ADULT INTELLIGENCE

A PSYCHOLOGICAL STUDY OF TEST PERFORMANCES

By Theodore Weisenburg, M.D.,
Anne Roe, Ph.D., and
Katharine E. McBride, Ph.D.

NEW YORK
THE COMMONWEALTH FUND
London · Humphrey Milford · Oxford University Press
1936

PUBLISHED FEBRUARY, 1936
SECOND PRINTING JULY, 1937

THE COMMONWEALTH FUND, 41 EAST 57TH ST., NEW YORK, N.Y.

PRINTED BY E. L. HILDRETH & CO., INC., BRATTLEBORO, VERMONT

PREFACE

THIS book is the report of an extensive study of adult intelligence and educational achievement by means of psychological tests. The group of cases examined, numbering seventy, is comparatively small, but the subjects were carefully selected and they represent a good sample of the middle levels of the population.

The selection of the cases and the nature of the examinations made are explained by the origin of the research. For many years before the work was undertaken, Dr. Theodore Weisenburg had been greatly interested in determining the characteristics of normal behavior as a basis for interpreting the abnormal. He urged further study of the various sensory modalities in normal persons before attempts were made to define the qualities or limits of the abnormal responses. He was particularly impressed by the uncertainty in differentiating between the normal and the abnormal in the fields of language and behavior disturbed in aphasia, apraxia, and agnosia. When Henry Head's volumes on *Aphasia and Kindred Disorders of Speech* appeared in 1926, Dr. Weisenburg turned his attention to Head's tests for aphasia and in 1929 he directed a study of these tests with non-aphasic cases.[1] The findings showed that superior adults made some of the errors which Head had found characteristic of aphasic patients, while low-grade adults made a greater number of such errors. It was evident that there were few, if any, tests which would provide for an absolute distinction between the aphasic and the normal, and that a better understanding of aphasia awaited a more complete knowledge of the abilities of normal adults.

In 1929 Dr. Weisenburg planned an extensive program of research which included both a study of sensation in normal persons, and a new attack on the problem of aphasia through a psychological study of normal as well as abnormal cases. The Commonwealth Fund of New York City generously provided a grant to support the work through 1932. During these years Dr. Weisenburg devoted much of his time to the supervision of the sensory research, to the development of instruments for testing sensation, to the examination of aphasic patients, and to the study of the problem of aphasia.

Earlier investigations had shown that aphasic disturbances extended far beyond simple speech responses and could be studied fully only by means of

[1] G. H. J. Pearson, B. J. Alpers, and T. H. Weisenburg, *Aphasia: A Study of Normal Control Cases,* Arch. Neur. & Psychiat., 1928, 19, 281–295.

an extensive battery of tests. The battery arranged therefore included tests of the simple speech processes; of more complicated processes requiring language facility and "verbal intelligence"; of attainment in reading, writing, spelling, and arithmetic; and of performance on non-verbal tests demanding intelligent behavior. On the basis of these tests, studies were made of 234 aphasic and control cases. The results for the aphasic and the apraxic or agnosic cases, as well as for the control cases of unilateral cerebral lesion without these disorders, have been reported in a separate volume.[2] In every case the abnormal performances were interpreted in the light of the behavior of the normal subjects studied by the same methods.

The work with the normal adults was begun in 1929 by Mildred Willard Gardiner, M.A., who examined approximately half of the subjects. The remainder were examined by Anne Roe, Ph.D.

In order to make satisfactory comparisons between normal and abnormal individuals, the groups had to be as nearly as possible equivalent in age and in educational, occupational, and social status. The majority of the aphasic and other cases with cerebral lesion were studied in three Philadelphia hospitals: the Orthopedic Hospital and Infirmary for Nervous Diseases, the Graduate Hospital of the University of Pennsylvania, and the Philadelphia General Hospital. However paradoxical it seemed, it was clear that the so-called normal adults would have to be selected from among the patients admitted to the same hospitals, for only thus could a random selection from similar levels of the population be obtained. As will be described in Chapter II, the normal cases were chosen from the surgical or orthopedic wards and were accepted for study only if the record was free of neurological or mental disorders and the physical condition satisfactory for good work.

For permission to study the patients selected for the normal group the authors are indebted to Dr. William J. Taylor, the late Dr. A. P. C. Ashhurst, Dr. A. Bruce Gill, and Dr. Ralph Pemberton at the Orthopedic Hospital and Infirmary for Nervous Diseases; the late Dr. J. B. Carnett, Dr. W. G. Elmer, Dr. W. E. Lee, and Dr. H. L. Bockus at the Graduate Hospital; and to Dr. E. L. Eliason, Dr. H. R. Owen, Dr. M. Behrend, Dr. J. O. Bower, Dr. J. T. Rugh, and Dr. H. M. Righter at the Philadelphia General Hospital. Dr. W. G. Turnbull at the Philadelphia General Hospital very kindly made the necessary arrangements there. The authors are

[2] T. Weisenburg and K. E. McBride, *Aphasia: A Clinical and Psychological Study.*

also grateful to the staffs in each of these hospitals, for their interest and co-operation made it easy to secure patients for the frequent work periods and to arrange the conditions necessary for satisfactory work.

Representing as they do a good sample of the middle levels of the population, the "normal" adults constitute an important group and one rarely available for psychological study. Unless complete surveys of a community are made by house-to-house testing of every individual, it seems almost impossible to secure an unselected sample of the population. The brighter individuals are more willing to demonstrate their powers than the dull, and the young apparently respond more readily than the old. The selective factors thus introduced are at a minimum in the hospital situation. The preliminary condition of a disorder requiring admission for surgical or orthopedic treatment probably has little, if any, relation to intelligence or educational achievement. The further condition that only patients who wanted to take the studies were included had some but comparatively little influence as a selective factor in the group, for only 21 per cent of the patients who were approached refused. This percentage was, however, higher for the subjects in the fourth and fifth age decades than for those in the second and third.

The hospital situation has another important advantage if extensive psychological studies are required, as they were in the case of this work. When subjects are recruited in other ways, their time is usually limited. Few active, employed persons have sufficient interest to give more than an hour or two for psychological examinations. The unemployed can be induced to take long examinations for some remuneration, but these cases alone probably constitute a below-average sample of the population. The hospital group not only includes a random sample of average and near-average classes of the population, employed as well as unemployed, but it is composed of persons who find time heavy on their hands and are usually willing to spend an almost unlimited amount of it on psychological examinations.

The results reported in this monograph include, first, data on the average performance level and the variability in performance for a large number of mental and educational achievement tests, together with facts as to the qualitative characteristics of "normal" responses. This material is important as a basis for future work with groups of normal adults as well as with the many pathological cases where there are changes or possible

changes in mental functioning. From the point of view of work with the normal, much of the material is particularly timely for it can be applied in the adult education projects which are now assuming such a prominent place in American life. In many cases adults, like children entering a school, must be given preliminary tests to provide for a satisfactory grouping of individuals of different levels of development. In all cases preliminary examination is required to determine their comprehension in reading, their language facility, and various other factors on which their success in the work will depend. From the point of view of studies of the abnormal, it must be noted particularly that standards of intellectual performances for the "average man" are still little known and that they have not before been determined for most of the individual tests included in this battery. Yet some of these are tests of primary importance in work with the abnormal, where the purpose is not only the determination of the intellectual level, but the qualitative analysis of the patient's performance, his methods of work, and his attitudes.

The monograph also contains analyses of the results for their bearing on a number of important problems: the comparative findings for the groups of men and women, the relationship of age and test performance, the factor of education, the differences in test performance in certain occupational classes, the relative standing of the individual on a wide variety of tests, and, finally, the relationship between the different test performances for the group as a whole. The findings on all these questions receive general discussion in Chapter IX, and the reader who is interested in the major contributions of the study rather than the detailed results is referred especially to this chapter.

It had been Dr. Weisenburg's plan to write the preface for this monograph, and in it to stress the practical significance of the results for the psychiatrist, the educator, and the psychologist. With his death, before the final draft of the book was completed, the preface had not yet been written. I have included the points he had intended to cover, in so far as I know them, but I deeply regret that the work has not had the benefit of his final criticisms, for one of the qualities which made him great as a director of research was his unusual ability to synthesize the results of a scientific investigation and foresee both their theoretical and their practical import.

K. E. McB.

April, 1935

CONTENTS

TABLES

FIGURES

I

A CRITICAL SURVEY OF STUDIES RELATING TO ADULT INTELLIGENCE

1. INTRODUCTION

WITHIN the last few years the problems of adult intelligence have for the first time in the history of "mental testing" been receiving the study their importance merits. It is true that they came into great prominence during the War when Yerkes and others undertook the colossal task of arranging intelligence tests for all grades of adults from the dullest to the most gifted, and the further task of having these tests quickly and well administered and interpreted.[1] But the primary purpose of the Army survey was a rough determination of the grade of intelligence for each of the nearly two million men tested, and there was no time to give to finer points. Thus while the Army data demonstrate the general mental level of the average man to the comparative satisfaction of the psychologist, if not of the layman, they do not provide material for study of the further problems which are coming to occupy the center of the field. The question of the relationship of different mental abilities, which attracts so much research at the present time, has been studied on the basis of the Army data, notably by Spearman, Kelley, and Brigham; but most investigators agree with Kelley that the material is not really suitable for such analysis.[2] The question as to the course of different test performances with age cannot be solved on the basis of the Army findings, for at different ages there were probably differing selective factors which made the various age groups incomparable.

The recent studies of adult intelligence have yielded important data on

[1] R. M. Yerkes, *Psychological Examining in the United States Army*, Mem. Nat. Acad. Sci., 1921, 15.

[2] T. L. Kelley, *Crossroads in the Mind of Man*, 1928, 215.

the relationship between age and various test performances. The two outstanding contributions are the large survey of a New England community made by Jones and Conrad[3] by means of the Army Alpha Test and the Maturity Study made by Miles[4] at Stanford. The first is unique among studies of adults because it is a survey of an almost complete community. For the given racial and cultural group it yields decisive information on the typical adult performance and the range of performances. Furthermore, in so far as the eight sub-tests of the Army Alpha permit, the study furnishes data on the characteristics of different types of test performance. The length of the examination had of course to be limited to make possible the testing of large numbers at one sitting and to ensure their cooperation.

The Maturity Study at Stanford is also based on large numbers of adults, but they are probably less representative of the population at large than those examined by Jones. In the first testing program in 1930 approximately 800 took the brief intelligence test, a fifteen-minute form of the Otis Self-Administering Test, and in addition certain brief tests of other functions, including perception, motility, memory, and imagination. Other large groups of subjects were given various tests in the following years. The testing program as a whole was more extensive than that of any other investigation. It must be noted however, first, that the attempt to cover a large number of different types of performance in a short time resulted in the use of many extremely brief tests; and, second, that the results of certain parts of the investigation must be interpreted separately, for not all of the groups entering into the study were comparable.

Both these investigations will be discussed in greater detail in later sections of the chapter, which are intended to survey the various lines of work contributing directly or indirectly to the study of adult intelligence.

2. DEVELOPMENT OF MENTAL TESTING IN COLLEGE GROUPS

Test situations which would bring out individual differences in mental traits are by no means inventions of modern science. One recalls Gideon's method of reducing his force of 10,000 before the attack on the Midianites,

[3] H. E. Jones and H. S. Conrad, *The Growth and Decline of Intelligence,* Genet. Psychol. Monog., 1933, 13, 223–298.

[4] C C. and W. R. Miles, *The Correlation of Intelligence Scores and Chronological Age from Early to Late Maturity,* Amer. J. Psychol., 1932, 44, 44–78; W. R. Miles, *Age and Human Ability,* Psychol. Rev., 1933, 40, 93–123.

when he led the whole force to a stream to drink, and selected as the best warriors those who did not for a moment leave themselves unprotected. It was a long time, however, before test situations were more than gross indices, of practical value in that they set apart some of the most pronounced mental abilities or defects. The first scientific interest may be safely dated from Galton, who studied a number of functions in which individuals differed, chiefly in the field of sensation. Most of these functions would not be accepted today as indices of intelligence, but Galton's purpose was essentially that of the modern investigator, "to obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points."[5]

It is only with regard to the studies of adult intelligence that the development of mental testing will be discussed here, for the subject as a whole has been reviewed many times.[6] In the United States, where so much of the early work was confined to the college laboratory, students were regularly pressed into service as subjects. The choice was made rather because of convenience than because of a dominant interest in mature intelligence, but in any case many of the earliest "mental tests" were administered to the adults of the group represented in college.

The tests included physical and psychophysical as well as more strictly psychological measurements, and indeed the latter were usually allotted less time than the former. The first series was suggested by Cattell in 1890, in a paper in which the term "mental test" made its appearance.[7] It was undoubtedly Cattell who set the ball rolling and who dominated the first decade of mental testing in America. The tests he used, as reported in the 1890 paper, included experiments on two-point discrimination, pressure causing pain, least noticeable difference in weight, dynamometer pressure, rate of movement, and reaction time for sound, as well as tests of "association time" in color naming and of immediate memory for letters. It is interesting to find that in this first paper Cattell clearly emphasized the necessity for standardized procedures.

In the next year Jastrow reported the results of another series of tests, mainly of sensation and bilateral movement.[8] He had given these tests to

[5] From Galton's remarks on J. McK. Cattell's paper, *Mental Tests and Measurements*, Mind, 1890, 15, 380.

[6] For an account of the early work, see J. Peterson, *Early Conceptions and Tests of Intelligence*, 1925, chaps. V and VI.

[7] Cattell, *op. cit.*, 373–380.

[8] Peterson, *op. cit.*, 82–83.

college students, but two years later he attempted to extend his research to other groups, setting up a psychological laboratory at the Columbian Exposition in Chicago. There, as in the laboratory Galton had established in the South Kensington Museum, anyone interested could take the tests and determine his position in relation to the standards which were being established. Again the largest number of the tests were of sensory functions, but there were in addition studies of reaction time and memory. The results of this work at the Exposition should have been interesting, but they were never reported.

In the year of the Exposition, Gilbert published a paper which is important for the study of adult intelligence even though the subjects were children, for the results showed a constant improvement with age up to maturity, at which time the advance seemed to be retarded.[9] This was apparently the first of the studies of mental development. The improvement from year to year through childhood undoubtedly confirmed the general opinion that the simpler functions were valuable indices of mental ability, but the slowing up of the progress at maturity did not seem to suggest that these functions were of little significance after a certain level of development.

In 1896 Cattell published another important paper, written in collaboration with Farrand.[10] The tests then reported were approximately those Cattell had outlined in 1890, with a few additions, and the subjects were again college students. The results interested the authors much less than the methods. They wanted to describe their system of physical and psychophysical as well as mental measurement so that other laboratories could repeat some of the tests, and could also supplement the existing knowledge of individual differences by making experiments on functions and traits which Cattell had not included in his survey. Galton's idea of "sinking shafts" at a few critical points had apparently fallen before the desire for knowledge about all varieties of individual differences, physical as well as mental. Such an extensive program required cooperation, and Cattell, who had been appointed chairman of a committee on mental tests at the previous meeting of the American Psychological Association, undertook to secure that cooperation. Despite the criticisms which Binet and Henri[11] made of his tests,

[9] *Ibid.*, 84.

[10] J. McK. Cattell and W. Farrand, *Physical and Mental Measurements of the Students of Columbia University*, Psychol. Rev., 1896, 3, 618–648.

[11] A. Binet and V. Henri, *La psychologie individuelle*, Année psychol., 1895, 2, 411–465.

he was apparently successful in convincing American psychologists of their value; and for a few years longer the Americans concentrated largely on tests of the simpler functions while French and some of the German investigators worked to secure adequate tests of the more complex processes. It is interesting to speculate what the Americans would have done if they had been faced by the practical problems which confronted Rieger, or Ebbinghaus, or Binet, and whether the Americans too would have turned to the more complex test situations, even though these involved comparatively unknown psychological processes.

It is not to be supposed that investigators in the United States ignored the French point of view entirely. Sharp,[12] working in Titchener's laboratory, experimented with the tests Binet and Henri[13] had suggested in 1895, using advanced students in psychology as subjects. In general her findings were not encouraging; she reported that the positive results of her work had been "wholly incommensurate with the labor required." At the same time she made the interesting discovery that the more complex of the tests she used, the tests of imagination and of the description of pictures, were important in differentiating between individuals, even in the case of the individuals of her homogeneous group. Her experience with the tests as a whole led her to agree with Binet and Henri "that individual psychical differences should be sought for in the complex rather than in the elementary processes of mind, and that the test method is the most workable one that has yet been proposed for investigating these processes."

The construction of tests for complex processes was of course a slow and tedious procedure. Wissler in 1901 reported that it was proposed to add tests of attention, suggestion, and apperception to Cattell's series, and that the supplying of missing words in a paragraph had been tried as a test of apperception; but these tests were still "in the nature of an experiment."[14] The actual data he had to report were the results of the application of Cattell's earlier tests to 267 Columbia and 42 Barnard students, and the intercorrelations of these. Not only were the correlations between these "laboratory mental tests" and those between the mental tests and the physical tests low, but the markings of students in college classes, while they corre-

[12] S. E. Sharp, *Individual Psychology: A Study in Psychological Method,* Amer. J. Psychol., 1898–99, 10, 329–391.

[13] Binet and Henri, *op. cit*, 411–465.

