 303, 304
Chesler, D. J., 504, 514
Christal, R. E., 192, 194, 220, 234
Cioffi, F., 197, 234
Clark, R. D., 459, 468
Clarke, W. V., 115, 116, 117, 140, 141
Cobb, P. W., 521, 551
Cole, R. J., 348, 362
Common, J., 356, 362
Comrey, A. L., 105
Conger, J. J., 87
Cook, D. W., 387, 388
Cooke, J., 589
Coulson, J. E., 265, 272
Covner, B. J., 506, 514, 549, 550, 551
Cowles, J. T., 343, 362
Crawford, Dorothea M., 112, 140
Crawford, J. E., 112, 140
Crissey, O. L., 387, 388
Crissy, Wm. J. E., 153, 154, 155, 172
Crockett, W. H., 373, 374, 388
Cronbach, L. J., 58, 86, 104, 140
Crowden, G. P., 534, 535, 551
Crutchfield, R. S., 278, 305
Culbert, S. S., 563, 569
Curnow, R., 562, 570

Dabas, Z. S., 370, 388
Dailey, J. T., 153, 155, 173
Daniels, H. W., 145, 171
Darlington, R. B., 58, 609, 618
Dashevesky, S. G., 593, 606
Datta, Lois Ellen, 130, 140
Davis, K., 488, 489
Dawis, R. V., 156, 173
Dawson, R. I., 265, 272
Day, R. H., 245, 272
Deese, J., 605, 606
Dennis, J. P., 567, 569
Denton, J. C., 209
DeSilva, H. R., 525, 536, 551
Dickinson, T. L., 459, 468
Dickson, W. J., 18, 21, 290, 305, 306, 311, 327
Dill, D. B., 533, 551
Dill, W. R., 478, 490
Dingman, H. F., 219, 236
Disney, Walt, 270
Dollar, J., 338, 363
Doob, W., 338, 363
Dorcus, R. M., 106, 107, 108, 140
Drake, C. A., 522, 551
Dubin, R., 473, 474, 476, 478, 489, 500, 514
Dudycha, A., 352, 363
Dudycha, Linda, 458, 459, 469
Duffy, Elizabeth, 561, 569
Dugan, R. D., 222, 223, 234
Duncan, C. P., 245, 272
Dunlap, J. W., 592, 597, 598, 606
Dunn, R. E., 127, 141
Dunnette, M. D., 7, 21, 80, 82, 83, 86, 116, 117, 141, 160, 162, 171, 172, 189, 194, 205, 234, 343, 344, 363, 378, 388, 499, 514
Dunteman, G. H., 485, 486, 490
Dvorak, A., 588
Dvorak, B. J., 507, 508, 509, 514

Eagly, A. H., 279, 305
Eaton, W. H., 339, 363
Eckstrand, G. A., 253, 272
Edwards, A. E., 209, 234
Edwards, A. L., 294, 304
Edwards, W., 454, 469
Efferson, C. A., 449
Egerman, K., 260, 262, 272
Elliott, D. N., 209
Ellison, G. D., 599, 600, 606
Ellison, R., 128, 142, 225, 226, 230, 235
Endler, O. L., 105
England, G. W., 58, 159, 171, 499, 514
Erickson, J., 162, 171
Errett, W., 163, 172
Estep, M. F., 146, 171
Evans, C. E., 301, 304
Ewen, R. B., 378, 388

Farquhar, Barbara, 594, 607
Feinberg, M. R., 19, 21
Ferguson, L. W., 16, 17, 21
Ferree, C. E., 564, 565, 569
Festinger, L., 278, 279, 304, 305, 466, 467, 469
Fiedler, F. E., 427, 429, 432, 448
Findikan, F., 486, 487, 490
Fine, S. A., 511, 515
Fisher, V. E., 387, 388
Fiske, D. W., 155, 172, 194
Fitts, P. M., 238, 246, 271, 590, 594, 606
Flanagan, J. C., 115, 141, 205, 234, 505, 514, 530, 551
Fleishman, E. A., 159, 160, 171, 188, 194, 205, 234, 238, 239, 258, 259, 272, 374, 388, 421, 423, 424, 425, 432, 433, 446, 447, 448, 449, 599, 600, 606
Flory, C. D., 12, 21
Forbes, T. W., 526, 551
Ford, A., 563, 569
Ford, Henry, 355
Forehand, G. A., 232, 234
Fosdick, S. J., 304
Frankmann, J. P., 604, 606
Fredericksen, N., 80, 86, 230, 231, 234
French, J. R. P., 233, 234, 440, 449
Friedlander, F., 378, 388
Friedman, L., 465, 469
Friesen, E. P., 301, 304
Fruchter, B., 188, 194
Fry, E. B., 265, 272
Fryer, D., 385, 386, 388
Fuchs, A. H., 240, 272