[14] C. Wissler, *The Correlation of Mental and Physical Tests,* Psychol. Rev., Monog. Suppl., 1901, 3, No. 16.

lated with themselves to a considerable degree, did not correlate with the mental tests. The almost complete lack of correlation was a serious blow to the laboratory mental tests, for it seemed obvious that the grades were in large part dependent upon intelligence.

Need for the development of more satisfactory tests then became even more evident, and gradually, with increasing work on the more complex processes, such tests began to appear. In 1910 Goddard revised the Binet Scale for use in the United States. In 1913 Calfee, Waugh, Bingham, Bell, Kitson, and others reported the findings for college students on complex tests such as directions, sentence completions, mixed relations, and logical memory. All this work has been well reviewed by Carothers as a preliminary to a report of her own study of 200 Barnard freshmen in 1915 and 1916.[15] She gave nineteen tests which differed considerably in their nature and complexity from those Wissler had used fifteen years before. Carothers' tests fell, according to their intercorrelations, into five groups:

I. Coordination and Tapping
II. Cancellation, Checking, Color Naming, Word Naming, and Substitution
III. Directions, Opposites, Verb-Object, Mixed Relations, Word Building, and Completion
IV. Word Recollection, Word Recognition, Logical Recollection, and Logical Recognition
V. Information and Vocabulary
Miscellaneous: Digit Span and Knox Cube

It is interesting to note that the author looked for a general common factor but found no evidence of one. Her correlations with a composite score[16] of all the groups of academic subjects were higher than Wissler's, a natural result of the types of tests used. They ranged from +.14 to +.27.

In 1918, largely as a result of the successful use of the Thorndike Test of Mental Alertness with the Student Army Training Corps, the Columbia faculty voted an alternative method of admission, to include an intelligence test.[17] From that time on intelligence tests became more and more important as supplementary criteria, and sometimes constituted the chief criterion

[15] F. E. Carothers, *Psychological Examinations of College Students,* Arch. Psychol., 1922, 6, No. 46.

[16] She called this composite score "Intelligence quotient."

[17] Wood made an enthusiastic report on the results of intelligence testing at Columbia during the next five years. B. D. Wood, *Measurement in Higher Education.*

for college admission. The Army Alpha and the Thorndike College Entrance Examination were the most widely used, particularly in the earlier years. At present some colleges have their own tests, modeled largely upon these two. The colleges which are members of the College Entrance Examination Board require for admission not only subject achievement tests but also the Scholastic Aptitude Test, the first form of which was constructed for the Board in 1926 by Brigham, Angier, McPhaill, Rogers, and Stone. This test, which at first had nine sub-tests, has been subjected to more extensive and elaborate statistical analysis than any other ever introduced, and has been continuously modified. At the present time it consists of two sections, verbal and mathematical. The verbal section contains three sub-tests: 1) Antonyms, four words, among which a pair of opposites is to be found; 2) Double definitions, a modified completion test; and 3) Paragraph reading. The mathematical section consists of 100 items.[18] The large population (8,000 to 9,000 a year) to which these tests have been given is, of course, a highly selected one, and the results show only the intellectual characteristics of the young adult of superior mental ability.

3. INTELLIGENCE TESTS IN THE ARMY

The first and largest study of adult intelligence ever undertaken was made in the United States Army during the World War. It was a radical innovation in military procedure. In the years just before the War there had been suggestions in both France and Germany that mental test methods would aid in the selection of men fit for military service. The French authorities, however, saw no merit in the scheme which Binet and Simon proposed in 1910,[19] and apparently no extensive use was made in Germany of any of the procedures suggested for determining the intellectual status of conscripts and eliminating the low grade.[20]

When the United States entered the War on April 6, 1917, a group of psychologists then meeting at Harvard appointed a committee to gather information on the part psychology could play in the preparations for war. In

[18] The tests are not available for reference, but yearly reports are published; the reports through 1931 are reprinted in C. C. Brigham's book, *A Study of Error.*

[19] A. Binet and Th. Simon, *Sur la necéssité d'une méthode applicable au diagnostic des arriérées militaires,* Annales médico-psychol., 1910, 11, 411–465.

[20] Anonymous, *The Experience of the German Army with the Defective and Feeble-minded,* J. Psycho-Asthenics, 1911, 16, 68–76.

the early summer a group mental examination was arranged by Yerkes, Bingham, Goddard, Haines, Terman, Wells, and Whipple, and tried out with such dispatch that by August first a report embodying the plan for extensive testing and presenting the results of the trial examinations was in the hands of the Surgeon General. A detailed review of the problems involved has been presented by Yerkes in the fifteenth volume of the Memoirs of the National Academy of Sciences.

Between September 1917 and January 1919, 1,726,966 men were examined.[21] Most of them were given the Army Alpha, some the Beta, and some were also given a series of individual tests—the Stanford Binet and a long series of performance tests. The tremendous number of records obtained on the Alpha Test makes it of greatest interest for any study of adult intelligence. The test in its final form was composed of the following sub-tests, the origins of which are also given below:[22]

1. Oral directions: adapted from Abelson, Otis, Woodworth, and Wells.
2. Arithmetical reasoning.
3. Practical judgment: adapted from Binet "comprehension questions" and Bonser "selective judgment" tests.
4. Synonym-antonym: Otis test.
5. Disarranged sentences: Otis adaptation of the Binet test.
6. Number series completion: adapted from Rogers' missing number test.
7. Analogies: Otis, Bingham, and Thurstone form of test.
8. Information: suggested by Wells, Bingham, and Whipple.

The distribution of intelligence in the United States, as indicated by the Army results, showed a marked positive skew, a fact which has led to controversy as to whether the true distribution of intelligence, which is apparently normal for children, was actually skewed in adult groups, or whether the results depended partly on factors in the sampling. Terman contended that the Army draft was not representative because, in his opinion, a large

[21] Four groups were chosen for statistical analysis: white draft, Negro draft, white officers, and a selected group in permanent organizations. These groups were selected from fifteen camps, pro-rated by states, giving 96,354 records for white recruits, 25,392 for Negro recruits, 15,528 for white officers, and 24,205 for white permanent organizations. To these records were added those of a special experimental group of 1,047 white recruits; these brought the grand total to 162,526.

[22] Mem. Nat. Acad. Sci., 1921, 15, 300–302.

The original test (Army test a) consisted of ten sub-tests. After a preliminary trial these were cut to eight sub-tests with a total of 212 items, and the tests retained were slightly revised to correct faults found in the original series.

proportion of those granted exemptions were men of superior intelligence.[23] Lincoln disagreed and presented arguments to show that the exempted groups were not superior.[24] There are undoubtedly points on either side, but it seems probable that Lincoln's position is the more justifiable and that the sample is fairly representative and on the whole the most satisfactory ever obtained.

The median Army Alpha score for the native-born white men was 58.9, while that for the foreign-born white was 46.7, and that for the Negro considerably lower (38.6 for Negroes from northern states and 12.4 for Negroes from southern). On the basis of Stanford Binet Tests of 653 men, Army Alpha scores were transmuted into Stanford Binet equivalents. The central result of this analysis, namely, that the intelligence of the white draft was not much above a thirteen-year level, is too well known to necessitate restatement. It has not only been discussed pro and con by psychologists, but has been seized upon by popular writers who have made good copy of the intellectual level of the average man. The height of the popular indignation was reached in Walter Lippmann's papers in the New Republic in 1922 and 1923.[25] They expressed many criticisms of the mental testing movement which psychologists in general would reject, but they sounded a note of caution which psychologists as well as laymen were finding extremely necessary, namely that the mental-age method of expressing intelligence test results was likely to be misleading when applied to the scores of adult subjects.

Terman contended that the group of 653 men given the Stanford Binet Scale was not representative of the entire draft, while Lincoln found that it had a distinctly higher percentage of A and B men (as determined by Alpha scores). The question is of little importance in relation to that of the character of the draft as a whole, for there is nothing to be gained by transmuting Alpha into Stanford Binet Scores, but it seems clear that the groups on which the Army standards were established were more nearly representative of the population at large than those on which the Stanford Binet norms were based.

This point was clearly made by Symonds in an article in which he re-

[23] L. M. Terman, *Mental Growth and the I. Q.*, J. Educ. Psychol., 1921, 12, 325–341.

[24] E. A. Lincoln, *The Mental Age of Adults*, J. Educ Res., 1922, 6, 133–144.

[25] W. Lippmann, *The Mental Age of Americans*, New Republic, 1922, 32, 213–215, 246–248, 275–277, 297–298, 328–330; 1923, 33, 9–10.

ported a second approximation to the curve of the distribution of intelligence of the general population.[26] Assuming the normality of the distribution in each occupational group, he weighted the distributions for different groups in the Army according to the numbers allocated to each by the 1910 census. He found that the total curve thus obtained was strongly skewed to the right; and that "the numbers in low intelligence occupations . . . outweigh the numbers in high intelligence occupations." His final Alpha median was 48, and sigma 43. These figures corresponded almost exactly to the figures obtained for the Army data when medians for native-white, foreign-born, and Negro groups were weighted and combined. Symonds' comments are particularly arresting:

> Where shall we go for the average man? The average man has an Army Alpha score of about 48. Fryer's table [see page 48] shows that representative occupations around this level are masons, hospital attendants, station agents, miners, teamsters, riggers, boilermakers, airplane workers, factory storekeepers, horse shoers, salesclerks, hostlers, barbers, stationary engineers, cobblers, horse trainers, caterers, bricklayers, auto truck chauffeurs, farmers, concrete workers, printers, and bakers. To obtain representative figures children and adults should be tested coming from social groups of which the above listed occupations are typical, not high school students and business men. There are as many in the population who are less intelligent than the average semi-skilled workmen in the occupations above as there are of those who are more intelligent. Every person who wishes to obtain norms or standards representative of the total population can do no better than to consider carefully the occupation groups in which he proposes to do his testing. There seems to be no better ready criterion.

Another important study based on the Army material, and appearing in the same year as Symonds' article, is Brigham's analysis of the mental levels in the various racial groups.[27] From the point of view of this research the interest lies not so much in the actual findings for the different racial groups then compared, as in Brigham's later criticism of the work and of the Army Alpha Test in general.[28] This criticism is twofold, the first point being that a tetrad analysis of the sub-tests shows disparate group factors and the second that scores in different sections of the Alpha Test were derived from different sub-tests. For these reasons Brigham concluded that the

[26] P. M. Symonds, *A Second Approximation to the Curve of the Distribution of Intelligence of the Population of the United States, with a Note on the Standardization of the Binet-Simon Scale,* J. Educ. Psychol., 1923, 14, 65–81.

[27] Brigham, *A Study of American Intelligence.*

[28] Brigham, *Intelligence Tests of Immigrant Groups,* Psychol. Rev., 1930, 37, 158–165.

scores in the eight sub-tests should never have been added to obtain a total score.

The reaction of the Army examiners to the fact that the different sub-tests did not contribute proportionately to the score at different levels was just the reverse of Brigham's.[29] They stated:

> The alpha examination itself is not a homogeneous scale, but a composite of eight short scales, each much more nearly homogeneous than the composite total alpha. Alpha must be treated as a composite for two reasons: (1) Certain of its component tests are very much more difficult than others, so that examinees of less than average intelligence do not in general register their ability in all of the tests. (2) Owing to the fact that the total time allowed for the alpha examination is rigidly apportioned among the eight tests, . . . the collection of the eight tests into an "examination" is mainly a matter of administrative convenience, with relatively little implication of quantitative or qualitative equivalence. It follows from what has just been stated that subjects who are able to score in all of the alpha tests obtain total scores that are not comparable with total scores earned by individuals who failed in some of the tests. The reason for this difference is that in obtaining an individual's total score by adding together the scores in each test, scores of 0 are treated arithmetically the same as other scores. . . .
>
> The fact that zero points are disposed at different levels of intelligence for different tests would be of no consequence if negative scores could be registered; but since they can not be, the individual who fails to earn a positive score and is marked zero is actually thereby given a bonus varying in value directly with his stupidity.

It should be admitted that in general practice, similar scores have been taken to represent similar performances on the tests, but it is difficult to see how this situation could invalidate the constants of a frequency distribution obtained from large numbers of cases.

As to Brigham's first criticism, that there are disparate group factors in the test, one may point out that "g" from a statistician's point of view and general intelligence from a clinician's point of view are different concepts. Brigham stated the situation clearly when he wrote:[30]

> Factors have figured so largely in the work of Spearman, Kelley, and others that it is important to understand their nature. Factors are not to be regarded

[29] Mem. Nat. Acad. Sci., 1921, 15, 622.

In determining interrelationships between the tests, the effects of the piling up of zero scores in certain of the sub-tests were discounted by statistical procedures which redistributed these cases in accordance with the normal correlation surface. This procedure, of course, did not adjust the total score for any individual.

[30] Brigham, *A Study of Error*, 38.

as psychological realities, or existences, but merely as certain decimal multipliers which in a given system might be used to account for certain correlation coefficients. One should hunt for factors in the realm of roots, relations, and postulated existences and not in the central nervous system or the introspective report. The materials from which factors are derived are coefficients of correlation between tests.

The flaws in the Alpha Test and the difficulties in its interpretation are not sufficient ground for throwing the Army results out of court. They are not suited to many of the types of analysis one might wish to make, but they still represent the most satisfactory, and indeed the only index of the intelligence of the American population at large. The importance of the Army Alpha has recently been emphasized by Wells, who revised the test for the Psychological Corporation.[31] He expressed his belief that the old norms were still the best available and so arranged his revision that the new test would be directly comparable to the old. His views on the character of the test are evident in the following statements:

The numerous and active progeny which have issued from Army Alpha during its decade and a half of existence, may find the informed reader somewhat puzzled as to the reason for its rejuvenation. The reasons center around the considerations that for all its technical defects it is still much the most widely standardized of "intelligence" tests, depending for this fact on considerations which it is to be hoped will never be reproduced; it is also the most widely known, if not always the best understood with reference to superior educational and vocational groups, and the only one readily available in several alternate forms of substantially equal difficulty. . . .

The evidence is striking that the relative difficulties with the Army Alpha, even in its present cumbrous and incongruous, often laughter-provoking state, are still those of administration rather than of meaning.

4. INTELLIGENCE TESTS IN INDUSTRY

The field of industrial psychology has much to contribute to a knowledge of certain skills, types of behavior, and special abilities at adult levels, but less to contribute on the problems of adult intelligence. Except for a short period just after the War, intelligence tests have been little used.[32] They have been shown to be of value in establishing the limit of intelligence below which success in the particular occupation is decidedly improbable, and

[31] F. L. Wells, *Army Alpha, Revised,* Person. J., 1932, 10, 411–417.

[32] A. W. Kornhauser and F. A. Kingsbury, *Psychological Tests in Business,* 1924.

in some cases they have established a limit above which interest and satisfaction in the job are unlikely.[33] Except perhaps in clerical occupations they have shown low correlations with success in work. On the whole their uses have not been of sufficient immediate value to lead to extensive surveys, and the findings available are on small groups, selected in one regard or another, and not satisfactory material from which to draw conclusions as to the level or characteristics of mental functioning in the adult.

The wide range of test intelligence found among individuals in one type of occupation was evident in the Army data. Fryer's study of this material summarized the findings for ninety-six occupations, giving interquartile score ranges as well as median Army Alpha scores for each.[34] A similarly wide range has been found in a recent investigation by Cattell, in which he attempted to standardize an intelligence test for adults.[35] He listed twenty-five occupations from which he had examined 12 or more members; in spite of the fact that the numbers in some of these groups were small and that his test[36] differed somewhat in nature from the Alpha, there is a fairly close correspondence between his listing of occupations in order of intellectual level and Fryer's listing.

5. SEX DIFFERENCES IN INTELLIGENCE TEST PERFORMANCES

From the time of Thompson's[37] pioneer work until recent years, almost all studies of sex differences in intelligence had been based on the findings for children. Garrett summarized these as follows:[38]

Most investigators have found the average boy to be superior to the average girl in tests involving numerical or mathematical relationships, as well as in the ability to employ spatial and geometric concepts. Girls, in turn, are usually su-

[33] For studies in this field, see:
K. M. Cowdery, *Measures of General Intelligence as Indices of Success in Trade Learning,* J. Appl. Psychol., 1922, 6, 311–330.
E. Burr, *Minimum Intellectual Levels of Accomplishment in Industry,* J. Person. Res., 1924, 3, 207–212.
For a summary discussion, see M. S. Viteles, *Industrial Psychology,* 121–127.

[34] D. Fryer, *Occupational-Intelligence Standards,* Sch. & Soc., 1922, 16, 273–277.

[35] R. B. Cattell, *Occupational Norms of Intelligence and the Standardization of an Adult Intelligence Test,* Brit. J. Psychol., Gen. Sec., 1934–35, 25, 1–28.

[36] The sub-tests were synonyms, classification, opposites, completion, and inferences—the types of performance which had, in Cattell's opinion, been shown to correlate most highly with "g" without overlap.

[37] H. B. Thompson, *The Mental Traits of Sex,* 1903.

[38] H. E. Garrett and M. R. Schneck, *Psychological Tests, Methods, and Results,* Part II, 198.

perior to boys in tests requiring memory and in the ability to employ language relationships quickly and precisely . . . the consistency of the results reported by different investigators lends weight to the belief in the existence of a true difference.

In a study of over 2,000 school children Pressey found that the girls' lead in verbal material was gradually lost with age, but the boys were superior in arithmetic at all ages.[39] A number of other studies have shown, however, that the girls' superiority in verbal material persists at the college level.

The most conclusive recent evidence is that from the results of the Scholastic Aptitude Test. Each year since the mathematical section was added, the girls have been superior in mean scores on the verbal and the boys on the mathematical portion of the test.[40] In most years the population was about 8,000 and the differences in means were clearly significant. It is to be noted that Brolyer considers the group of girls more highly selected than the group of boys, a condition which would tend to produce a difference smaller than the true one in mathematical material and larger in verbal.

Evidence of reliable differences is not so clear among adults. In his study of family resemblances, Carter found that sex differences in adult scores were slight in comparison with individual differences.[41] It is to be noted, however, that the difference divided by the sigma of the difference for the computation tests, on which the men surpassed the women, was 7.74, while this ratio for the vocabulary test, on which the women were superior, was 1.00.

Conrad and Jones, whose important study of adult intelligence has already been mentioned, found slight superiority on the part of the females; it was greater during adolescence than in later years but in their opinion not sufficient to be of any practical importance.[42] The males were rather consistently inferior on those four tests of the Army Alpha which the authors thought most "strongly verbal" (common sense, opposites, disarranged sentences, and analogies). The differences on the other tests are reported in the following statements:[43]

[39] L. W. Pressey, *Sex Differences Shown by 2,544 School Children on a Group Scale of Intelligence, with Special Reference to Variability,* J. Appl. Psychol., 1918, 2, 323–340.

[40] C. R. Brolyer, *Fifth–Ninth Reports of the Commission on the Scholastic Aptitude Test,* 1930–1934.

[41] H. D. Carter, *Family Resemblances in Verbal and Numerical Abilities,* Genet. Psychol. Monog., 1932, 12, 1–104.

[42] *Op cit.,* 265-266.

[43] Conrad, Jones, and Hsiao, *Sex Differences in Mental Growth and Decline,* J. Educ. Psychol., 1933, 24, 166.

In test 1 (oral directions) the males are in early adolescence as inferior as in the more highly verbal tests; but at later ages they acquit themselves relatively much better, overtaking the females in the period of middle maturity, and again dropping slightly behind in later maturity. Test 8 (information) is one which has commonly been regarded as favoring males. . . . Our obtained sex differences are relatively slight in this test, and with frequent intersections of the growth curves. In test 6 (number series completions) a similar condition occurs. . . . [In 2] (arithmetic problems) we find the only instance in which males are fairly consistently superior to females. The male superiority does not clearly emerge until later adolescence. . . .

Small differences, varying in direction according to the nature of the material and dissimilarities in interest and practice, were also found for the moving picture test which Conrad and Jones reported.[44]

Miles and Miles discovered irregular trends in the average scores for men and women of successive decade groups on a fifteen-minute form of the Otis Self-Administering Tests of Intelligence, but they were inclined to regard the differences as "possibly due to errors in sampling."[45] The men were superior in most of the age groups. When comparisons were made between five large-range age groups, the men were superior in four instances and in two of these the differences were great (with a difference divided by the sigma of the difference of 3.58 for ages fifteen to twenty-nine and 2.61 for ages fifty to sixty-nine). It is to be noted that these comparisons are based on an omnibus test, and differences between the sexes in particular types of performance, such as Jones and Conrad found, may thus have been hidden. Other parts of the Stanford Maturity Study, that is, investigations of other activities besides those involved in the Otis Test, have shown few significant or large sex differences, however; and most of these investigations are of specific types of activity in which differences between groups should have been detected if they had existed.[46]

6. THE GROWTH AND DECLINE OF INTELLIGENCE

The attention of psychologists, particularly of those concerned with the construction or interpretation of mental tests, has long been focused on the problem of the growth of intelligence.[47] It has been attacked by comparing

[44] Conrad and Jones, *Psychological Studies of Motion Pictures. V. Adolescent and Adult Sex Differences in Immediate and Delayed Recall*, J. Soc. Psychol., 1931, 2, 433–459.

[45] *Op. cit.*, 64.

[46] See pages 31–32.

[47] For discussion of the earlier contributions, see F. N. Freeman, *Mental Tests*, 1926, chap. XIII, and F. D. Brooks, *The Psychology of Adolescence*, 1929, chaps. IV and V.

the test performances of different groups of children at successive ages, and by retesting the same groups of children at an interval of one or more years. There has been careful work on sex differences in mental development, on the relations between physical and mental growth, and on many other aspects of the problem. The question of the limit of mental growth has of course been a part of the developmental studies of children, but until recently no studies were continued through successive ages beyond adolescence. Consequently, it was never possible to determine the actual course of any test performance with age. The only data available were the Army findings, with different age groups which were probably not comparable, and a small number of studies of special groups of adults—firemen, policemen, prisoners, professors, and various other types of employed and unemployed.[48] None of these data demonstrated the peak of development or the rate or characteristics of the decline beyond maturity.[49] The first important work on these problems began in the late 1920's with Thorndike's experiments on adult learning, Willoughby's study of family resemblances, and Jones and Conrad's investigation of the relation between age and test performance in a large group of adults. The results of this work must be reviewed in some detail, as well as the findings of Miles' Stanford Maturity Study.

[48] Most of these studies are on selected groups, which are difficult to locate in the population at large and will not be discussed in this chapter. If a sufficient number of such groups were studied by comparable methods, they would, of course, be valuable as component parts of a larger population which could be weighted to approximate that of the census figures

For examples of these special studies, the reader is referred to the following articles:

H. H. Caldwell, *Adult Tests of the Stanford Revision Applied to University Faculty Members,* J. Exper. Psychol., 1922, 5, 247–262.

R. Pintner and A. Toops, *A Mental Survey of the Population of a Workhouse,* J. Delinq., 1917, 2, 278–287.

L. M. Terman, et al., *A Trial of Mental and Pedagogical Tests in a Civil Service Examination for Policemen and Firemen,* J. Appl. Psychol , 1917, 1, 17–29.

F. D. Fry, *The Correlation of the Reverse Audito-Vocal Digit Memory Span with the General Intelligence and other Mental Abilities of 308 Prisoners of the Eastern State Penitentiary of Pennsylvania,* Psychol. Clin., 1931, 19, 156–164.

[49] There were a few early studies of senile decline, notably those by Beeson (*Intelligence at Senescence,* J. Appl. Psychol., 1920, 4, 219–234) and Foster and Taylor (*The Applicability of Mental Tests to Persons over Fifty Years of Age,* J. Appl. Psychol., 1920, 4, 39–58); but the application of the findings has been limited by the fact that the groups were selected. Interestingly enough G. Stanley Hall did not mention these two papers in his large volume on *Senescence* in 1922, perhaps because he did not consider the existing mental tests applicable to adults. He wrote: "There are no mental tests of generally recognized validity above the teens, so that we have no criteria for determining psychological age for even the elderly . . ." (page 196).

Before discussing the data on mental growth or decline, however, it is first of all necessary to note that any findings must be interpreted in the light of the type and difficulty of the material the test presents and of the scaling method employed. The results should be determined separately for different types of test, and the test should always be suited to the mental maturity of the subject. The scaling method must, of course, be independent of assumptions as to the nature of mental growth.

The mental-age method of scaling, with the linear relationship between mental and chronological age, is therefore unavailing in the study of mental growth. The early investigators, when they did not report results in terms of mental age, reported raw scores. Thurstone and Ackerson,[50] in connection with a report of Binet Tests scaled by Thurstone's method of absolute scaling[51] note that the raw score is also an unsatisfactory unit of measurement for it is probably not the same in different parts of the scale. The studies based on a method of absolute scaling are few, however, and of comparatively recent date.

One of the most widely quoted early studies is Teagarden's.[52] She tested 408 children over twelve and a half years of age at Mooseheart, a home maintained by the Loyal Order of Moose for orphans of its deceased members. She considered the group low-average in relation to the population at large, but relatively unselected up to eighteen years. Of the results reported, those for the Stanford Binet Scale are unsatisfactory because of the mental-age scaling, but those for the Army Alpha furnish interesting material for further analysis. Teagarden used raw and T-scores for each half-year group, preferring the T-scores although the results for the raw and T-scores did not differ markedly. Sigmas were also given for the raw scores, however, so that these are more satisfactory for analysis.

Teagarden used the average-of-three method of smoothing her curves and thus eliminated the irregularity of the progression from one half-year group to the other which is evident when the series of medians is surveyed.[53] None of the differences between medians is statistically significant. By taking twelve instead of six months as the unit, so as to increase the number of cases in each group, the irregularities may be eliminated and the general trend brought out more

[50] L. L. Thurstone and L. Ackerson, *The Mental Growth Curve for the Binet Tests,* J. Educ. Psychol., 1929, 20, 569–583.

[51] Thurstone, *A Method of Scaling Psychological and Educational Tests,* J. Educ. Psychol., 1925, 16, 433–451.

[52] F. M. Teagarden, *A Study of the Upper Limits of the Development of Intelligence,* Teach. Coll. Contri. to Educ., 1924, No. 156.

[53] If the distributions were, as Teagarden believed, approximately normal, means and medians would not differ appreciably.

clearly.[54] The new figures are shown in Table I. The reliability of the differences between the medians (here treated as means) for succeeding years is also given.

The difference between the medians which most nearly approaches statistical reliability is the rise between the groups with midpoints at fifteen and sixteen years. There is a slight rise in median score between sixteen and seventeen, but none at all between seventeen and eighteen.[55] These figures would indicate an

TABLE I

THE RELIABILITY OF DIFFERENCES IN MEDIAN ARMY ALPHA SCORES BETWEEN SUCCESSIVE ONE-YEAR GROUPS (TEAGARDEN)

AGE	N	MEDIAN	SIGMA	$\frac{D}{\sigma_D}$	CHANCES IN 100 OF A SIGNIFICANT DIFFERENCE
12–6 to 13–5	87	64	38.40		
				0.94	83
13–6 to 14–5	93	69	32.18		
				1.60	94
14–6 to 15–5	76	78	39.62		
				2.80	99.7
15–6 to 16–5	60	97	39.18		
				1.08	86
16–6 to 17–5	42	106	42.61		
				0	0
17–6 to 18–5	31	106	44.93		

earlier cessation of growth than Teagarden reported, a cessation between sixteen and seventeen rather than between seventeen and eighteen. They would also indicate a positively accelerated curve between thirteen and sixteen, whereas Teagarden concluded that this curve was negatively accelerated. It is hard to say how much of the negative acceleration she found depended on the average-of-three method of smoothing the curve.

The analysis Thurstone and Ackerson made of the results of Binet Tests scaled by the method of absolute scaling showed a positively accelerated curve up to the age of about eleven years with negative acceleration there-

[54] Age groups above eighteen were omitted because they were small and probably selected.

[55] An analysis of the Stanford Binet along the same lines as this analysis of the Alpha casts some doubt on Teagarden's conclusion that "the stopping place in the curve is reached earlier in Alpha than in the other tests" (page 62), since the most reliable rise in the Binet occurs a year earlier than that in the Alpha.

after.[56] The population they studied was slightly retarded, and more variable than the average; the subjects were 4,208 white children, three to seventeen years of age, entering the Institute for Juvenile Research. The growth curve obtained was asymptotic to absolute zero and asymptotic to the level of test maturity. The authors noted that they had of course no rational procedure for locating the level of maturity according to the Binet Test as for locating the absolute zero. They found that the brighter children reached the inflection point of about eleven years sooner than the dull, and that they reached test maturity sooner than the dull. The authors stated that: "The absolute variability of test intelligence increases with age until adult intelligence is attained."

Richardson and Stokes, on the basis of results obtained for the Simplex Junior Intelligence Scale given to 12,000 children of ages six to fourteen, the total population between these ages in one town in England, supported Thurstone's conclusions as to variability.[57] They found that the absolute variability of intelligence increased with age, and that the spread of intelligence at any age was proportional to its mean level at that age. Consideration of the growth curve indicated that the sigma divided by the mean was a constant equal to about .18. The mean growth curve obtained was an almost perfect G curve, with the inflection point at about four years.[58] Richardson noted that "from 6 years onward we are dealing with the upper half of the developmental curve which has reached by 11 years a level of over 80% of maturity. Hence the differences we are accustomed to measure with our tests, between one year and another, and one individual and another, will often be quite small when viewed against the background of the absolute scale." He speculated as to the probable form of the growth curve below and above the ages examined, and suggested with regard to the upper end that development probably reached 96 per cent by eighteen years and 99 per cent by twenty-four.

Another study by means of the method of absolute scaling has been carried out by Odum.[59] His data were also very extensive and based on the re-

[56] *Op. cit.*, 569–583.

[57] C. A. Richardson and C. W. Stokes, *The Growth and Variability of Intelligence*, Brit. J. Psychol., Monog. Suppl., 1933, 18, 1–83.

[58] The growth curves for the separate items were also G curves, but most of them showed a slight kink somewhere between nine and a half and thirteen and a half.

[59] C. L. Odum, *A Study of the Mental Growth Curve with Special Reference to the Results of Group Intelligence Tests*, J. Educ. Psychol., 1929, 20, 401–416.

Different tests were used at different ages, the total series including the Dearborn Scale, Series I and II; the Otis Group Scale, Primary and Advanced; the Illinois Group Intelligence

sults of group tests. His population ranged in age from five to nineteen years. Odum found that the growth curve showed negative acceleration, and that the variability in the case of each test increased as chronological age increased. He concluded that the ability of children to score on group intelligence tests did not "stop growing before the age of 17, and very likely not until a later age."

The crucial ages from fourteen to eighteen have been the subject of special study as a part of Thorndike's work on mental growth. In 1922 and 1923 he gathered material on retests of 8,000 high school students, grades nine to eleven and ages fourteen to eighteen, and, after allowing for the effect of practice, found evidence of a steady increase in score from age to age. He concluded that any decrease in gain with age was offset by increasing selection of the group.[60] To study the influence of age freed from the influence of selection, he later compared the gains made in a year by a large number of thirteen-year-olds from the ninth grade and fourteen-year-olds from the tenth, the basic assumption being that about the same amount of ability was required to be in the tenth grade at fourteen as to be in the ninth at thirteen.[61] The results showed a slightly larger gain in the thirteen-year group.

Thorndike is one of those who have most emphatically contended that mental growth as indicated by test performances continues beyond adolescence. In *The Measurement of Intelligence,* he stated:[62]

> Neither in our CAVD results, nor in the National-Otis-Haggerty estimate, nor in Brooks' results is there any justification for the doctrine that the gain in altitude of intellect of the sort measured by existing intelligence tests is zero after 14, or after 15, or even after 16. It decreases, but it should not become inappreciable until 18 or later. According to our results the decrease from 14 to 18 is not an

Scale; and the National Intelligence Test, Scale A. The numbers taking each of these tests ranged from 2,000 to 40,000.

[60] E. L. Thorndike, *On the Improvement in Intelligence Scores from Fourteen to Eighteen,* J. Educ. Psychol., 1923, 14, 513–516.

[61] Thorndike, *On the Improvement in Intelligence Scores from Thirteen to Nineteen,* J. Educ. Psychol., 1926, 17, 73–76.

[62] Thorndike, *The Measurement of Intelligence,* 467.

It is to be noted that Thorndike suggested that there was a "differential gain from the age of fourteen to twenty and beyond, whereby some individuals increased these abilities very greatly, whereas others increased them little or not at all." On this basis he would account for the change from symmetry and normality characterizing the distributions of test intelligence for children to the positive skewness shown in the distribution of the sort of intellect measured by Alpha for the adult population. (Page 293.) See page 8.

abrupt slowing up of a gain that has been steady hitherto, but is part of a general negative acceleration which began long before the age of 6½.

The type of curve Thorndike described continued to rise longer than that found by Thurstone, a difference which may well depend on the difference in discriminative power between the Binet and the CAVD tests for the older subjects. At the younger ages Thorndike's curve is not unlike that obtained by Richardson and Stokes: both showed positive acceleration in the very early years, followed by negative acceleration. Thurstone would place the inflection point later, about eleven years. Odum found no inflection point between five and nineteen and described a negative acceleration throughout this age period.

Thorndike's studies of learning also indicated a limit of growth which fell considerably beyond early adolescence. He set the age somewhere near twenty.[63] As to the characteristics of learning in the years beyond twenty, he concluded:[64]

> The general tendency from all our experiments is for an inferiority of about 15 per cent as a result of the 20 years from twenty-two on. Learning representing an approximation to sheer modifiability unaided by past learning shows considerably more inferiority than this. Actual learning of such things as adults commonly have to learn shows considerably less. . . .
>
> Adults may be expected to learn harder things than they could have done in the years from six to sixteen and to learn the same things in ways requiring greater powers of abstraction and reasoning.

One of the first studies of adult intelligence after the Army survey was the investigation Willoughby made in 1927, which had as its immediate purpose an analysis of family resemblances.[65] The families selected for examination were those having children aged thirteen in the schools around Stanford University. The territory is probably a superior one, and a further factor in the selection of the group is that almost 50 per cent of the families refused to cooperate; these may have included a larger proportion of dull subjects. While the test levels probably should not be taken as representative of the population at large, the findings on the course of various test performances with age are of great importance. The peaks of ability on the different tests ranged from seventeen to twenty-seven years:

[63] Thorndike, *Adult Learning*, 129. [64] *Ibid.*, 106 and 154.