Gadel, M. S., 374, 388
Gagné, R. M., 245, 249, 250, 271, 272
Galanter, E., 278, 304
Gardner, G., 197, 234
Gardner, J. E., 560, 561, 569
Gaudet, F. J., 127, 141
Gaylord, R. H., 78
Gekoski, N., 209
Geldreich, E. W., 540, 541, 551
Gellerman, S. W., 127, 141
Georgeopoulos, B. S., 227, 228, 229, 235
Ghiselin, B., 128, 142, 225, 226, 230, 235
Ghiselli, E. E., 80, 81, 87, 92, 93, 96, 109, 117, 122, 140, 141, 188, 189, 190, 194, 197, 234, 261, 273, 420, 421, 449, 514
Gibbs, C. B., 245, 272
Giese, W. J., 374, 388, 401, 413
Gilbert, A. C., 80, 86
Gilbreth, F. B., 355, 578, 579, 606
Gilbreth, Lillian M., 355, 539, 578, 579, 606
Giorgia, J. M., 220, 236
Glanzer, M., 261, 262, 272
Glaser, R., 260, 261, 262, 263, 272
Glennon, J. R., 164, 165, 171, 172
Gleser, G. C., 58, 86
Goheen, H. W., 168, 169, 170, 171, 172
Goldberg, M. H., 265, 272
Goldstein, I. L., 605, 606
Goldstein, L. G., 529, 530, 531, 551, 552
Gomberg, W., 505, 514, 581, 586, 606
Gonyea, G. G., 512, 513, 514
Gordon, L. V., 170, 172, 209
Gordon, O. J., 397, 413
Gough, H. G., 209
Greenly, A. J., 386, 390
Greenspoon, J., 148, 171
Greenwood, M., 521, 551
Greiner, L. E., 430, 433, 448
Griffith, J. W., 540, 542, 551
Griggs, R., 360, 363
Gross, M. L., 98, 141
Guetzkow, H., 478, 490
Guilford, J. P., 105, 114, 115, 141, 199, 206, 209, 234, 294, 305
Guion, R. M., 104, 141, 174, 183, 184, 189, 194, 392, 413
Guttman, L., 300, 305

Habbe, S., 443, 449, 506, 514
Hagen, E., 41, 43, 58
Haire, M., 188, 194, 197, 234, 449, 474, 480, 489, 490
Hall, R. L., 261, 273
Hall, W. A., 597, 598, 606
Halpin, A. W., 490
Hamidud-Din, Mohammed, 374, 388
Hammond, K. R., 298, 305, 455, 458, 459, 469
Haner, C. F., 531, 532, 551
Hanna, J. V., 387, 388
Hanson, P. G., 233, 235
Harber, H. B., 220, 236
Harrell, T. W., 166, 171, 232, 234
Harris, E. F., 258, 259, 272, 374, 388, 424, 425, 432, 449
Harris, F. J., 209, 528, 551
Harrison, R., 370, 389
Harter, N., 238, 271
Hartmann, G. W., 381, 389, 390
Hatch, R. S., 209
Hathaway, S. R., 114, 141
Hauty, G. T., 245, 272
Hay, E. N., 123, 124, 141
Hebb, D. O., 605, 606
Heermann, E., 260, 262, 272
Heise, G. A., 477, 478, 490
Helfrich, M. L., 447, 448, 449
Hempel, W. E., Jr., 238, 239, 272
Hemphill, J. K., 421, 487, 490, 499, 500, 514
Henry, M. M., 420, 449
Heron, A., 287, 305, 374, 389

Hershey, M. S., 284
Herzberg, F., 343, 371, 373, 376, 377, 378, 388, 389
Hess, E. H., 278, 304
Highland, R. W., 213, 214, 234, 513
Hilton, J., 356, 357
Himelstein, R., 166, 171
Hinrichs, J. R., 171, 500, 501, 502, 514
Hoffman, B., 98, 141
Hoffman, P. J., 56, 58, 609, 618
Hollenbeck, G. P., 136, 140
Holliday, F., 105
Holtzman, W. H., 155, 172
Hoppock, R., 366, 368, 369, 370, 380, 389
Horne, J. H., 500, 514
Horrocks, J. E., 260, 262, 272
Horst, P., 59, 87, 209
Hoslett, S. D., 449
Hotelling, H., 186, 194
Hovland, C. I., 108, 142
Howe, E. S., 214, 234
Howell, M. A., 209
Howell, W. C., 243, 272, 605, 606
Hsieh, K., 56, 58, 609, 618
Hughes, J. L., 263
Hulin, C. L., 116, 141, 370, 372, 376, 389
Hull, C. L., 23, 24
Humm, D. G., 115, 141
Hursch, C. J., 455, 469
Hursch, J. L., 455, 469
Huse, E., 155, 172
Huxtable, Z. L., 540, 541, 551

Indik, B. P., 227, 228, 229, 235, 374, 389
Irion, A. L., 244, 273
Izard, B. R., 209, 214, 234

Jackson, J. J., 549, 551
Jacobson, T. L., 226, 230, 235
James, V. A., 144, 172
Janney, E. T., 12, 21
Jeffrey, T. E., 350, 363
Jenkins, J. G., 397, 398, 413, 438, 439, 449
Jenkins, W. O., 593, 594, 606
Jenkins, W. S., 219, 236
Jenusaitis, E., 378, 389
Jerdee, T. H., 438, 449
Jerison, H. J., 604, 605, 606
Johnson, H. M., 521, 551
Johnson, K. K., 157, 172
Johnston, W. A., 605, 606
Jones, F. E., 130, 141
Jones, L. V., 350, 363
Jones, M. H., 105, 106, 107, 108, 140, 141
Jurgensen, C. E., 349, 363