[65] R. R. Willoughby, *Family Similarities in Mental-Test Abilities*, Genet. Psychol. Monog., 1927, 2, 239–277.

Type of test	*Source of test*	*Peak of ability*
Vocabulary	Stanford Achievement	25
Science-nature information	Stanford Achievement	25
History-literature information	Stanford Achievement	22
Number series completion	Army Alpha and Beta	18
Form combination	Army Alpha and Beta	19
Digit-symbol substitution	Army Alpha and Beta	18
Arithmetic reasoning	National Intelligence	20+
Opposites	National Intelligence, with the addition of harder items	27
Symbol series completion	National Intelligence, with the addition of harder items	17
Verbal analogies	National Intelligence, with the addition of harder items	18
Comparison	National Intelligence, with the omission of pictorial elements	23

It is evident that the tests showing the earliest peaks were the number and non-verbal tests and also the analogies, while those showing the latest peaks were tests of vocabulary and information.

The tests were differentiated not only with regard to peak of development, but also with regard to two other characteristics of the growth curves: the extent of senescent decline and the sharpness or roundness of the peaks.[66] A decline which carried the curve back to pubertal or prepubertal levels appeared in the case of the opposites, the number and the symbol-series completion, the analogies, the symbol-digit substitution, the history-literature information, and the comparison. There was less decline for the vocabulary, the form-combination, and the science-nature information tests, and least of all for the arithmetic reasoning. The two tests showing the sharpest peaks were number-series completion and analogies, while that showing the roundest was arithmetic reasoning. In this connection Willoughby suggested that the sharp peaks were probably to be found for test abilities "acquired more or less as a matter of normal growth and comparatively independent of school training, while a long, rounded peak would seem to be the result of laborious and long-continued effort against obstacles."

The next important study of adult intelligence after Willoughby's was that made by Jones and Conrad in New England.[67] It has already been de-

[66] Willoughby's curves are continued to sixty years.

[67] *Op. cit.*, 223–298.

scribed as unique among studies of adults because it represents a survey of a complete or almost complete community. Possible adult resistance to mental testing was overcome by propaganda and the offer of a free movie. Cases obtained in this way were supplemented by house-to-house testing. It is possible that there was some selection in the age groups over forty, for there was an increasing number of refusals beyond this age, but on the whole the population is one of the most satisfactory ever obtained, and one of the few providing trustworthy indices of the typical adult test performances at different ages.[68] The only tests given, however, were a test on the movie and the Army Alpha, and consequently the extent of the data on adult test performances is rather limited.

The general characteristics of the growth curve on the total Army Alpha for the 1,191 subjects, aged ten to sixty, were found to be "linear growth to about 16 years, with a negative acceleration beyond 16 to a peak between the ages of 18 and 21."[69] Following this peak there was a decline "which is much more gradual than the curve of growth, but which by the age of 55 involves a recession to the 14-year level."[70] The difference between test lev-

[68] The figures are, of course, representative only for a geographical area similar to the one surveyed, where about 87 per cent of the population came from old New England stock.

The occupational distribution was similar to that of the Vermont Census for 1920 the agricultural groups, however, were more heavily weighted in Jones' sample.

[69] Jones and Conrad noted that this growth curve was similar to that reported by Teagarden and discussed above They found close agreement in the actual scores up to sixteen, and suspected that Teagarden's groups were selected beyond this age (*Op cit*, 239, 241)

[70] *Ibid*, 239.

Hsiao reported the same data in an M A thesis at Columbia in 1927, *Performance of Army Alpha as a Function of Age* There are some discrepancies between his data and those given by Jones and Conrad The number of cases was the same for the two studies in only two age groups, the medians the same in only three There are differences in the growth curves which depend on the fact that Hsiao gave equal distances on the base line to each step in his distribution, while Jones and Conrad allowed for the unequal steps. Furthermore, Hsiao smoothed his curves by the average-of-three method, whereas Jones and Conrad took as original data the average of the mean and the median and then smoothed the curves by eye

Hsiao reported the reliability of the differences between total Alpha scores in successive age groups. The differences between the successive year groups from eleven to sixteen are nearly reliable in four instances and reliable in one, that between thirteen and fourteen. In the adult period, none of the differences is reliable, but the decline for the group at forty-five to forty-nine as compared with that at forty to forty-four is almost reliable There are rises, although neither is reliable, in the semi-decade thirty-five to thirty-nine and again in the semi-decade forty to forty-four. In all the sub-tests except Test 7 the medians for the age range forty to forty-four are higher than those for the age range thirty-five to thirty-nine.

The rise in the period forty to forty-four also appeared in Mursell's examinations of prisoners by the Kuhlmann-Anderson and Kuhlmann-Binet Tests (*Decrease in Intelligence with Increase in Age among Inmates of Penal Institutions*, J. Juv. Res, 1929, 13, 197–203). As in the study by Jones and Conrad the rise was followed by a sharp drop after forty-five years.

els in the different age groups led Jones and Conrad to make the following significant statement:

> A distinction . . . needs to be made between peak of development, and the "mental age of adults." The latter phrase is as a matter of fact no more justifiable than would be the expression the "mental age of children." In both cases (with possibly rare exceptions) the mental age changes significantly with the progress of the years: among the children there is growth; among the adults decline.[71]

Consequently they gave no median mental age for the adult group as a whole.[72] The T-scores were, however, based on the variability within an age range of fifteen years, namely twenty-five to thirty-nine, and thus significant growth or decline within this period were in effect excluded.

The growth curves for the separate sub-tests of the Army Alpha showed considerable diversity. The most rapid declines appeared for Tests 7 (Analogies), 3 (Common Sense), and 6 (Number Completions). Jones and Conrad suggested that these should probably be considered the best tests of the Army Alpha as measures of basic intelligence, that is, the tests most free from the effect of environmental variables. On this point they found themselves in accord with Thorndike's conclusion that "age exerts its most adverse influence upon native capacity or 'sheer modifiability.'" Test 1 (Oral Directions) showed a sudden, sharp drop from the early peak, but thereafter reached a level which was well maintained till sixty. Tests 2 (Arithmetic Problems) and 5 (Dissected Sentences) showed a gradual but not marked decline with age. The two tests manifesting little or no decline with age were 4 (Opposites) and 8 (General Information). The authors

[71] Jones and Conrad discussed at some length the additional factors which might influence the declining curves. (*Op. cit.*, 254–257.) They considered the possibility of slighter motivation on the part of the older subjects, poor eyesight or hearing, the remoteness of school training, and disuse of functions, an inability to work rapidly, and a difficulty in understanding questions. Comparisons of the cases tested in the meeting halls and the cases tested at home, together with comparisons of the findings on different types of tests, some of them less influenced than others by school training, speed, etc., seemed, however, to demonstrate that the declines depended not on any of the above-mentioned factors, but on increasing age.

[72] While the median score on Army Alpha for the total native-white draft was 59, the median for the 802 cases from Vermont taking Alpha only was 67, for the 710 cases from New Hampshire 62, and for the 1,134 from Massachusetts 72. (Mem. Nat. Acad. Sci., 1921, 15, 682–683.) It is natural, therefore, that Jones' medians should be slightly higher than the total median, but it is surprising that they should be so much higher; for his groups over thirteen years old the medians ranged from 71 for the age period fifty to fifty-four to 93 for the age period nineteen to twenty-one. The sample is such a complete one that the higher medians must indicate either a community group which is actually superior to the population at large or better testing conditions or both.

suggested that these tests, because of increasing experience, should really show increased scores with age; in other words, that the sustained rise was not of the same significance as a sustained rise in tests of more clearly "basic intelligence" would be.[73]

A comparison of the curves obtained by Willoughby and by Jones and Conrad for tests similar in type of performance required shows certain differences in peak of development and rate of decline but no marked disparities. In the study made by Jones and Conrad there is a sharper decline for analogies and considerably less decline for opposites. In both, number-series completions manifest a definite drop while arithmetic reasoning and information are relatively well maintained.

A summary of the data with regard to the peak of development, or the ages at which mean scores are highest, is presented in Table II. Whenever possible tests were grouped as verbal or non-verbal, and when no such distinction was possible the test was classified as "doubtful."[74] On the basis of

[73] Jones and Conrad stated that both of these were information tests, which should probably not be included in a battery to be used for groups of different ages. It seems probable that the difficulty is one of interpretation, that both are useful tests providing satisfactory norms are established for different age levels.

Wilson recommended Test 8 (Information) as an adult test. (*Information as a Measure of Intelligence and Maturity*, J. Educ. Psychol., 1924, 15, 309–312.) Conrad, using Hsiao's data, also recommended it. (*General-Information, Intelligence, and the Decline of Intelligence*, J. Appl. Psychol., 1930, 14, 592–599.) Conrad argued that it was a test acceptable to adults, and suitable for them since their scores did not decline with age, and that it correlated fairly well with total Alpha scores. He pointed out, however, that an apparent increase in scores on Test 8 may mean an actual decline, if the older patient shows less increment than his age itself should bring about. It is to be noted, however, that Spearman found the information test unsuitable for intelligence testing (*The Abilities of Man*, 270–277, 362), and that Wells deleted it in his revision of the Army Alpha.

As to the continued use of opposites tests in the study of adult intelligence, there would probably be a greater number of ayes than nays. The test has generally been found to correlate highly with measures of so-called general intelligence. The Army examiners found the opposites test of their first scale the best in the examination "as a measure of the traits which enter into a Stanford Binet mental age." (Mem Nat. Acad. Sci., 1921, 15, 332.) A version of the opposites is one of the three sub-tests remaining in the verbal section of the Scholastic Aptitude Test.

The correlation between Test 4 and Test 8 is the highest of the intercorrelations of the Alpha sub-tests for the Army data. While Test 4 has the highest correlation with Stanford Binet, Test 8 has only the fifth highest. The high correlation between them may possibly be a spurious one, resulting from their similar correlation with age. The fact that they have similar age curves does not, of course, indicate that they are tests of closely related performances.

[74] Of the third group, the arithmetic reasoning test is clearly distinct from the verbal, and yet can hardly be classified with the tests which are in the non-verbal group. In view of Kelley's analysis of the factors in the Army Alpha, the directions test does not seem to fit in with the other verbal tests, nor are the analogies tests altogether satisfactory in that class. (*Op. cit.*, 216.)

the existing data it is difficult to judge to what extent the differences in the ages at which the peaks appear depend on inadequacies in certain of the tests, and to what extent they depend on actual differences in the maturation rate of different abilities.[75] The present results show that the peaks appear later for most of the verbal test abilities than for the predominantly non-verbal, and that the analogies test, as well as the arithmetic reasoning

TABLE II

AGES AT WHICH HIGHEST MEAN SCORES ARE FOUND IN DIFFERENT TYPES OF TESTS (WILLOUGHBY AND JONES AND CONRAD)

VERBAL		NON-VERBAL		DOUBTFUL	
Test	*Age*	*Test*	*Age*	*Test*	*Age*
		WILLOUGHBY			
Opposites	27	Number-Series Completion	18	Arithmetic Reasoning	20
Vocabulary	25	Symbol-Series Completion	18	Analogies	18
Science Information	25	Form Combination	19		
Comparison	23	Digit-Symbol Substitution	18		
History and Literature Information	22				
		JONES AND CONRAD			
Practical Judgment	18	Number-Series Completion	18	Directions	21
Synonym-Antonym	18 and 40–44			Arithmetic Reasoning	17
Dissected Sentences	19–21			Analogies	16
Information	35–39				

and possibly the directions and practical judgment, are to be classed with the non-verbal tests in regard to the early age at which the maximum mean score appears.

As to the data on variability Jones and Conrad reported that a differential growth rate in adolescence was clearly indicated by increase in sigma not only for Alpha but for every sub-test. Except in Test 7 (Analogies), they found no evidence for a differential duration of growth.

Although Thurstone and Ackerson specifically limited the application of their law of absolute variability, stating that absolute variability of test in-

[75] A further indication of differential gain appears in a study of the growth of intelligence at the college level by Rogers at Bryn Mawr. (*The Growth of Intelligence at the College Level*, Sch. & Soc., 1930, 31, 693–699.) She found the most marked gain in reading ability.

telligence increased with age until adult intelligence was attained,[76] Jones and Conrad seemed to consider the extension of the law. They stated:[77]

> It is interesting to note that fulfillment of Thurstone's law of absolute variability would require that the decline in intelligence-test scores be accompanied by corresponding decline in absolute standard deviations—implying more rapid decline of intelligence among the brighter than among the duller of the adults. Using raw Alpha scores, we have failed to obtain a general linear relationship between mean and standard deviation. Whether or not the same absence of a general relationship would be found with Alpha scores in "absolute" scaling remains to be determined.

Certainly neither common sense nor the evidence from scientific studies would indicate that the senile decline began earlier in the more intelligent.

The Stanford Maturity Study is, as already noted, somewhat less satisfactory than that of Jones and Conrad from the point of view of the sampling of the population. Subjects were secured through church groups, philanthropic societies, clubs, and lodges, and their services were remunerated by payment to the organization rather than to the individual.[78] The method was ingenious in that it aroused group interest and brought in a much larger proportion of the total community than would have been secured if subjects had been solicited individually. It is probable, however, that the groups of adults who have joined themselves to any organization, with the possible exception of the church, represent selected samples.[79]

A much more ambitious program of testing was planned than that carried through by Jones and Conrad, for the purpose of the Maturity Study was to determine adult performances not only on intelligence tests, but also in perception, motility, memory, the understanding of spoken directions, and other fields. In the testing program of 1930, Miles planned as comprehensive a test battery as could be administered in a two-hour period. He found that under proper conditions all subjects could be induced to give

[76] *Op. cit.*, 582. [77] *Op cit*, 269 [78] Miles, *op. cit*, 107

[79] Evidence that the percentages of adults belonging to clubs or other organizations differ markedly according to social and financial status has recently appeared in the report of the White House Conference on Child Health and Protection, Section III A. In the questionnaire study of personality development in about 13,000 public school children of the eighth, ninth, and tenth grades, it was found that 53 per cent of the poor mothers belonged to no organization—church, social, or educational. This surprisingly high percentage fell in the case of the lower-middle-class mothers to 31; in the case of the upper-middle-class mothers to 12; and in the case of the wealthy only 7 per cent belonged to no organization. (*The Adolescent in the Family*. New York: D. Appleton-Century, 1934. Table 47, page 359.) Corresponding data were apparently collected for fathers but were not given in the report.

this amount of time to the work; he noted, however, a danger of "waning interest" even within these two hours, and stated further that not all adults who serve once will serve twice and fewer still a third or fourth time.[80] The returning groups must therefore have been increasingly selected if new subjects were not added. Many new subjects were taken on in the testing program of 1932, and the investigation as a whole is based upon a number of different groups of subjects, constituting altogether the most extensive survey of adult performance ever obtained, with the exception of the Army survey. The work has so many ramifications that a complete estimate is not yet possible.

The parts which are of greatest interest in connection with the present study of adults are the results, first, of the intelligence test given to the main group of 823 subjects, aged seven to ninety-four, and, second, of certain of the particular studies, notably those of the simpler perceptual and motor abilities, the understanding of spoken language, and the adaptability to test situations.

The intelligence test employed was very brief, a form of the Otis Self-Administering Test of Intelligence, Higher Examination, Form A, given as a fifteen-minute test.[81] In justification of such a short test Miles and Miles reported that it showed a correlation of $+.86 \pm .02$ with the thirty-minute form in the limited range of superior ability represented by 121 university students, and gave other similar data. It cannot be denied, however, that a fifteen, or even a thirty-minute test gives a scanty indication of the so-called general intelligence, and such tests of course show little as to the relationship between different types of performance or their course with age.

In order to make some comparison with the Army results, the scores on the fifteen-minute Otis Test were transmuted into equivalent scores on a twenty-minute abbreviated form earlier used by Miles and Miles; these were then transmuted into equivalents for the original Otis Test, and these finally transmuted into the equivalent Army Alpha scores given by Otis. Whatever the effect of so many transmutations, the results showed clearly that the Army Alpha equivalents for the median Otis scores for Miles' group were much higher than the Army median of 59 for the native-white draft.[82] In short, the judgment as to the level of intelligence of the sample

[80] Miles, *op. cit.*, 108.

[81] Miles and Miles, *op. cit.*, 46–47.

[82] The difference is even more marked if the comparison is limited to those of Miles' subjects who were within the age range of the Army draft.

which would be indicated by the nature of the selection seems to be confirmed by these test findings.

TABLE III

SCORES BY DECADES FOR 616 ADULTS, CITY B (TABLE VI, MILES AND MILES, ABBREVIATED)

AGE PERIOD	N	MEAN SCORE ON ABBREVIATED OTIS TEST	SIGMA	$\frac{D}{\sigma_D}$	SCORE ON ARMY ALPHA
15–19	51	38.5	8.0		126
				—.11	
20–29	80	38.7	7.9		126
				2.53	
30–39	87	35.1	9.6		108
				.17	
40–49	90	34.2	9.2		108
				3.27	
50–59	119	29.9	9.7		88
				2.86	
60–69	103	26.1	10.6		72
				2.05	
70–79	68	23.1	8.7		61
				3.54	
80–89	18	14.7	9.3		26

On the question of the growth and decline of intelligence, as indicated by test scores on this fifteen-minute test, Miles and Miles stated:[83]

From a maximum score at the age of 18 yr. the curve of intelligence maintains almost a plateau through the "twenties," begins then to fall, and persists in its downward course, at first slowly, then more rapidly until the end of the life span.

An abbreviated form of one of their tables is reproduced in Table III to show the changes in test score from decade to decade, and the reliability of the differences.[84]

[83] *Op. cit.*, 60.

[84] The findings are given for City B only; the age range in City A was much shorter, from fifty to ninety years.

The mean scores for the age periods fifteen to nineteen and twenty to twenty-nine are almost identical; the decline then begins; it seems to be arrested in the forties but is marked in each successive decade thereafter. The arrest in the forties is interesting in connection with the fact that Jones and Conrad found a slight rise in the age period forty to forty-five.[85] No such rise appeared in Willoughby's results. An interpretation of the finding is difficult for in neither case does it seem to depend on fluctuations of sampling.

In summarizing the results of the Stanford investigation in his presidential address at Ithaca in 1932, Miles reported as follows on the relation of age and intelligence test performance for a larger group than that discussed above:[86]

> The age-score curve from performance on a standard intelligence test of 2500 adults, (approximately 250 over 70 years of age) weighted to represent the level of the general population, shows a top mean of 15 to 16 year mental age at life-age 18, persistence at practically this same mental level through two decades, then gradual decline to old age, registered in the Pearson correlation coefficient —.50 for age-score in the period of adulthood. Stated in another way, the test ability gain of the last three years in the period of mental age growth is gradually lost in the next three-score years.

Miles then stated that results for 400 subjects with the speed factor eliminated showed that speed did not slacken before mental power. In a later paper, however, C. C. Miles reported that comparisons of the fifteen-minute Otis speeded and unspeeded indicated that in early and middle maturity speed declined faster than power, whereas in late maturity the decline of power was relatively more apparent.[87]

Despite the brevity of the Otis Test and the comparatively small number of items, these were analyzed with a view to determining their course with age. On this question W. R. Miles reported as follows:[88]

[85] See also Mursell's similar finding, footnote 70, page 23.

[86] Miles, *op cit*, 116

[87] C. C. Miles, *Influence of Speed and Age on Intelligence Scores of Adults*, J Gen. Psychol, 1934, 10, 208–210.

Another interesting finding is reported in this paper. After noting that in all populations the average individual loses at the rate of about six or seven months mental age per decade, Miles stated that decreases were apparently similar if individuals were retested. This finding for the 190 individuals tested in 1930 and retested in 1932 is important even though the test was brief, for there have been very few retests of adults.

[88] *Op. cit.*, 116–117.

. . . verbal associations, generalizations, interpretations of meaning, and recognition of relations show marked resistance to the influence of age. Speed, organization and recall of unfamiliar material, and difficult logical procedures involving a relatively wide immediate memory span show speedier decline. Perhaps I should call the decline registered in our curve not one in intelligence as such but rather a diminution in reaction speed and sum of energy available for new work types. This would mean that the decline correlates with physiological rather than psychological deterioration.

It seems unlikely that the question of the relation of age and particular intelligence test performances can be pursued very far on the basis of the items in the brief Otis omnibus test. Willoughby's investigation and that of Jones and Conrad have more to contribute on this subject. As already noted, however, the Stanford Study included many other tests besides the "general intelligence" test, and the results of these present important data on a number of traits or types of performance, many of which do not enter into the usual form of intelligence test and have rarely been studied at the adult level.

Tests of visual perception, reaction speed, motility, and complex motor skills all gave similar age curves, characterized by highest performances in the late teens or early adult years, followed by slow decline to about sixty and more rapid decline thereafter.[89] The curves for memory and learning ability dropped more rapidly than those for the perceptual and motor activities. The grasping of spoken directions showed a decline amounting to approximately a third of the standard deviation per decade, and a correlation with age (−.51) about that of the abbreviated Otis Test.[90] A series of non-language tests constructed or selected to indicate "adaptability" to new situations, and correlating from +.30 to +.51 with intelligence, yielded correlations with age ranging from −.03 to −.56.[91] In Marsh's report on these tests, he concluded that human adaptability showed only a slight tendency to decrease with age. The point is an extremely important one, for if adaptability to new situations falls off rapidly with age, this condition may explain the declines on many tests, for example, the code test used by Miles in studying memory. Finally, certain of the performances or traits studied showed little or no change with age. Aside from the verbalizations,

[89] *Ibid.*, 112–114.

[90] B. Price, *The Grasping of Spoken Directions as an Age Function in Adults,* Psychol. Bull., 1933, 30, 588–589.

[91] C. J. Marsh, *Human Adaptability as Related to Age,* Psychol. Bull., 1933, 30, 589.

generalizations, and so forth as indicated by the Otis items, these included imagination,[92] persistence in work, and the four traits studied by the Bernreuter test—neurotic and introvert tendencies, self-sufficiency, and dominance.[93]

The demonstration that different test performances reach their maximum development at different ages and show different courses with age through maturity is one of the most interesting discoveries of the recent work. It is a problem which challenges future work as well, because the facts are a necessary basis not only for an understanding of mental functioning in the adult years, but also for the solution of practical questions as to the type of work the older person can best handle and the difficulties he is likely to meet. In this connection, however, there are two important findings which must not be overlooked. One is that even those performances which decline rapidly after maturity in comparison to other test performances, actually decline slowly in comparison to the rate of development. As general observations would of course suggest, the changes during the adult years, at least up to the age of sixty, are slight in relation to the changes during childhood. A second important finding, which Miles and other investigators have stressed, is that individual variations are large. A gradually declining curve from the twenties or thereabouts is apparently the true condition, but it must be noted that many individuals in the later decades surpass the average scores for the twenties or thirties on a long series of performances.

A satisfactory attack on the problems involved calls for the study of different types of activity, each of which must be tested separately and at sufficient length to afford a reliable estimate. Few investigations have as yet met these conditions adequately, but they are basic for much of the important work which remains to be done, including the study of the relationship of mental abilities at the adult level. There are still a number of activities which are probably significant in the adult period but have not yet been investigated; and it is evident that the satisfactory determination of tests for adults, the nature of which will differ according to the purpose of the work, awaits the study of these activities.

[92] Studied by the kinephantom, a revolving fan the silhouette of which the subject had to interpret; see W. R. Miles, *Movement Interpretations of the Silhouette of a Revolving Fan,* Amer. J. Psychol., 1931, 43, 392–405.

[93] C. C. Miles, *Age and Certain Personality Traits of Adults,* Psychol. Bull., 1933, 30, 570.

II

THE SELECTION AND COMPOSITION OF THE GROUP STUDIED

1. Selection of the Group
   a. Problems involved
   b. Criteria for selection
   c. Method of approach
2. Composition of the Group
   a. Distribution by diseases
   b. Distribution by race
   c. Distribution by age
   d. Distribution by education
   e. Distribution by occupation
3. Comparative Findings in Four Diagnostic Groups
4. Comment on the Sample

1. SELECTION OF THE GROUP

a. *Problems involved*

IN selecting normal adults to serve as a control group in the study of aphasia, the first requirement was to obtain subjects similar in age and in educational, occupational, and cultural status. Survey of the possible sources for such a group showed that the most nearly similar sample of the population, and therefore the best control, was to be found in the patients admitted to the same hospitals in which the aphasic patients were studied but not suffering from any nervous disorder or any condition known to affect mental functioning. Permission was therefore secured to study patients in the wards of the hospitals in which the majority of the aphasic patients were being examined: the Orthopedic Hospital and Infirmary for Nervous Diseases, the Graduate Hospital, and the Philadelphia General Hospital. Samples from the wards of these hospitals were satisfactory as controls because almost all the aphasic patients were also ward cases.

The advantage of this method of selection is immediately obvious. The patients admitted to one department of a hospital are more likely to be similar to those admitted to another department in all factors except the

type of disease which required their admission, than is any other limited group of the population which could be studied.[1] The disadvantage, when the purpose is the study of mental functioning in *normal* adults, is also obvious, for a number of hospital patients certainly cannot be considered normal either mentally or physically. At the same time there are many patients whose hospitalization is enforced by the need for some particular treatment rather than by a serious protracted illness or general debilitation, and who, in the opinion of the authors, would be accepted by both medical men and psychologists as normal from the point of view of mental functioning. The problem in the selection of the normal control groups therefore resolved itself into the choice of those patients whose disorders were least likely to have any effect on the functions required in the various mental and educational achievement tests.

Neurological conditions were immediately excluded from the normal control group and in addition it seemed wise to reject all cases of glandular disorder, tuberculosis, or syphilis.[2] Aside from these, few types of disease had been studied with regard to possible mental changes and there was little material available to determine the choice of patients.[3] Data from tests of intelligence in cases of children with diseased adenoids and tonsils were not of immediate value because patients coming in for treatment for these conditions had not been hospitalized long enough to be studied thoroughly. Many of the subjects finally accepted had had tonsillectomies at some previous time; very few showed diseased tonsils or adenoids at the time of the examination.

In view of the fact that orthopedic cases were available for study and were particularly suitable because they were in good working condition

[1] For a discussion of the possible slight differences in economic factors, see Weisenburg and McBride, *Aphasia A Clinical and Psychological Study*, 123.

[2] It is known that certain glandular conditions are accompanied by definite mental changes, and various investigations, particularly in the German literature, have indicated that there are characteristic mental disturbances and defects in speech and writing in tuberculosis. For a recent paper, see J. Kollarits, *Uber Sprach- und Schreibstorungen im allgemeinen und als "kleine Zeichen der Geistesschwächung" bei Tuberkulose und im Alter*, Arch. f. Psychiat. u. Nervenkh., 1933, 99, 109–196.

Syphilis is apparently characterized by slightly inferior intelligence test performance, but this condition, as explained in the Army Memoirs, may as well indicate that those of lower intelligence are more likely to expose themselves to infection or to neglect prophylactic measures as that syphilis results in lower test performance. (Mem. Nat. Acad. Sci., 1921, 15, 811.) In any case, the matter is controversial, and syphilitics could not safely be accepted for the so-called normal group.

[3] For a recent discussion of the problem, see G. C. Schwesinger's excellent summary in *Heredity and Environment*, 307–325.

and had unlimited time, the study of crippled children made by Fernald and Arlitt was of special interest.[4] These investigators found an average Stanford Binet I.Q. of 82.35 for a group of 194 children, a figure which would seem to indicate somewhat inferior intelligence. Analysis showed that this mean I.Q. was lowered by the inclusion within the group of 27 cases of spastic birth paralysis (mean I.Q. 69.11) and 15 cases of central nervous system involvement, including congenital lues (mean I.Q. 75.93). The groups of poliomyelitis, tuberculosis, nutritional disorders including rickets, infections not involving the central nervous system, and trauma ranged from 83.79 to 86.53 in I.Q. These groups were still slightly inferior, but apparently not because of the disease conditions, for similar ranges in I.Q. were characteristic of the siblings of the crippled children. In short, it seemed probable that the crippled children, who were all institutional cases, were drawn from inferior family groups in which I.Q.'s of about 85 would be typical.

Another study of interest in view of the large number of gastro-intestinal cases available for examination was Paulsen's report on mental and motor efficiency in cases of intestinal toxemia.[5] The group under treatment for intestinal toxemia gained 16 per cent more than the control group on reexamination by a battery of mental tests and tests of steadiness, motor control and coordination. The author concluded that this net gain represented the increase in mental and motor efficiency resulting from the treatment. This conclusion might well be questioned, however, in view of certain unsatisfactory conditions in the experiment. The test group was selected on the basis of the presence of harmful bacteria in sufficient numbers in the feces, and the majority when questioned "confessed" the symptoms of intestinal toxemia. The control group was matched as nearly as possible in chronological age, test intelligence, training, and environment; but no attempt was made to determine the presence or absence of intestinal toxemia in these cases. A further, though probably less important, defect in the experimental conditions is a difference of motive in the two groups: the test cases were promised assistance in bettering their intestinal conditions in return for work on the tests while the controls "entered the experiment as a matter of interest and desire to be of assistance to the examiner."

[4] M. R. Fernald and A. H. Arlitt, *A Psychological Study of a Group of Crippled Children of Various Types*, Sch. & Soc., 1925, 21, 449–452.

[5] A. E. Paulsen, *The Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency*, Arch. Psychol., 1924, 11, No. 69.

As these reports indicate, there was little evidence at the beginning of this research as to the effect on intelligence of disease, other than glandular disease or disorders of the central nervous system. During the course of the work, however, there appeared a very important study by Dawson and Conn, which is more extensive and more decisive than any previous investigation.[6] An intelligence test, the Burt Revision of the Binet, was given to over a thousand children suffering from types of disease which included pneumonia, rheumatism, nephritis, glandular disorders, epilepsy, encephalitis lethargica, and other brain conditions. The findings show clearly that non-brain cases made scores on intelligence tests equivalent to those made by the healthy population to which they belonged, while the brain cases were on the average inferior. In other words, disease, apart from disease of the brain or ductless glands, "does not appear to have any appreciable effect on the intelligence." An interesting analysis of mental test performance in acute, non-acute, and chronic conditions revealed the fact that the average scores were actually higher for patients examined when fevered and acutely ill; they were slightly lower for patients classified as chronic than for those in the afebrile but non-chronic group, but the difference was not significant.

The subjects selected for this research were studied during periods of less serious illness than those whom Dawson and Conn examined. With so little positive data on the problem, however, it was considered safest to restrict the cases to patients with bone fractures or dislocations and to patients recuperating from some surgical treatment and in good working condition at the time of the examination.

b. *Criteria for selection*

For the reasons discussed above, the normal adults were chosen from the orthopedic and surgical wards, with further limitations which would make them satisfactory as controls for the aphasic group. These limitations were as follows:[7]

1. Freedom from any present or earlier neurological, mental, or glandular disease
2. Satisfactory vision and hearing
3. Age under sixty years
4. English as the native tongue.

[6] S. Dawson and J. C. M. Conn, *Intelligence and Disease,* Med. Research Council, Spec. Report Series, No. 162, 1931.

[7] Work was begun at the Orthopedic Hospital where a number of patients were examined

c. *Method of approach*

When cases which seemed to fill the requirements had been selected from the hospital records, or the physician's report, the examiner proceeded very informally to explain the problem and ask for the patient's cooperation. He was told that studies were being made of patients who had speech difficulties resulting from a "stroke" or injury to the brain, and that these studies were handicapped by lack of knowledge as to just how much the average man remembered of his schooling, or how much he had learned without schooling. It was further explained that we were as yet unable to do a great deal for the aphasic patient because we did not know how the ordinary man worked on tests of reading, writing, and arithmetic. Two points were always made as clear as possible: 1) that the "ordinary man" was the one we wanted to study, not the well-educated man; and 2) that while each person must do his best, it was not his individual score, but the general results for the group which interested us. Sometimes one of these points was more useful in clinching the bargain, sometimes the other; and occasionally our objectives were attained when our explanations did not seem to be comprehended at all.

All investigators report a certain number of refusals; Miles and Miles speak of the "difficulty which we and others have encountered with adult subjects in administering intelligence tests." In this work, of the total number of patients asked to take the examinations, only 21 per cent refused.[8]

The figures for the three hospitals are: Philadelphia General, 24 per cent; Graduate, 19 per cent; and Orthopedic, 18 per cent.

The motives which led to acceptance in the remaining 79 per cent of the cases are of some interest. Three types were distinct, but all of these were probably present to some degree in most cases. The first was a willingness

who did not meet the requirements for selection which soon had to be laid down. These patients were eliminated from the group, but when the final calculations had been made and the complete hospital records were again checked, it was found that the group still included 9 patients who did not meet the rigid requirements: 2 of these patients were being treated for neuritis at the time of the examination, 3 were cases of deformity resulting from poliomyelitis, and 4 were cases of tuberculosis of the bone or joint. The mean score on the Stanford Binet Scale for these 9 patients was 162 and the mean score on the Pintner Non-Language Mental Test 346. Since the means for the total group on these two tests were 164 and 356 respectively, and since the group of 9 fell close to the average of the total group on the other tests as well, it seemed clear that the inclusion or exclusion of their records would not affect the results materially.

[8] The patients gave various reasons for refusing. Some thought they were too sick; some had families who would not approve, and some simply "did not want to be bothered."

to do anything which would help someone else. The second was boredom; the patient welcomed any occupation to fill in time. The third was more personal, arising from a reluctance to say "no" to the examiner and pride at being selected. At the Philadelphia General and Orthopedic Hospitals, where work was carried on over long periods of time, there was some good-natured banter about "taking lessons," but this publicity was an advantage rather than a disadvantage; patients felt that there was a certain amount of prestige attached to taking the examination, and a number who had not been selected offered themselves as subjects.

It will have been evident that in approaching the patient, the emphasis was placed rather upon achievement in reading, writing, and arithmetic than upon intelligence. During the course of the examination, a few of the patients asked if some of the tests were intelligence tests, but were not disturbed upon being told that they were.

Whether the tests were designed for group or individual application they were all given as individual tests, a very important point since individual work enabled the examiner to become well acquainted with each patient, to observe his methods of work, and to ensure his steady cooperation. Occasionally a patient was reluctant to try some particular test; if he objected seriously the test was omitted, but in most cases he could be persuaded to attempt the work, and once persuaded, he seemed to take pride in doing his best. Cooperation was splendid, and the contact between patient and examiner all that could be desired. The advantages which accrue from a favorable personal relation cannot be overemphasized.