Kahn, R. L., 370, 389, 424, 426, 449
Kallejian, V., 442, 449
Karr, C., 214, 234
Kates, S. L., 383, 389
Katz, D., 305, 424, 426, 449
Katzell, R. A., 376, 389
Kay, E., 233, 234
Keating, E., 156, 172
Keller, L., 87
Kelley, T. L., 609, 618
Kelly, E. L., 155, 172
Kelman, H. C., 279, 305
Kendall, L. M., 370, 376, 389
Kerr, W. A., 293, 305, 522, 541, 542, 551, 556, 557, 558, 559, 560, 561, 562, 569
Kerrich, J. E., 521, 522, 550
Kidd, J. S., 257, 258, 260, 273
Kinkade, R. G., 257, 258, 260, 273
Kirchner, W. K., 80, 86, 116, 117, 137, 138, 141, 160, 162, 171, 172, 205, 234, 499, 514
Kish, L., 282, 305
Klaus, D. J., 260, 262, 272
Knight, R., 544, 551
Knutson, A. L., 379, 380, 389
Kolbe, L. E., 105
Kolstad, A., 291, 292, 293, 296, 305
Kornhauser, A., 287, 288, 305, 378, 380, 381, 389
Krech, D., 278, 305
Kreidler, D. L., 243, 272
Kress, A. L., 502, 514
Kriedt, P. H., 374, 388
Kronenburg, B., 142
Krug, R. E., 260, 262, 272
Krugman, H. E., 302, 304
Kubis, J. F., 511, 515
Kuder, G. F., 113, 141, 387
Kunze, K., 342, 343, 363

LaForge, G. E., 56, 58, 609, 618
Lamouria, L. H., 232, 234
Lang, E. S., 245, 272
Larson, R. L., 450, 469
Laseau, L. N., 301, 304
Lauer, A. R., 530, 551
Lawler, E. E., 353, 363
Lawrence, H. G., 153, 155, 173
Lawshe, C. H., Jr., 108, 142, 374, 389, 491, 503, 504, 505, 514
Leavitt, H. J., 467, 468, 469, 470, 477, 478, 479, 481, 490
Lefkowitz, J., 19, 21
Leftwich, W. H., 209
Lemhau, P., 142
Lepkowski, J., 209
Levinson, H., 479, 480
Lewin, Kurt, xii, 337, 413
Lifson, K. A., 586, 606
Likert, R., 109, 141, 424, 427, 428, 449, 479, 480, 490

Lindahl, L. G., 267, 273
Lippitt, R., 415, 416, 448
Livingstone, D. W., 267, 273
Locke, E. A., 116, 141, 370, 376, 389
Lockman, R. F., 3, 4, 7, 21
Lockwood, H. C., 166, 172
Lodahl, T. M., 378, 389, 421, 449
Lopez, F. M., 374, 389
Lord, F. M., 71, 87
Lukaszewski, J. S., 209
Lumsdaine, A. A., 263, 272
Lundberg, G., 575
Lunneborg, C. E., 512, 513, 514
Lupton, T., 500, 514
Lytle, C. W., 341, 342, 363

McBain, W. N., 562, 569
McCarton, M. A., 540, 551
Maccoby, N., 424, 449
McCollough, C., 263, 273
McCollum, I. N., 7, 21
McCormick, E. J., 7, 21, 152, 153, 173, 175, 194, 251, 273, 500, 515, 567, 569, 598, 599, 602, 606
McCullock, T. A., 358
McFarland, R. A., 525, 526, 539, 551, 594, 595, 597, 598, 606
McGehee, W., 267, 273, 560, 561, 569
McGeoch, J. A., 244, 273
McGrath, J. J., 606
McGregor, D., 431, 449, 479, 481
McKinley, J. C., 114, 141
MacKinnon, D. W., 155, 173
Mackworth, N. H., 605, 606
McMurray, R. N., 151, 172, 430, 449
McNamara, W. J., 24, 25
Madden, J. M., 220, 234, 235
Magee, J. F., 453
Mager, J. F., 452, 453, 469
Mahler, W. R., 266, 273
Mahoney, T. A., 58, 438, 449
Maier, N. R. F., 255, 273
Maleski, A. A., 503, 514
Mandler, G., 278, 304
Mangelsdorf, J., 374
Mann, F. C., 374, 389, 549, 550, 551
Mannion, H., 357
Marks, M. R., 232, 235
Marrow, A. J., 396, 413
Maslow, A. H., 331, 332, 353, 363, 375
Massarik, F., 256, 273, 431, 443, 449
Mausner, B., 343, 371, 373, 376, 377, 389
Mayfield, E. C., 145, 153, 154, 172
Mayo, E., 306, 327, 479, 480
Mayo, T. B., Jr., 541, 542, 551
Meehl, P. E., 147, 172
Melton, A. W., 603, 606
Melville, S. D., 80, 86
Meranda, P. F., 116, 117, 141
Meyer, H. H., 233, 234, 235, 437, 438, 449
Miller, A. M., 376, 389
Miller, G., 477, 478, 490
Miller, N., 209
Miller, N. E., 338, 363
Miner, J. B., 385, 389
Mintz, A., 519, 520, 521, 552
Mogensen, A. H., 580, 581, 582, 606
Monroe, W. S., 266, 273
Moore, B. V., 148, 171
Moreno, J. L., 397, 413, 438
Morgan, J. J. B., 563, 569
Morrison, R. F., 164, 165, 172
Morse, N. C., 360, 363, 380, 381, 382, 389, 424, 449
Morsh, J. E., 494, 499, 514
Morton, R. B., 233, 235
Mosel, J. N., 168, 169, 170, 171, 172
Moseley, A. L., 525, 551, 597, 598, 606
Mosier, C. I., 86, 87, 178, 194
Mossin, A. C., 374, 389
Mosso, A., 534, 552
Mouton, Jane S., 429, 430, 433, 448, 479, 481
Mower, O. H., 338, 363
Mullins, C. J., 130, 141
Mundel, M. E., 586, 606
Münsterberg, Hugo, 17, 21, 22, 88, 306, 326
Murray, H. A., 115, 141
Murray, W. I., 209
Muscio, B., 533, 552
Myers, C. S., 550, 551
Myers, J. H., 163, 172
Myers, M. S., 378, 389