2. COMPOSITION OF THE GROUP[9]

a. *Distribution by diseases*

The distribution of cases according to the pathological condition necessitating admission to the hospital is presented in the following outline:[10]

DISEASES OF THE SKIN	3
Due to infection: carbunculus	1
Due to trauma or physical agents: third degree burns of legs	1
ulcer of leg	1

[9] All cases at the Orthopedic Hospital and two at the Graduate Hospital were examined by Mrs. Gardiner, the remaining cases by Dr. Roe.

[10] The terminology and form of the classification are those given in a *Standard Classified Nomenclature of Disease*, H. B. Logie, ed., 1933.

DISEASES OF THE MUSCULO-SKELETAL SYSTEM 37

Of the bones

Due to infection: chronic osteomyelitis of right or left femur . . 2
chronic periostitis of right femur 1
tuberculosis of vertebra 1
Due to trauma: fracture of vertebra 1
fracture of right or left femur 5
fracture of right or left tibia and fibia 2
fracture of coccyx 1
Due to unknown causes: osteitis deformans of left femur . . . 1

Of the joints

Due to prenatal influences: congenital dislocation of hips . . . 1
Due to infection: ankylosis of right hip, tuberculous 2
ankylosis of left knee, tuberculous 1
arthritis of hip, non-tuberculous 3
arthritis of knee 1
chronic arthritis 2
arthritis of sacro-iliac joint 2
Due to trauma: sacro-iliac subluxation 2
Due to unknown causes: osteoarthritis 2
osteoarthritis of knees 4

Of the muscles

Due to disturbances of innervation: paralysis of muscle, post-poliomyelitic 3

DISEASES OF THE CARDIO-VASCULAR SYSTEM 1

Of the veins

Due to trauma: varicose veins of the leg 1

DISEASES OF THE DIGESTIVE SYSTEM 21

Of the small intestine

New growths: adenocarcinoma of intestine[11] 1
Due to unknown causes: duodenal ulcer 2

Of the colon

Due to disturbances of innervation: spasticity of colon 1

Of the appendix

Due to infection: acute appendicitis 3

[11] At the time of examination, a year before this diagnosis was made, the diagnosis was "chronic intestinal obstruction."

Of the anus

Due to static mechanical abnormality: hemorrhoids 1

Of the bile passages

Due to infection: acute cholecystitis 1
Due to static mechanical abnormality: cholelithiasis 4

Of the abdomen and peritoneum

Due to static mechanical abnormality: femoral hernia 1
inguinal hernia 7

DISEASES OF THE NERVOUS SYSTEM 2

Due to unknown causes or causes not determinable:
trigeminal neuralgia 1
left median and ulnar neuritis 1

UNCLASSIFIABLE, OBSERVATION 6

b. *Distribution by race*

The group reported in this monograph is composed entirely of white subjects.[12] All of them had been born in an English-speaking country and 65 of the 70 were natives of the United States. Of the remaining cases four, 2 men and 2 women, had been born in Ireland and one, a woman, in Scotland. Of the 65 subjects, only 2 were Jewish and only 2 came from homes where a language other than English was spoken regularly, the language in each case being Italian.

[12] A Negro group studied by the same methods numbered only 15, and it has not been considered worth while to report the results for such a small group in detail. In view of the fact that there are few studies of adult Negroes, however, several important findings are worth summarizing. It must be noted that the hospital seems to afford a good solution to the problem of securing comparable groups of whites and Negroes in the adult period, for the patients admitted to the wards of most hospitals, whether they are white or Negro, are likely to be derived from closely similar levels of the population. The probability that the groups will be comparable is particularly high when the studies are made in a city hospital, like the Philadelphia General, or in a hospital like the Graduate which serves a district containing a large proportion of Negroes of various economic strata.

The mean age of the Negro subjects, 11 male and 4 female, was 40 years. The mean school grade completed was 6, and the mean occupational rating on the Barr Scale 7.13. These two figures may of course reflect inequality of opportunity between the Negro and the white, but in any case the Negroes were inferior on all the chief tests except the Porteus Maze, and significantly inferior on some, notably the Pintner Non-Language Test and the Stanford Binet Vocabulary. The superiority of the Negroes on the Porteus Maze was not statistically significant, but it is an interesting difference, which has not, to the knowledge of the authors, been reported by other investigators.

The Negro subjects cooperated even more freely than the white subjects. They were immediately and obviously interested in the tests and they were apparently less self-conscious than the white subjects.

TABLE IV

DISTRIBUTIONS BY AGE FOR THE THREE HOSPITAL GROUPS AND FOR THE TOTAL TEST GROUP

AGE	PHILADELPHIA GENERAL HOSPITAL		GRADUATE HOSPITAL		ORTHOPEDIC HOSPITAL		ALL THREE HOSPITALS		
	M	*F*	*M*	*F*	*M*	*F*	*Male*	*Female*	*Total*
50–59	3	4	1	1	1	3	5	8	13
40–49	3	1	3	1	5	1	11	3	14
30–39	8		2	1	7	3	17	4	21
20–29	2		3	2	3	5	8	7	15
10–19	1	1	1		4		6	1	7
Number	17	6	10	5	20	12	47	23	70
Mean	38.9	47 2	34.3	36.2	32.5	34.1	35.2±1.8	38.0±2.8	36.1±1.5
Sigma	11.1	13 4	13.0	11 2	11.7	12.8	12.2±1.4	13.8±2.0	12.8±1.1

TABLE V

DISTRIBUTIONS BY SCHOOL GRADE REACHED FOR THE THREE HOSPITAL GROUPS AND FOR THE TOTAL TEST GROUP

SCHOOL GRADE REACHED	PHILADELPHIA GENERAL HOSPITAL		GRADUATE HOSPITAL		ORTHOPEDIC HOSPITAL		ALL THREE HOSPITALS		
	M	*F*	*M*	*F*	*M*	*F*	*Male*	*Female*	*Total*
15–16						1		1	1
13–14		1			1		1	1	2
11–12			1		7	2	8	2	10
9–10			2		2	2	4	2	6
7–8	5	4	5	1	6	5	16	10	26
5–6	6		2	2	3	2	11	4	15
3–4	2				1		3		3
1–2									
0		1						1	1
Unknown	4			2			4	2	6
Number	17	6	10	5	20	12	47	23	70
Mean	6.9	7.3	7 8	6 0	9.0	9.1	8.0±.4	8.1±.6	8.1±.3
Sigma	2.2	4 1	1.8	1.4	2 9	2 9	2.7±.3	3.3± 5	3.0±.2

c. *Distribution by age*

Table IV shows the age distributions by decades for the male and female sub-groups in each hospital and for the total group. The mean age for the patients at the Philadelphia General Hospital, 41.1 years, was higher than that for the total group, while the mean ages for the patients at the Graduate and Orthopedic Hospitals, 34.9 and 33.1 years respectively, were lower. The differences, however, were not great.

d. *Distribution by education*

The data on educational level are expressed in terms of the grade the subject had reached before leaving school.[13]

Distributions for the separate hospital groups and for the total group are presented in Table V.[14] There were some differences between the mean levels for the groups in the three hospitals: the Philadelphia General fell lowest with a mean school grade of 7.0, the Graduate next with a mean of 7.4, and the Orthopedic considerably higher with a mean of 9.0.

e. *Distribution by occupation*

Two occupational scales were employed, the Taussig and the Barr.[15] The Taussig is a discrete five-division scale based on general type of occupation: I Professional, II Semi-professional and business, III Skilled labor, IV Semi-skilled labor, and V Unskilled labor. The Barr Scale was constructed on the basis of a list of a hundred representative occupations, each of which was rated by thirty judges according to the grade of intelligence believed to be necessary for that particular occupation.[16] Thus it gives an index of

[13] Adjustment was made for the older patients who went to school at a time when twelve terms were roughly equivalent to eight grades today, and for the English, Irish, or Scotch-born, whose school terms differed in length. When, as in many cases, the patient stated he had finished "two or three" or "four or five" grades, the lower number was always recorded.

It would have been interesting to record also the number of grades repeated, but this additional fact was not obtained in all cases and indeed it is likely to have been forgotten in many.

[14] In 3 of the 4 cases at the Philadelphia General Hospital in which no statement as to education could be elicited, the level was probably four years or below. The fourth case, however, was one of the most intelligent men of the group, and it is probable that he had had at least an eighth-grade education.

The 2 women at the Graduate Hospital whose records for education are incomplete probably had had the equivalent of five years of schooling.

[15] Terman, *Genetic Studies of Genius,* I, Stanford University: Stanford University Press, 1925, 63–64; 66–69.

[16] Ratings for occupations other than those listed on the scale may be derived by matching the occupation to be rated with that which seems most similar to it, or by calculating an intermediate value between two occupations it resembles.

TABLE VI

DISTRIBUTIONS BY TAUSSIG OCCUPATIONAL CLASSIFICATION FOR THE THREE HOSPITAL GROUPS AND FOR THE TOTAL TEST GROUP

TAUSSIG CLASSIFICATION	PHILADELPHIA GENERAL HOSPITAL		GRADUATE HOSPITAL		ORTHOPEDIC HOSPITAL		ALL THREE HOSPITALS		
	M	*F*	*M*	*F*	*M*	*F*	*Male*	*Female*	*Total*
I						1		1	1
II	4		2		5	4	11	4	15
III	6		5		2	2	13	2	15
IV	6		3	1	4	4	13	5	18
V	1				3		4		4

TABLE VII

DISTRIBUTIONS BY BARR OCCUPATIONAL RATINGS FOR THE THREE HOSPITAL GROUPS AND FOR THE TOTAL TEST GROUP

BARR RATING	PHILADELPHIA GENERAL HOSPITAL		GRADUATE HOSPITAL		ORTHOPEDIC HOSPITAL		ALL THREE HOSPITALS		
	M	*F*	*M*	*F*	*M*	*F*	*Male*	*Female*	*Total*
13.00–13.99						1		1	1
12.00–12.99					1	1	1	1	2
11.00–11.99	1		1		1	2	3	2	5
10.00–10.99	2		2		4	1	8	1	9
9.00– 9.99	3		2		1	1	6	1	7
8.00– 8.99		5	1	4		1	1	10	11
7.00– 7.99	5		2		2	1	9	1	10
6.00– 6.99	1		1	1	2	3	4	4	8
5.00– 5.99	3		1		2		6		6
4.00– 4.99	1						1		1
3.00– 3.99	1				1		2		2
Number	17	5	10	5	14	11	41	21	62
Mean	7.6	8.4	8.6	7 9	8.2	9.4	8.0±.4	8.8±.5	8.3±.3
Sigma	2.1	0.0	1.8	0.9	2.6	2.4	2.3±.3	2.1±.3	2.3±.2

the intellectual level characteristic of the occupation, while the Taussig Scale simply classes together occupations similar in type of work involved.

Distributions on the Taussig Scale are given in Table VI.[17] Inspection of

[17] At the Philadelphia General Hospital 6 women, 5 housewives and one student, could not

the table will show that the Orthopedic Hospital group contained the largest proportion of cases from the professional and semi-professional classes. The Philadelphia General and Graduate groups contained relatively larger proportions of cases in classes III and IV, skilled and semi-skilled labor. As a whole the distribution on the Taussig Scale shows that the subjects are fairly equally divided among the semi-professional and business, the skilled labor, and the semi-skilled labor classes.

Ratings on the Barr Scale are presented in Table VII.[18] The three occupational distributions, like the educational, overlap considerably, with the mean for the Graduate Hospital (8.36) higher than that for the Philadelphia General (7.73), and the mean for the Orthopedic (8.75) higher than that for the Graduate. This overlapping is just what would be expected from a knowledge of the populations served by the three hospitals.[19]

3. COMPARATIVE FINDINGS IN FOUR DIAGNOSTIC GROUPS

The majority of the cases, 45 of the 70, fall into three main groups according to their medical diagnosis: injuries of the bone and dislocations, infections of the bones or joints, and diseases of the digestive system. The last group may be divided into two if the cases of hernia or hemorrhoids are separated from those of duodenal ulcer, intestinal obstruction, appendicitis, cholecystitis, and cholelithiasis. Before the group of 70 cases is considered as a whole, a preliminary comparison of the status of these four diagnostic groups is worth while.

Table VIII shows the mean levels for the four groups in age, education, occupational rating on the Barr Scale, and scores on the Stanford Binet and Pintner Non-Language Tests, together with the sigma deviations from the

be classified on the Taussig Scale. At the Graduate Hospital 4 women could not be classified; all were housewives. At the Orthopedic Hospital one man and one woman, each crippled and unable to work, and 5 young men, all students, could not be classified.

[18] All cases were rated with the exception of one subject at the Philadelphia General and 7 subjects at the Orthopedic Hospital. These were either students or unemployed cripples.

After some consideration housewives were assigned a rating of 8.37. This was obtained by averaging the Barr ratings for the cook and the tailor, a process which seemed to be indicated by the nature of the occupation. In some cases the rating is too low and in others obviously too high. It would probably have been fairer to give each housewife the rating which would have been assigned her husband, but this solution did not suggest itself until the opportunity for obtaining the necessary information had passed.

[19] Very few of the ward patients at the Philadelphia General Hospital pay any fees, and so far as is known, none of those included has. At the Graduate Hospital, 45 per cent of the ward patients paid an average fee of $1.00 per day. At the Orthopedic Hospital, 52 per cent of the ward patients paid an average of $1.27 per day.

means for the total group in education, occupational rating, and the two mental tests.[20] It will be evident that the groups of patients with arthritis and bone fractures and dislocations make higher mean scores on the Stanford Binet and Pintner Non-Language Tests than either of the groups of

TABLE VIII

COMPARATIVE FINDINGS FOR FORTY-FIVE CASES DISTRIBUTED BY MEDICAL DIAGNOSIS

	FRACTURES AND DISLOCATIONS	ARTHRITIS	GASTRO-INTESTINAL OTHER THAN HERNIA AND HEMORRHOIDS	HERNIA AND HEMORRHOIDS
Number of cases	11	13	12	9
MEAN LEVELS				
Age	31.2	36.1	37.2	39.6
School grade reached	7.9	10.2	7.1	7.4
Occupation: Barr Scale	7.9	9.6	8.5	8.6
Stanford Binet Scale	168.1	178.5	152.5	155.0
Pintner Non-Language Test	353.3	377.0	347.1	353.0
SIGMA DEVIATIONS FROM THE MEANS FOR THE TOTAL GROUP				
School grade reached	−.1	+.7	−.3	−.2
Occupation: Barr Scale	−.2	+.6	+.1	+.1
Stanford Binet Scale	+.2	+.6	−.4	−.3
Pintner Non-Language Test	0.0	+.2	− 1	0.0
PERCENTAGE DISTRIBUTION OF THE GROUP BY HOSPITAL				
Philadelphia General	36	8	42	78
Graduate	19	8	50	22
Orthopedic	45	84	8	0

patients with gastro-intestinal disorders. The first two groups surpass the mean for the total number of patients on the Stanford Binet Scale while

[20] The Stanford Binet and the Pintner Non-Language represent only two of the large number of tests used in the study, but are sufficient for this brief comparative survey of the mental levels of four diagnostic groups.

the second two groups fall below it. The differences are similar to the differences in education, and the latter depend on the differing proportions from the three hospitals in the various diagnostic groups. Since the Orthopedic Hospital is superior to the other two in the educational level of the patients studied, the arthritic group, with 11 of 13 patients from the Orthopedic, reaches a higher mean level in education than the others. Mean scores on the Pintner Non-Language Mental Test, which are probably less affected than the Binet scores by educational and cultural status, are similar in all four groups.

The slight differences between the mental levels in the four diagnostic groups apparently reflect only differences in selection; it seems clear that they do not depend on the positive or negative effects of any one or more disease conditions. This is an important point, for there may be less controversy as to the normality of mental functioning in one diagnostic group than in another. In cases of bone injury or of dislocation, when the patient has been in good health up to the time of the accident which caused his injury, when he is without pain and the treatment is proceeding well, there can be no doubt that he reacts normally to the tests given him. In the gastro-intestinal cases, in some instances after an abdominal operation, there may be some question as to whether the patient reacts normally. It is important therefore to note that the patients of this character reached the levels expected of them on the basis of their educational status just as those of any other group did. In view of the fact that each of the four groups is small, the figures are indeed extremely consistent. It goes without saying that in the case of all patients every precaution was taken to ensure that the findings would be as trustworthy as possible, and examinations were made only when the patient was in good working condition and when he felt ready for work.

4. COMMENT ON THE SAMPLE

Two criteria were available for placing this small sample of adult subjects in relation to larger populations: first, the occupational distribution of the 1930 census for the city of Philadelphia and, second, the educational and occupational data for the very large groups of adult men examined during the War. The first criterion, of course, provides only an index of the similarity between the occupational distributions for the test group and the city population; it does not indicate, as the second does, anything as to the educational or intellectual level of the test group.