Nader, R., 525, 552
Nagle, B. F., 183, 184, 185, 186, 194, 374, 389
Nash, A. N., 438, 449
Naylor, J. C., 46, 51, 52, 58, 161, 162, 172, 191, 192, 194, 245, 246, 247, 248, 250, 260, 261, 262, 271, 273, 352, 363, 458, 459, 468, 469, 549, 552, 603, 606, 609, 618
Nealey, S. M., 350, 351, 363
Neel, R. G., 127, 141
Nevis, E. C., 127, 128, 142, 166, 173
Newcomb, T., 381, 389, 390
Newman, S. H., 209
Norman, W. T., 215, 216, 217, 218, 235
Nuckols, R. C., 293, 305
Nunnally, J., 87
Nye, C. T., 197, 235

Opsahl, R. L., 343, 344, 363
Orr, D. B., 511, 512, 514
Osgood, C. E., 278, 300, 305

Otis, A. S., 107, 108, 142
Otis, J. L., 105, 145, 171, 507, 514
Owens, W. A., 164, 165, 166, 171, 172
Oxley, W., 357

Palmer, G. J., Jr., 485, 486, 487, 490, 500, 515
Park, J. F., Jr., 563, 569
Parker, T. C., 376, 389
Parsons, S. O., 166, 172
Pashalian, S., 153, 154, 155, 172
Paterson, D. G., 108, 109, 142, 156, 172, 507
Payne, M. C., Jr., 563, 569
Peel, E. A., 209
Pepinsky, H. B., 421
Peres, S. H., 209
Peterson, C. R., 458, 469
Peterson, R. O., 371, 373, 389
Pickett, R. M., 605, 606
Pigors, P., 434, 449
Plag, J. A., 153, 155, 172
Planty, E. G., 443, 449
Porter, L. W., 261, 273, 332, 333, 334, 335, 336, 363, 378, 389, 420, 449
Posner, M. I., 563, 569
Pressey, S. L., 263, 264, 273
Price, P. B., 226, 230, 235
Prien, E. P., 153, 155, 171, 172, 500, 515

Quasha, W. H., 109, 141

Radom, M., 515
Raines, G. N., 153, 155, 172
Rand, G., 564, 565, 569
Randle, C. W., 417, 419, 420, 449
Rich, J., 285, 286, 291, 297, 305
Richards, J. M., 226, 230, 235
Rizzo, J. R., 205, 235, 482, 490
Roach, D. E., 370, 389
Robinson, H. A., 368, 389
Robinson, P., 551
Roethlisberger, F. J., 17, 21, 290, 305, 306, 311, 327, 402, 413
Rogers, C., 288, 305
Rogers, M., 245, 272
Rohrer, J. H., 153, 155, 172
Rorer, L. G., 56, 58, 135, 142, 609, 618
Rorschach, H., 115, 142
Rosenberg, M. J., 278, 304
Rosenberg, S., 209, 214, 234, 261, 273
Roslow, S., 355
Ross, P. F., 179, 194
Ross, S., 3, 4, 7, 21
Rothaus, P., 233, 235
Rothe, H. F., 197, 235, 503, 515
Rundquist, E. A., 106
Russ, J., 345, 346, 362
Russell, J. T., 51, 52, 58
Ruter, H. W., 374, 388, 401, 413

Ruttenberg, H., 352, 363, 574
Ryan, T. A., 539, 552

Sartain, A. Q., 106, 124, 142
Satter, G. A., 491, 503, 505, 514, 515
Saunders, D. R., 230, 234
Sawrey, W. L., 87
Sawyer, J., 6, 7, 21
Sayles, L. R., 472, 490
Schaffer, R. H., 382, 389
Scheidt, V., 546, 552
Schenck, E. A., 458, 469
Schenet, N. G., 549, 552
Schlaifer, R., 461, 462, 469
Schmid, J., Jr., 370, 390
Schmidt, R. N., 452, 469
Schultz, D. G., 196, 227, 235
Schumard, F. W., 579, 580, 606
Schwartz, M. M., 378, 389
Scollay, R. W., 166, 172
Scott, W. D., 16, 17, 21, 22
Scott, W. E., 339, 340, 363, 561, 569
Seagoe, M. V., 246, 247, 273
Sears, R. R., 338, 363
Seashore, S. E., 14, 21, 227, 228, 229, 235, 424, 470, 490
Seidman, J. M., 383, 384, 389, 390
Sells, S. B., 155, 172, 486, 487, 490
Selltiz, C., 408, 413
Selover, R. B., 106
Severin, D., 382, 390
Shartle, C. L., 105, 205, 235, 421, 482, 485, 486, 490, 498, 511, 515
Shaw, J., 153, 155, 172
Shepard, H. A., 479, 481
Shine, L. C., 46, 51, 52, 58, 609, 618
Shubik, M., 454, 469
Shuman, J. T., 124, 142
Siegel, A. I., 196, 227, 235
Silberman, H. F., 265, 272
Silverstein, A. B., 214, 234
Simon, H. A., 454, 455, 469, 479
Sirota, D., 374, 390
Sisson, E. D., 211, 214, 235
Skinner, B. F., 263, 273
Sleight, R. B., 564, 569
Smith, H. C., 559, 560, 561, 570
Smith, K. R., 507, 514
Smith, M., 52, 58
Smith, P. C., 370, 372, 376, 389, 562, 570
Smith, S. L., 593, 594, 595, 596, 607
Smith, W. J., 164, 172
Smith, W. R., 128, 142, 225, 226, 230, 235
Snow, A. J., 386, 390
Snow, C. E., 307, 327
Snyderman, B. B., 343, 376, 377, 389
Sobol, R., 197, 234
Spade, Mark, 342, 585, 586, 607

Spray, S. L., 500, 514
Spriegel, W. R., 144, 172
Springbett, B. M., 157, 171
Stafford, A. R., 153, 155, 173
Stagner, R., 7, 21, 98, 99, 142, 285, 286, 291, 297, 305, 349, 350, 363, 392, 393, 394, 413
Stander, N. E., 209
Stark, H., 378, 389
Stauffer, G. F., 58, 609, 618
Stead, W. H., 105
Stern, F., 170, 172
Sternberg, J. J., 153, 173
Stevens, S. S., 606
Stirling, J., 357
Stockford, L. O., 342, 343, 363
Stogdill, R. M., 375, 376, 390, 421, 427, 428, 449
Stone, C. H., 156, 172
Strauss, G., 472, 490
Strong, E. K., Jr., 113, 142, 383, 387
Strupp, H. H., 153, 173
Suci, J., 278, 300, 305
Suman, J. R., 444, 449
Summers, D. A., 458, 459, 469
Summers, S. A., 458, 469
Surgent, L. V., 106
Swets, J. A., 209
Symonds, P. M., 149, 173, 200, 235

Taft, R., 7, 9, 21
Tannenbaum, A. A., 470, 490
Tannenbaum, L R., 256, 273, 431, 442, 443, 449
Tannenbaum, P. H., 278, 300, 305
Taylor, C. W., 128, 129, 130, 142, 225, 226, 230, 235
Taylor, E K., 127, 128, 142, 166, 173, 177, 179, 181, 187, 194, 214, 235
Taylor, F. V., 602, 605
Taylor, F. W., 329, 341, 427, 572, 575, 576, 577, 578, 583, 607
Taylor, H. C., 51, 52, 58
Thomas, D. W., 593, 594, 595, 596, 607
Thompson, J. D., 479, 482
Thorndike, R. L., 41, 43, 58, 77
Thurstone, L. L., 108, 142
Thurstone, Thelma G., 108, 142
Tiffin, J., 108, 111, 142, 152, 153, 173, 175, 194, 251, 273, 386, 390, 564, 569
Todd, F. J., 455, 469
Tolle, E. R., 209
Toops, H. A., 82, 87, 184, 185, 194
Topal, J. R., 541, 542, 551
Trankell, A., 155, 173
Trattner, M. H., 511, 515
Triandis, H. C., 487, 490
Trites, D. K., 155, 173
Trumbo, D., 153, 154, 173
Tucker, L. R., 455, 469
Tuden, A., 479, 482
Tupes, E. C., 153, 155, 173
Turner, W. W., 220, 221, 235
Twery, R., 370, 390

Uhlaner, J. E., 529, 530, 552
Uhrbrock, R. S., 145, 173, 203, 204, 235, 295, 296, 300, 305
Ulrich, Lynn, 153, 154, 173

Van Ark, G., 443, 449
Van Atta, L., 263, 273
Van Steenberg, N. J., 529, 530, 552
Van Zelst, R. H., 398, 413
Vernon, A. M., 534, 552
Verplank, W. S., 148, 173
Vincent, N. L., 161, 162, 166, 167, 172, 549, 552
Viteles, M. S., 426, 449, 498, 506, 507, 515
Vroom, V. H., 79, 336, 337, 343, 363, 370, 373, 374, 375, 389, 390, 486, 490, 563, 570

Wadsworth, G. W., 115, 141
Wagner, E. E., 127, 142
Wagner, R., 152, 153, 154, 173
Walker, W. B., 233, 235
Wallace, S. R., 193, 194
Walton, E., 378, 388
Ward, B., 230, 234
Warner, C. G., 563, 570
Waters, L. K., 214, 215, 235, 248, 249, 271
Watson, G., 383, 384, 390, 396, 413
Watson, J. H., 358
Webster, H., 209
Weiss, D. J., 156, 173
Weiss, R., 360, 363
Weitz, J., 170, 173, 182, 194, 293, 305, 382, 390
Wernimont, P. F, 162, 163, 173, 378, 390
Weschler, I. R., 256, 273, 298, 300, 305, 431, 442, 443, 449
Westie, C. M., 487, 490
Wherry, R. J., 179, 191, 192, 194, 209, 214, 235, 370, 390
Wherry, R. J., Jr., 209
White, M. H., 540, 551
White, W. C., 358
Whitehead, T. N., 314, 327
Whyte, W. H., 98, 142
Wiest, W. M., 261, 273
Wiley, L., 219, 220, 235, 236
Williams, J. V., 153, 173
Wilson, C. E., 543, 552
Wilson, R. F., 503, 504, 514