The 1930 census of the United States, which was made during the first of the three years in which these subjects were examined, shows occupational distributions for the city of Philadelphia which are expressed in Table IX in terms of the percentages of the native-white male, native-white female, and total native-white population over ten years of age and gainfully em-

TABLE IX

PERCENTAGE DISTRIBUTION BY OCCUPATION OF THE NATIVE-WHITE POPULATION OF PHILADELPHIA AND OF THE EMPLOYED MEMBERS OF THE TEST GROUP

	MALE		FEMALE		TOTAL	
	Philadelphia population	*Test group*	*Philadelphia population*	*Test group*	*Philadelphia population*	*Test group*
Agriculture	–	0	–	0	–	0
Forestry and fishing	–	0	–	0	–	0
Extraction of minerals	–	2	–	0	–	2
Manufacturing and mechanical industries	42	39	29	18	37	35
Transportation and communication	11	25	3	9	10	21
Trade	20	22	11	9	15	19
Public service (not elsewhere classified)	5	0	–	0	3	0
Professional service	6	2	13	18	7	5
Domestic and personal service	3	8	12	28	12	12
Clerical occupations	13	2	31	18	15	6

NOTE: Dash (–) = less than one per cent.

ployed in the ten occupational divisions.[21] The corresponding percentages in each of the occupational divisions for the test group are placed beside those for the city population. Inspection of the table shows a marked simi-

21 *Fifteenth Census of the United States, 1930, IV, Occupations by States,* Washington, U.S. Gov't Printing Office, 1933, Table 12, 1412–1415.

The percentages falling in each of the ten occupational divisions were calculated separately for the native-white males, the native-white females, and the total native-white population. Data for the foreign born were discarded because there were only five foreign born in the test group, and these were all natives of English-speaking countries and probably better classed with the native-white than with the foreign-born population.

larity between the occupational distributions for the test group and for the city population. For both the total test group and the white males the percentages in the divisions of manufacturing and mechanical industries and of trade are very nearly the same as the percentages for the city population in these two divisions. The only discrepancies which are at all noteworthy are in the division of transportation and communication where the percentages for the test group exceed those for the city population and in the clerical occupations where the opposite relationship appears. On the whole it is clear that the test group, small as it is, represents a good sample of the city population from the point of view of the occupational divisions listed in the census report. Whether the test group samples the upper, middle, or lower strata of these occupational divisions cannot, of course, be determined from these figures, and for evidence on this point one must turn to the analyses which have been made of the occupations and Army Alpha scores of the recruits examined during the War.

A convenient method of estimating a mean Barr Scale rating for the occupations of the men tested by the Army Alpha during the War is possible on the basis of Fryer's and Symonds' work with the Army data. Fryer reported the corrected Army Alpha scores for 96 occupations.[22] From these Symonds selected the occupations representative of a score of about 48 on the Army Alpha, that is, a score which appeared to be typical of the average man.[23] These occupations, except for a few which were not clearly designated, have each been assigned a Barr rating. They are as follows:

Masons	8.58
Hospital attendants	6.26 (estimated)
Station agents	10.26
Miners	4.29
Teamsters	5.41
Riggers	8.00 (estimated)
Boilermakers	9.06 (estimated)
Airplane workers	8.65 (estimated)
Factory storekeepers	10.61 (estimated)
Horse shoers	5.97 (estimated)
Salesclerks	9.72
Hostlers (R.R.)	6.91 (estimated)
Barbers	6.92
Stationary engineers	9.51 (estimated)
Cobblers	6.85

[22] Fryer, *op. cit.*, 275–276.

[23] Symonds, *op. cit.*, 65–78.

Horse trainers	7.24 (estimated)
Caterers	9.97 (estimated)
Bricklayers	7.77
Auto truck chauffeurs	7.23 (estimated)
Farmers	7.24
Bakers	7.91
Mean	8.12

The range of these occupations on the Barr Scale is surprisingly wide, 4.29 to 10.61. The mean rating is 8.12. This figure is very close to the mean Barr rating of 8.01 which was found for the men of the test group.[24] The fact that the men of the test group fall so near the mean rating for the occu-

TABLE X

PERCENTAGE DISTRIBUTION BY SCHOOL GRADE REACHED OF MALES IN THE TEST GROUP AND OF THE NATIVE-WHITE DRAFT OF THE AMERICAN ARMY

SCHOOL GRADE REACHED	TEST GROUP	ARMY
15–16	0.0	2.0
13–14	2.3	3.3
11–12	18.5	6.8
9–10	9.3	11.1
7–8	37.3	39.8
5–6	25.6	20.1
3–4	7.0	11.2
1–2	0.0	3.8
0	0.0	1.9
Number of subjects	43	60,250
Mean grade reached	8.0	6.9

pations which have been thought to characterize the man of average intelligence as indicated by Army Alpha score, together with the similarity in occupational distribution shown between the test group and the city popu-

[24] The massing of most of the women as housewives at a Barr rating of 8.37 is the chief factor in the rise of the mean for the total group over that for the men only. As noted before, this is an unsatisfactory rating for the housewives. It is probable that the mean for the men is more truly representative of the level of the group.

lation, is evidence that the test group is a fairly good sample with regard to level of occupation, that is, grade of intelligence required, as well as type of occupation.

The education of the adults in the test group may be compared directly with the education of the native-white draft of the Army.[25] The percentages of each group reaching the various school grades, which are presented in Table X, are very similar.[26] The mean for the test group is somewhat higher than that for the Army, but the chief reason for this difference may well be the fact that the test group was studied fifteen years after the Army group.

The various comparisons which have been made show that the test group is a good sample of the occupational distribution of the city of Philadelphia, that it has a mean Barr rating which is very close to that of the occupations considered to be representative of the man of average intelligence as indicated by Army Alpha score, and finally that, although its mean school grade is somewhat higher than the Army mean, it is nevertheless very similar to the large Army sample from the point of view of educational distribution. The findings are sufficient to indicate that it is a good sample of the middle levels of the population, and as such a valuable basis for the study of mental functioning in the adult period.

[25] The test group, as noted above, contained five foreign-born subjects, but these were all from English-speaking countries and nearer the general level of the native-white than of the foreign-born population.

[26] The statistics for the Army are derived from Table 302 of the Mem. Nat. Acad. Sci., 1921, 15, 758.

III

THE TESTS

1. Basis for Selection
2. Outline of Tests
3. Examination Periods

1. BASIS FOR SELECTION

THE tests were chosen to meet the problem of the aphasia study, that is, to establish and analyze for the aphasic patients whatever changes might occur in language and thought processes.[1] The complexity of the changes in aphasia, as they had repeatedly been demonstrated in the literature, made it necessary to study a wide variety of performances, ranging from the simplest speech functions to complicated processes of abstract thought. Only the simplest speech processes, such as repeating easy words and phrases or naming common objects, could be assumed to be entirely within the ability of all adults except the low-grade feebleminded and the mentally deteriorated. For all the more complicated performances, including responses on both intelligence tests and educational achievement tests, tentative adult norms had to be established before the aphasic performances could be evaluated. Consequently all the tests of the aphasic battery, with the exception of the simplest speech tests, were given to the normal group.

[1] The fact that the tests were selected primarily for the study of aphasia resulted in the inclusion of tests which would not have been chosen for a study of adult intelligence alone. (The battery used with the non-aphasic adults included the Stanford Binet Scale, only a few separate tests of which were employed in the study of aphasic patients.) There were included, first of all, tests for specific processes, such as the particular reading abilities analyzed by the Gates Tests of Visual Perception and Selection. Performances on these tests are interesting in cases of aphasia where reading difficulties are pronounced but comparatively uninteresting in normal adults, and results of such tests will not be presented. Secondly, since the understanding of spoken language and of printed material is sometimes unequally disturbed in cases of aphasia, both oral and printed forms of several tests were arranged. Both forms were given to the normal control group, and the results are reported in this study. Thirdly, in view of the aphasic limitations in speaking and writing, it was necessary to include tests which would not be greatly influenced by particular handicaps in either of these methods of response. There are two cases in point here: first, the use of both the Morrison McCall Spelling Scale, a word dictation test, and the Gates Oral Spelling Test; and, second, the use of the Chapman Unspeeded Reading-Comprehension Test in addition to the Thorndike McCall Reading Scale, because reading comprehension on the former can be indicated without a verbal response, simply by pointing to the incorrect word.

While some few intelligence tests for adults exist, notably the Army Alpha and the various examinations for high school or college students, none was sufficiently comprehensive or sufficiently analytic for the aphasia study. In any case, all the tests for college students could immediately be eliminated as too difficult for the average adult. It was therefore necessary to construct the battery largely from the mental and educational achievement tests in use with children. Among these tests, an effort was made to choose those with content best adapted to adult interests. In their use with children most of these tests have been shown to be reasonably valid and reliable. For others no data have been presented. It must be noted, however, that the tests cannot be assumed to be valid and reliable for adults on the basis of results for children.

2. OUTLINE OF TESTS[2]

In the following outline of tests and in the presentation of results in Chapter IV, the tests have been divided into three groups: Language Intelligence, Educational Achievement, and Non-Language. This division is a more or less artificial one, and there are many points of overlapping between the groups, but it is a convenient practical arrangement.

I. Language Intelligence Tests
  a. Controlled Association Tests
    1. Opposites Tests
      Oral Form: Whipple's List IV
      Printed Form: new list constructed to approximate List IV in difficulty
    2. Mixed Analogies Tests
      Oral Form: constructed from Van Wagenen's Test B
      Printed Form: constructed from Van Wagenen's Test A
  b. Kelley Trabue Completion Exercise Beta
  c. Vocabulary Tests
    1. Oral: Stanford Binet Vocabulary Test
    2. Printed: Thorndike Test of Word Knowledge
  d. Stanford Binet Scale
  e. Miscellaneous Tests
    1. Absurdities Tests: Oral and Printed Forms constructed from absurd statements suggested by Ballard and others
    2. Pintner-Toops Revised Directions Test

[2] For sources and for construction and content of new tests, see Appendix.

II. Educational Achievement Tests
  a. Reading Tests
    1. Gray Oral Reading Tests
    2. Thorndike McCall Reading Scale
    3. Chapman Unspeeded Reading-Comprehension Test
  b. Spelling Tests
    1. Gates Oral Spelling Test
    2. Morrison McCall Spelling Scale
    3. Stanford Achievement Dictation Test
  c. Writing a Letter
  d. Arithmetic Tests
    1. Stanford Achievement Computation Test
    2. Stanford Achievement Reasoning Test

III. Non-Language Tests
  a. Tests from the Pintner Paterson Performance Scale
    1. Mare and Foal
    2. Seguin
    3. Substitution
  b. Pintner Non-Language Mental Test
  c. Cancellation
  d. Immediate Recall: Kuhlmann Binet
  e. Drawing a Chair
  f. Goodenough Drawing Scale
  g. Porteus Maze Test

3. EXAMINATION PERIODS

Ten to fifteen hours were required to give the complete battery of tests. Since the normal working period was generally a half hour to an hour, with one and sometimes two periods a day, the examination required ten days to two weeks to complete. A number of patients were discharged before all the tests had been given, but since they were not given in a definite order all of the tallies are less than the total number of cases in the group.[3]

[3] Although no regular order was set up for all the tests, some particular rules were observed: where both oral and printed forms of a test were being used, the oral was given first; and the Pintner Non-Language Test was given before the Digit-Symbol Substitution of the Pintner Paterson Scale.

IV

TEST RESULTS

1. Language Intelligence Tests
2. Educational Achievement Tests
3. Non-Language Tests

1. LANGUAGE INTELLIGENCE TESTS

UNDER the heading of language intelligence tests are grouped the verbal intelligence tests of the generally accepted forms—opposites, sentence completion, vocabulary, and so forth—and also the Stanford Binet Scale. The latter does include non-verbal sub-tests, but is best considered in this first group because it is predominantly verbal. The findings on all the tests of the group are presented in the first section of Table XI, where the number of cases, the range of scores, the mean, sigma, and the standard errors of the mean and sigma are given for each test.[1]

Of the tests of this group the first two, the *Opposites,* were the only ones which were definitely too easy for the adults examined. The distributions of scores in terms of number of correct responses were markedly skewed.[2] The distributions of time scores were less skewed but it is evident that the tests as they stand are not satisfactory for normal adults.

The *Analogies Tests* have considerably greater discriminative value. The distribution for the Printed Analogies, which happened to be given to a larger number of cases, was better than that for the Oral, but the Oral was sometimes a more interesting test in that it gave an opportunity for close observation of the patient's reactions. The Analogies Tests, like the Opposites, were always well received by the adult subjects.

The *Kelley Trabue Completion Exercise Beta* proved a most satisfactory test from many points of view. It presented a minimum of administrative difficulties and was generally interesting to the adults of this group. The

[1] All calculations of means and sigmas were made from raw scores.

[2] As already indicated, there is some question about the true distribution of so-called general intelligence or of intelligence test performances at the adult level. Distributions of children's scores are approximately normal and there is no reason to suppose that adults would not show similar distributions for the majority of tests. In some instances, however, normal distributions may result from the operation of chance factors. In others, the nature of the performance studied may be such as to yield a negatively skewed distribution at the adult level despite the fact that the test is fairly adequate.

distributions are shown in Table XII. The total distribution on this test is one of those most nearly approximating the normal curve.

Both vocabulary tests, the *Stanford Binet* and the *Thorndike Test of Word Knowledge,* were easy to administer and fairly satisfactory in dis-

TABLE XI

RANGE OF SCORES, MEAN SCORES, AND SIGMAS ON ALL TESTS OF THE BATTERY

	N	RANGE	MEAN	SIGMA
I. LANGUAGE INTELLIGENCE TESTS				
Oral Opposites: Score	65	9–20	17.8±0.3	2.4±0.2
Time	65	141–37	71.0±2.9	23.5±2.1
Printed Opposites: Score	51	9–20	16.2±0.4	3.1±0.3
Time	51	130–30	74.2±4.9	34.5±3.4
Oral Analogies	41	6–31	20.6±1.0	6.5±0.7
Printed Analogies	60	7–34	21.0±0.9	6.7±0.6
Kelley Trabue Completion Beta	57	4.0–10.2	7.8±0.2	1.3±0.1
Stanford Binet Vocabulary	69	25–86	54.5±1.9	15.6±1.3
Thorndike Word Knowledge	54	18–98	61.8±2.8	20.4±2.0
Stanford Binet Scale	65	96–234	163.6±3.1	25.0±2.2
Oral Absurdities	44	0–10	6.1±0.4	2.4±0.3
Printed Absurdities	41	1–10	7.1±0.3	2.2±0.2
Pintner-Toops Directions	62	5–27	15.6±0.6	4.8±0.4
II. EDUCATIONAL ACHIEVEMENT TESTS				
Gray Oral Reading Paragraphs, Set IV	49	129–40	73.7±3.0	20.6±2.1
Thorndike McCall Reading	59	14–33	24.7±0.6	4.6±0.4
Chapman Reading-Comprehension	42	3–29	16.4±1.0	6.6±0.7
Gates Oral Spelling	53	9–36	26.1±1.0	6.9±0.7
Morrison McCall Spelling	59	9–49	39.8±1.1	8.3±0.8
Stanford Achievement Dictation	58	32–121	90.6±2.8	21.1±2.0
Writing a Letter	36	2.8–8.0	5.4±0.2	1.2±0.1
Stanford Arithmetic Computation	55	10–45	26.4±1.1	8.3±0.8
Stanford Arithmetic Reasoning	54	8–38	22.8±0.7	5.4±0.5

NOTE: Each mean and sigma is followed by its standard error.

TABLE XI (*cont.*)

	N	RANGE	MEAN	SIGMA
III. NON-LANGUAGE TESTS				
Pintner Paterson				
Mare and Foal: Time	50	52–8	28.6±1.3	10.5±1.1
Seguin	46	28–9	15.5±0.9	6.2±0.6
Substitution	45	325–50	104.8±6.5	44.1±4.4
Pintner Non-Language	56	80–526	356.4±14.4	107.8±10.2
Imitation	56	0–12	6.3±0.4	2.7±0.3
Easy Learning	56	8–50	37.8±1.5	10.9±1.0
Hard Learning	56	6–49	33.6±1.3	9.8±0.9
Drawing Completion	56	6–20	13.4±0.4	3.3±0.3
Reversed Drawing	56	0–9	5.5±0.3	2.0±0.2
Picture Reconstruction	56	2–28	12.8±0.7	5.4±0.4
Cancellation	46	5–34	19.6±1.0	6.8±0.7
Kuhlmann Immediate Recall	44	433–57	173.0±10.7	70.7±7.5
Drawing a Chair	27	6–26	15.7±1.0	5.2±0.8
Goodenough Drawing	53	4–50	28.2±1.5	10.6±1.0
Porteus Maze	45	48–216	169.5±5.5	34.6±3.7

NOTE: Each mean and sigma is followed by its standard error.

criminating the adult subjects. The distributions are presented in Tables XIII and XIV. The Stanford Binet, as is usually the case, furnished an excellent preliminary index of the patient's level together with various clues as to his cultural development, his intellectual reactions, and his interests and attitudes. Both tests were found to be fairly reliable. The correlation between the two lists of the Stanford Binet was +.95 with a standard error of .02 for 41 cases.[3] The correlation between odd and even items of the Thorndike Test of Word Knowledge was +.94 ±.02 for 52 cases.

The distributions of mental ages on the Stanford Binet Scale appear in Table XV.[4] Mental ages for these adult subjects are to be understood in a

[3] It is interesting to note that the mean scores on the two lists were almost the same: List 1, 27.76 and List 2, 26.34.