Windle, C. D., 219, 236
Wolfe, J. B., 343, 363
Wolins, L., 179, 194
Wollack, L., 209
Wonderlic, E. F., 108, 142
Woods, H. M., 521, 551
Woodworth, D. G., 155, 173
Worthy, J. C., 443, 449
Wrigley, C., 370, 390
Wylie, R. C., 245, 271

Yonge, K. A., 155, 173
Yuzuk, R. P., 397, 413

Zaccaria, M. A., 153, 155, 173
Zavala, A., 209, 212, 214, 236
Zerga, J. E., 491, 515
Zimmerman, W. S., 114, 115, 141

INDEX OF SUBJECTS

Absenteeism, 373
Accident proneness, 519–522
Accidents: actuarial method of reducing, 525–527
  causes of, 516–518
  data concerning, 523
  and job satisfaction, 373
  prediction of, 528–531
  reduction of, 524–528
  role of machines in, 523
  theory of, 522–523
Acquiescence, 135
Activation level, 561–562
Activity vector analysis, 115–117
Adaptability test, 108
American Institutes for Research, 12
American Psychological Association, 5, 7
Application blank, 158–167
  life history items, 163–167
  studies of, 159–167
  validity of, 162–163
Argyris theory of leadership, 431
Argyris value theories, 479, 480
Arousal theory, 561–563, 605
Attention theory, 605
Attitude: and behavior, 277–278
  change, 278–280
  defined, 274–275
  of employees, 280–282
  formation, 276–277
  interview, *see* Interview, attitude
  measurement of, 282–301
  scales, 294–300
  surveys, 301–304
Audiogram, 111
Audiometer, 111

Balance model, 278–279
Barnard-Simon theory, 479
Bass theory of leadership, 428
Bayes model, 459–460
Bennett test of mechanical comprehension, 110, 124
Bernreuter personality inventory, 115
Blake and Mouton value theories, 479, 481
Blake's managerial grid, 429–430, 432–433, 480–481
Brunswik lens model, 455–459

Cardall test of practical judgment, 124
Central tendency errors, 201
Check list method, 494, 499
Circle with a point, 411
Color, 566–567
Compensatory tracking, 602
Concurrent validity, 35
Congruity theory, 278–279
Construct validity, *see* Validity, construct
Consulting firms: general, 12
  testing, 11
Content validity, *see* Validity, content
Contrast errors, 201
Correlation, 28–32
  multiple, 62, 85–86
  partial, 38, 83–85
  semipartial, 83–85
Cost chart, 584
Counseling interview, *see* Interview, counseling
Criteria: leadership, 416–417
  training, 267
Criterion: actual, 176
  bias, 179–182
  composite, 184–187
  contamination, 177, 179–182
  deficiency, 176
  defined, 174–175
  development, 183–184
  dimensions of, 220–230

Criterion (*Continued*)
distal, 175
distortion, 177
dynamic, 188
equivalences, 179
extension, 178
immediate, 175
model of, 177–178
multiple, 187–189
proximal, 175
relevance, 53, 176–177
requirements, 182
static, 188
types of, 195–198
ultimate, 176
Criterion-group instruments, 296–300
Critical incident method, 205, 495, 505
Cross-section analysis, 488
Cross-validation, 85–86
Cumulated points scale, 203–204
Cutoff point location, 54

Decision errors, 48, 50, 53–54
Decision makers, types of, 453–454
Decision making: cognitive model, 466–467
normative models, 455–460
organizational, 467–468
rational versus nonrational, 454–455
rules for, 465
Decision process: problems, 450–451
steps in, 451–452
Decision theory, 53–58
Decision "trees," 452–453
Diary method, 495, 500
Dictionary of Occupational Titles, 510
Discrimination indices, *see* Indices, discrimination
Dissonance theory, 278–279, 460–461
Distributions: normal, 25
skewed, 25, 26
Drives, 330
Dunlap and Associates, 12

Ecco analysis, 488
Economic man, 454
Educational Testing Service, 11
Edwards Personal Preference Schedule, 136–137
Employee problem approach, 406–409
Equating indices, *see* Indices, equating
Error-choice method, 298–300
Exchange theory of organizations, 479–480
Exit interview, *see* Interview, exit
Expectancy theory, 605
Expert approach, 403

Face validity, 36
False negatives, 50, 53
False positives, 50, 53

Fatigue: and absenteeism, 549
defined, 532–534
and production, 543–548
related phenomena, 539–543
studies of, 534–539
Feedback, 241–242, 261
Fiedler's Contingency Model, 428
Flow chart, 583
Forced-choice method, 135, 209–218
indices, 211–213
rationale, 210–211
research, 213–218
steps in, 209–210
Forced distribution, 200
Frustration-aggression hypothesis, 338–339
Fryer name finding, 123

General Aptitude Test Battery, 125–126
Glare, 565
Goals, 339
Group dynamics, 409
Group interview method, 495
Group theories of organization, 479–480
Guided interview, *see* Interview, guided
Guilford Creativity Battery, 130
Guilford-Zimmerman Temperament Survey, 114
Guttman scale, 300

Halo effect, 200
Hawthorne Studies, 306–327
Heroic man, 454
Herzberg's theory, 376–378
Hours of work, 543–548
Human engineering, 589–592
coding controls, 592–594
coding displays, 592–593
design of vehicles, 593–594
Human performance, 571, 597–598
Human Performance Laboratory, 15
Human relations training, 255
Human Resources Research Office, 16
Humm-Wadsworth Temperament Scale, 115

Illumination, 564–566
Impressionistic method, 283–284
In-basket, 230–232
training, 255
Incentives: financial, 340–344
kinds of, 344–345
research on, 345–353
Indices: discrimination, 211
equating, 211–212
See also Forced-choice method, indices
Indirect analysis, 488
Indirect method, 300–301
Individual differences, 23–26
Individual interview method, 494

Industrial counselor method, 404–406
Industrial psychology: accreditation, 5
defined, 4
division of, 5, 6, 18
graduate training, 6
history and development, 16–18
problems, 19–20
scope, 7, 8, 9
Industrial spy technique, 403–404
Information processing, 603
Information retrieval, 603
Information storage, 603
Information utilization, 603–604
Inhibition theory, 605
Institute for Social Research, 13
Internal consistency, 42
Interview: assessment, 144
attitude, 143
counseling, 142
exit, 143
guided, 284–288
performance appraisal, 232–233
purposes of, 143–144
selection, *see* Selection interview
stress, 144
Inverted pyramid, 410–411
Ishihavi Test, 110

Job analysis examples of, 498–499
methods, 494–495
research, 499–502
uses, 491–494
Job enlargement, 340
Job evaluation, 502–504
Job satisfaction, 378–385
defined, 364–365
factors in, 365–370
and job behavior, 372–375
predicting, 385–388
relative importance of factors in, 371–372
theories of, 375–376

King Arthur's Round Table, 184
Knowledge of results, 241–242, 261
Kuder Preference Record, 113, 136
Kuder-Richardson formula, 43

Leadership: behavioral approach, 421–426
and communication, 433–434
criteria, 416–417
political atmosphere, 415–416
role of wife, 447–448
selection, 436–439
theories of, 427–433
training, 439–447
trait approach, 417–421
Learning: curves, 237–240
principles, 241–244, 249–250
schedules, 244, 246–249
Leavitt theory, 479, 481
Leniency errors, 199
Levinson theory, 479–480
Life Insurance Agency Management Association, 15
Likert Scale, 297–298
Likert's modified theory of management, 428–429, 479, 480
Linkage systems, 473–478
Living-in procedure, 487
Logical errors, 200–201

McGregor theory of leadership, 431, 479, 481
McMurray theory of leadership, 430–431
MacQuarrie test battery, 124
Management games, 255–256
Manpower allocation, 22–23
Maslow's need-hierarchy theory, 331–336, 353, 375, 420, 431, 479–480
Mayo theory, 479–480
Michigan leadership studies, 424–426
Miller Analogies Test, 105
Minnesota Mechanical Assembly Test, 109
Minnesota Multiphasic Personality Inventory, 114
Minnesota Paper Form Board, 109
Minnesota Spatial Relations Test, 108
Moderators, 78–80, 101
Monitoring behavior, 604
Morale: *defined*, 365, 391–395
determiners of, 395–397
measurement of, 397–402
methods of increasing, 403–413
Motivation, 249
complexity of, 328–331
and frustration, 338–339
overview, 360–362
Motives, 329
Multiple prediction, 60–62
Music: arousal hypothesis, 561–563
effect of, 556–561
in industry, 555–556
Myart versus Motorola, 138–140

National Industrial Conference Board, 15
Naylor part-whole hypothesis, 246–248, 260–261
Naylor-Shine tables, 46, 51–52, 609–618
Needs, 330–332
Noise: as a distractor, 563
effects of, 563–564
Nominating technique, 398
Numerical scales, 202

Objective probability, 460–461
Observation interview method, 495
Occupational clustering, 511–513
Occupational information, 510–511
Occupational Outlook Handbook, 510

O'Connor Finger Dexterity Test, 106, 118
Ohio State leadership studies, 421–424
Operant conditioning, 605
Organizations: communication in, 487–488
  decentralization, 488–489
  defined, 470–471
  dimensions of, 485–487
  structure, 471–478
  theories of, 479–482
  values, 482–485
O'Rourke Test of Mechanical Aptitude, 124–125
Ortho-Rater, 110
Otis Tests of Mental Ability, 107–108, 123, 124, 437

Paired comparison scale, 206–209
Parallel tests method, 41
Patterned interview, 151
Percentiles, 89
Perceptual motor skills, 597
  factor analytic studies, 599
  motor movements, 598–599
  tracking, 599–602
Performance appraisal: errors and bias, 199–202
  forced choice, 209–218
  interview, *see* Interview, performance appraisal
  ranking systems, 205–218
  rating systems, 202–205
Personality testing, 117
Placement, 22
Plateau, 238
Prediction: multiple cutoff system, 69–71
  multiple hurdle system, 75–77
  multiple regression system, 66–69
  profile matching system, 72–75
Predictive validity, *see* Validity, predictive
Prejudice, 275
Profile comparison, 73–75
Proximity errors, 201
Psychograph, 506–507
Psychological Corporation, 11
Psychologist: characteristics, 2
  geographical location, 2
  industrial, 5
  types of, 1
Psychology: graduate schools, 3, 4
  industrial, *see* Industrial psychology
  science of, 1
Purdue Pegboard, 111
Pursuit tracking, 602

Questionnaire, 291–293
Questionnaire method, 494

Rank-order scale, 205–206
Rater characteristics, 218–220
Rating scales, 294–295
References, 167–170
  validity of, 168–170
Regression, 32–34
  multiple, 60–66
Reliability, 39–41
  types of, 41–43
Research centers, 13–16
Response sets (styles), 134–138
Rest pauses, 548
Restriction of range, 77
Richardson, Bellows, Henry and Company, 12, 297–299, 305
Rohrer, Hibler and Replogle, 12
Role playing, 255
Rorschach, 115, 383

Safety education, 527
Sampling, 282–283
Satisficing man, 454
Scaled-item instruments, 295–296
Science Research Associates, 11
"Scientific" management, 575
Selection, 22
  basic model, 26–28
  decision theory model, 53
  Dunnette's model, 82
Selection interview, 143, 144
  accuracy, 155–157
  characteristics of, 147–150
  data sources, 146–147
  format, 150–152
  length, 145
  model, 145–146
  negative versus positive information, 157–158
  reliability, 152–154
  usage, 144–145
  validity, 154–155
Selection ratio, 46
Semantic differential, 300
Sensitivity training, 256–257
Shepard theory, 479, 481
Shrinkage, 85–86
Sight-Screener, 110
Significance, 34
Similarity errors, 201
Situational theories of organization, 479, 481–482
Snellin chart, 110
Social desirability, 135
Sociogram, 398–399
Sociometry, 397
Spearman-Brown prophecy formula, 42, 43
Split-half method, 42
Standard scales, 203
Standard scores, 89
Stanford-Binet Intelligence Scale, 91
Stogdill's theory of leadership, 375–376, 428

Stress interview, *see* Interview, stress
Strong Vocational Interest Blank, 113, 137
Structural theories of organizations, 479, 481
Subdivided test method, 42
Subjective probability, 460–461
Subjective utility, 461
Suppressor variable, 83
Synthetic validity, *see* Validity, synthetic
System controller, 599
System order, 601

Tannenbaum theory of leadership, 431
Task: complexity, 247–248, 260–261
  difficulty, 248
  fractionation, 248–249
  organization, 248, 260–261
  simplification, 248–249
Taylor-Russell tables, 51
Team training, 259–262
Technical conference method, 495
Telebinocular, 110
Test-retest method, 41
Testing. and civil rights, 102–104, 138–140
  creativity, 130
  examples in industry, 117–125
  executives, 126–128
  programs, 100
  scientists, 128–130
  United States Employment Service, 125–126
Tests: characteristics of, 88–89
  costs, 96
  dangers of, 94–95
  interest, 112–113
  mechanical aptitude, 108–110
  mental ability, 107–108
  motor ability, 111–112
  norms, 89–90
  personality, 113–116
  public attitude, 97–99
  sensory ability, 110–111
  sources of information, 104–107
  standards, 95, 99–100
  types of, 90–91
Thematic Apperception Test, 115
Therblig, 578–579
Thompson and Tuden theory, 479, 482
Threshold, auditory, 111
Thurstone Test of Mental Alertness, 108
Time and motion study, 571–572, 575–587
  failure of, 572–574
  in the home, 586–587
  principles, 582–584
  resistance to, 574–575, 588
  and unions, 581
Training: evaluation of, 265–267
  executive, 254–259
  versus experience, 251
  kinds of, 251–252
  programed instruction, 262–265
  staff, 270–271
  systems approach to, 252–253
  team, 259–262
  transfer of, 244–246
Transfer function, 602
True negatives, 50
True positives, 48
True score, 39
Turnover, 373

Unguided interview, 288–291
Unstructured interview, 151
Utility, 461
Utility of a predictor, 44
  effect of percent of current employees who are successful, 48–51
  effect of reliability of predictor and criterion, 52–53
  effect of selection ratio, 46–48
  effect of validity, 44–45

Validity, 44–45
  bias in, 37–38
  conditional, 79
  construct, 36
  content, 36
  decision theory model, 53
  effective, 50
  extension, 86
  generalization, 86
  predictive, 35
  relevance, 53
  and reliability, 52–53
  synthetic, 36
  type of test, 92
  types of, 35
  typical magnitudes, 80
    See also Selection interview, validity
Value theories of organizations, 479, 480–481
Vertical round table, 409–410
Vibration, 567
Vigilance: decrement, 604
  theories of, 604–605
Vroom motivational theory, 336–338, 375

Wagner Act, 138
Wherry-Doolittle test selection procedure, 63
Wonderlic Personnel Test, 108
Work environment, 552–555
Work participation method, 495

Ziegler Rate of Manipulation tests, 123