[4] The procedure in giving the Stanford Binet Scale was to include the tests from the X-year level through the year where all tests were failed. Tests below the X-year level were given when necessary to establish a basal age.

In 4 of the 65 cases only the starred tests were given.

The correlation between the scores on the complete scale and the scores on the abbreviated scale of starred tests is +.97 ±.01 for 61 cases. The mean score for the complete scale was 163.6, for the starred scale 165.4.

Since "scatter" of Stanford Binet scores is usually considered an important index in dealing with psychotics, it was thought interesting to compute the scatter for the normal adults of this

TABLE XII

DISTRIBUTION OF SCORES ON THE KELLEY TRABUE COMPLETION EXERCISE BETA

SCORE	MALE	FEMALE	TOTAL
10.0–10.49	2		2
9.5– 9.99		1	1
9.0– 9.49	5	2	7
8.5– 8.99	6	1	7
8.0– 8.49	6	6	12
7.5– 7.99	7	3	10
7.0– 7.49	5	1	6
6.5– 6.99		3	3
6.0– 6.49	4	1	5
5.5– 5.99		1	1
5.0– 5.49	1		1
4.5– 4.99			
4.0– 4.49	1	1	2
Number	37	20	57
Mean	7.89	7.62	7.80
Sigma	1.29	1.29	1.27

TABLE XIII

DISTRIBUTION OF SCORES ON THE STANFORD BINET VOCABULARY TEST

SCORE	MALE	FEMALE	TOTAL
84–89	2		2
78–83	2	1	3
72–77	2	2	4
66–71	3	3	6
60–65	7	1	8
54–59	9	4	13
48–53	8	2	10
42–47	7	3	10
36–41	3	3	6
30–35	2	4	6
24–29	1		1
Number	46	23	69
Mean	55.30	52.83	54.49
Sigma	13.05	15.10	15.59

limited sense, not as units of development corresponding to chronological age, but simply as scores. The highest possible Stanford Binet score is 234. Only one individual attained this maximum and only 4 others reached or exceeded 210. There may well be an end error in these 5 scores. Furthermore, there may possibly be an end error in the scores of the 15 other cases passing one or more tests at the XVIII-year level. A glance at the distribution, however, shows no piling up of scores in the higher steps.

TABLE XIV

DISTRIBUTION OF SCORES ON THE THORNDIKE TEST OF WORD KNOWLEDGE

SCORE	MALE	FEMALE	TOTAL
91–100	2	2	4
81–90	3	1	4
71–80	9	4	13
61–70	6	5	11
51–60	5		5
41–50	4	2	6
31–40	2	4	6
21–30	1	1	2
11–20	3		3
Number	35	19	54
Mean	61.54	62.21	61.78
Sigma	21.41	20.47	20.40

Despite all the justifiable criticisms which have been made of the Stanford Binet Scale, particularly for use with adults, it stands up well under analysis: inspection of the percentages passing each sub-test, as shown in Table XVI, indicates that few of the tests are badly misplaced.[5] For the

group. The measure used was Woodworth's so-called S D. (R. S. Woodworth, *Table for Finding the S.D. of the Mental Age Distribution*, New York, 1927; *Computing the Standard Deviations and Probable Error of a Binet Mental Age*, Psychol. Bull., 1928, 25, 167.) It should be recognized that as the basal age becomes higher, the possible size of the S.D. is progressively lessened and no correction for this change is possible. For the normal adults, the figures are:

	Mean S D.	*Sigma*	*N*
Men	31 74	11 11	42
Women	31.26	10.32	23

The measure of scatter plotted against Stanford Binet score gives a triangular-shaped distribution probably resulting in large part from the effect mentioned above of the raising of the basal age.

[5] Several of the tests require particular comment. The weights, IX 2, used at the Graduate

TABLE XV

DISTRIBUTION OF SCORES ON THE STANFORD BINET SCALE

SCORE	MALE	FEMALE	TOTAL
220–234	1		1
205–219	4	2	6
190–204	4	1	5
175–189	6	1	7
160–174	11	5	16
145–159	9	6	15
130–144	5	5	10
115–129	2	1	3
100–114		1	1
85–99		1	1
Number	42	23	65
Mean	167.9	155.8	163.6
Sigma	26.0	27.9	25.0

adults examined, the order of the tests from the most to the least difficult was as follows:[6]

1.	XVIII	6	Ingenuity
2.		4	Reproduces passage
3.		5	7 digits backwards
4.		1	Vocabulary (75)
5.		2	Paper cutting

and Philadelphia General Hospitals, were discovered, after much of the testing had been done, to have developed inaccuracies. Since they may have been inaccurate for some time, all the results which had previously been obtained were discarded. The rank of this test in the above outline is based on the records from 42 cases only, 34 at the Orthopedic and the balance at the Philadelphia General Hospital. Since the Orthopedic group is generally superior to the other two, the relative difficulty of this test does not depend on an overweighting of low-grade cases.

No rank was assigned to the Code Test, XVI 6. This was one of the few tests for which administrative difficulties were encountered. In each hospital many of the subjects refused to attempt it at all; others started in a half-hearted fashion but accomplished very little. The test was not only unsatisfactory as an index of intelligence for these adults, but it was so distasteful to many of them that it was eventually dropped from the series, and replaced for purposes of scoring by an alternate.

The Free Association, X 6, was a second test which presented some administrative difficulties. In attempting to produce as many words as possible in three minutes, the adults were decidedly more self-conscious than children. Some of them complained that the test made them "feel foolish." (Observations of behavior seemed to indicate that the test was sometimes an index of emotional stability.) Interestingly enough, the Negroes appeared to be more comfortable in this test situation than the whites.

[6] In the Army data the tests of an abbreviated Stanford Binet Scale are ranked in order of difficulty for white males. A similar ranking of tests for the white males of this group gives a rank correlation of +.86 with the Army list.

6.	XVI	Al.1	28 syllables
7.	XVIII	3	8 digits forwards
8.	XVI	1	Vocabulary (65)
9.		5	6 digits backwards
10.		2	Fables
11.		4	Enclosed boxes
12.		3	Difference between abstract words
13.	XIV	6	Clock
14.		Al.1	7 digits forwards
15.	XII	3	Ball and field
16.		6	5 digits backwards
17.	XIV	5	Arithmetic reasoning
18.	X	6	60 words
19.	XII	4	Dissected sentences
20.	XIV	1	Vocabulary (50)
21.	X	3	Designs
22.	XIV	3	President and king
23.		4	Problems of fact
24.	XII	5	Fables
25.		8	Similarities
26.	XIV	2	Induction
27.	X	4	Reading and report
28.	XII	7	Pictures, interpretation
29.	X	Al.2	20–22 syllables
30.	XII	1	Vocabulary (40)
31.	IX	6	Rhymes
32.	X	Al.1	6 digits forwards
33.	IX	2	Weights
34.		4	4 digits backwards
35.	XII	2	Abstract words
36.	X	2	Absurdities
37.	IX	5	3 words in sentence
38.	VIII	1	Ball and field
39.	IX	1	Date
40.	X	1	Vocabulary (30)
41.		5	Comprehension
42.	IX	3	Makes change
43.	VIII	2	Counts 20–1
44.		4	Similarities (2)
45.		5	Definitions
46.	VII	3	5 digits forwards
47.		6	Copies diamond
48.		Al.2	3 digits backwards
49.	VI	2	Mutilated pictures

TABLE XVI

PERCENTAGE OF THE TEST GROUP PASSING EACH ITEM OF THE STANFORD BINET SCALE

TEST	M	F	TOTAL
YEAR VI			
1. Right and left	100	100	100
2. Mutilated pictures	100	96	98
3. Counts 13 pennies	100	100	100
4. Comprehension	100	100	100
5. Names coins	100	100	100
6. 16–18 syllables	100	100	100
YEAR VII			
1. Fingers	100	100	100
2. Pictures, description	100	100	100
3. 5 digits forwards	100	96	98
4. Ties bow-knot	100	100	100
5. Differences	100	100	100
6. Copies diamond	100	96	98
Al.1. Days of week	100	100	100
Al.2. 3 digits backwards	100	96	98
YEAR VIII			
1. Ball and field	100	87	95
2. Counts 20–1	100	96	98
3. Comprehension	100	100	100
4. Similarities (2)	100	96	98
5. Definitions	100	96	98
6. Vocabulary (20)	100	100	100
YEAR IX			
1. Date	97	96	97
2. Weights	91	86	89
3. Makes change	97	100	98
4. 4 digits backwards	95	87	92
5. 3 words in sentence	97	91	95
6. Rhymes	85	86	86
Al.1. Months	100	100	100
YEAR X			
1. Vocabulary (30)	97	100	98
2. Absurdities	95	91	94
3. Designs	64	64	64
4. Reading and report	74	82	77
5. Comprehension	97	100	98
6. 60 words	62	61	62
YEAR X (*cont.*)			
Al.1. 6 digits forwards	90	87	89
Al.2. 20–22 syllables	82	86	84
YEAR XII			
1. Vocabulary (40)	90	74	85
2. Abstract words	97	87	94
3. Ball and field	60	48	55
4. Dissected sentences	62	65	63
5. Fables	71	74	72
6. 5 digits backwards	64	48	59
7. Pictures, interpretation	79	78	78
8. Similarities (3)	74	70	72
YEAR XIV			
1. Vocabulary (50)	69	56	64
2. Induction	85	61	77
3. President and king	76	48	66
4. Problems of fact	69	61	66
5. Arithmetic reasoning	71	44	62
6. Clock	55	30	46
Al.1. 7 digits forwards	64	39	55
YEAR XVI			
1. Vocabulary (65)	26	30	28
2 Fables	38	35	37
3. Difference between abstract words	36	44	38
4 Enclosed boxes	42	26	37
5. 6 digits backwards	31	26	29
6 Code	20	–	15
Al.1. 28 syllables	23	23	23
YEAR XVIII			
1. Vocabulary (75)	14	13	14
2. Paper cutting	20	22	20
3 8 digits forw. rd.	31	17	26
4. Reproduces passage	10	9	9
5. 7 digits backwards	14	0	9
6. Ingenuity	10	0	6

All the other tests were passed by each member of the group who fell sufficiently low to have been given them. The mean score for repeating digits forward was 6.69 ±.12 with a sigma of 1.02 ±.09, and the mean score for repeating digits backwards 4.87 ±.14 with a sigma of 1.16 ±.10. An end error was present in the distribution of the former, but not in that of the latter.

Neither of the *Absurdities Tests* is adequate in the present form. Of the two the Oral gave the better distribution. The construction of a longer absurdities test would probably be well worth while, for the detection of absurdities seems to be an excellent test problem for adults.[7]

The *Pintner-Toops Revised Directions Test* proved a convenient introduction to longer printed tests, and the total distribution of scores was satisfactory, but the diversity of the activities required by the various items made it difficult to interpret the total score.

2. EDUCATIONAL ACHIEVEMENT TESTS

The scores on the *Gray Oral Reading Test* are combined time and error scores.[8] The results reported in the second section of Table XI are based on only those subjects able to attempt reading as difficult as that of Set IV, that is, on 57 of the 68 subjects to whom the oral reading test was given, and the mean of 73.7 is consequently slightly lower than a mean for the total group would be. The distribution shows a marked negative skew.

The two reading comprehension tests, the *Thorndike McCall Reading Scale* and the *Chapman Unspeeded Reading-Comprehension Test*, were both fairly satisfactory for the adults examined. The first had a slight advantage from the point of view of the regularity of the distribution, which is presented in Table XVII.[9] The second, however, is a little better adapted to adults from the point of view of content.

The distributions for the three spelling tests all showed more or less negative skew. Those for the *Gates Oral Spelling* and the *Stanford Achievement Dictation Test* showed much less skew than that for the *Morrison McCall*.

[7] The combined scores on the Oral and Printed Absurdities Tests correlated fairly highly with the scores on the Stanford Binet Scale: +.75 with a standard error of .07 for 40 cases.

[8] Analyses of the errors made on the Gray Oral Reading Tests, Sets II, III, and IV, showed: 55 per cent mispronunciations, 22 per cent repetitions, 12 per cent omissions, and 11 per cent insertions.

[9] The Chapman Test was unfortunately given to only 42 cases, and this small number may account for some irregularities in the distribution.

TABLE XVII

DISTRIBUTION OF SCORES ON THE THORNDIKE MCCALL READING SCALE

SCORE	MALE	FEMALE	TOTAL
31–33	4	1	5
28–30	8	6	14
25–27	9	1	10
22–24	8	8	16
19–21	4	2	6
16–18	4	3	7
13–15	1		1
Number	38	21	59
Mean	24.92	24.24	24.68
Sigma	4.75	4.40	4.63

The Gates Test was quickly administered and readily accepted by adult subjects. The correlation of alternate items is .86 ±.04. For the 53 adults examined the order of difficulty of the words is as follows:

1. it
2. but
3. are
4. day
5. nine
6. card
7. me
8. do
9. mail
10. catch
11. teach
12. prison
13. built
14. travel
15. factory
16. marriage
17. afraid
18. visitor
19. measure
20. information
21. circular
22. amusement
23. estimate
24. elaborate
25. arrangement
26. approaches
27. necessary
28. restaurant
29. difficulty
30. acquaintance
31. magnificent
32. extraordinary
33. miscellaneous
34. architecture
35. conscientious
36. hippopotamus

The Morrison McCall Spelling Scale is not sufficiently difficult for adults from average or near-average levels of the population,[10] but the Stanford Dictation Test is fairly satisfactory from the point of view of its difficulty. A certain amount of negative skewness, such as appeared on both the Gates and the Stanford Tests may be characteristic of the scores for adult subjects no matter what the content of a legitimate spelling test, for a large number of the words which may justifiably be included, without making the test a

[10] The Morrison McCall Scale is satisfactory for low-grade subjects and is useful in patho-

TABLE XVIII

DISTRIBUTION OF SCORES ON THE STANFORD ACHIEVEMENT ARITHMETIC COMPUTATION TEST

SCORE	MALE	FEMALE	TOTAL
45–48		1	1
41–44	2		2
37–40	2		2
33–36	4	2	6
29–32	5	1	6
25–28	11	5	16
21–24	6	6	12
17–20	6	2	8
13–16	1		1
9–12		1	1
Number	37	18	55
Mean	26.92	25.33	26.40
Sigma	6.99	7.35	8.31

stunt performance, are within the ability of many adults. In other words, there is reason to believe that spelling is an activity which is fairly adequately acquired at adult levels by persons of moderate as well as persons of higher intelligence and education.

logical conditions where sentence dictation tests are impossible. The correlation of alternate items is $+.94 \pm .02$.

The words of List 1 fall in the following order of difficulty:

1. run
2. top
3. red
4. book
5. play
6. mine
7. with
8. sea
9. lay
10. add
11. easy
12. done
13. body
14. tired
15. alike
16. shut
17. reason
18. perfect
19. nearly
20. fifth
21. friend
22. led
23. folks
24. anyway
25. desire
26. therefore
27. written
28. omit
29. arrange
30. popular
31. celebration
32. search
33. interest
34. height
35. reference
36. divide
37. minute
38. pleasant
39. necessary
40. annual
41. career
42. character
43. separate
44. committee
45. immense
46. acquaintance
47. discipline
48. principle
49. judgment
50. lieutenant

The correlation between the odd and even sentences 3 through 34 was +.88 ±.03 for the 55 subjects examined.

The order of difficulty of the words is as follows:[11]

1. he
2. ran
3. well
4. back
5. plant
6. boat
7. pick
8. party
9. eat
10. three
11. apples
12. work (instead of
13. stand school)
14. name
15. rubber
16. race
17. fence
18. yellow
19. church
20. tonight
21. took
22. chicken
23. frock
24. peach
25. health
26. offer
27. dinner
28. step
29. ahead
30. built
31. judge
32. cousin
33. farce
34. honor
35. merchant
36. family
37. recovered
38. objection
39. avoid
40. woman
41. different
42. praise
43. famous
44. gentleman
45. quickly
46. gloomy
47. virtue
48. prefer
49. ought
50. notified
51. series
52. trifle
53. illness
54. employee
55. jealous
56. excellent
57. competition
58. religious
59. occupants
60. especially
61. foreign
62. mysterious
63. alleged
64. valuable
65. secretary
66. minority
67. unusual
68. humorist
69. professor
70. ceremonies
71. solemn
72. develop
73. eminent
74. schedule
75. immediately
76. coaxing
77. politician
78. proprietor
79. acquaintance
80. correspondent
81. fatiguing
82. poultice
83. guarantee
84. grippe
85. occurred
86. nuisance
87. caucus
88. chauffeur
89. undoubtedly
90. incidentally
91. zephyr
92. ptomaines
93. nausea
94. parliamentary
95. aqueous
96. paroxysms
97. anhydrous
98. antonyms
99. fuchsia

[11] An analysis of errors on the Stanford Dictation Test showed the following percentages: omissions, 37 per cent; substitutions, 42 per cent; additions, 9 per cent; reversals, 3 per cent; unclassified, 9 per cent. Of these errors 12 per cent were phonetically justifiable, that is, spellings which rendered a correct version of the word as it is pronounced.

Scores for the test of *Writing a Letter* are scale values of the Nassau County Supplement to the Hillegas Scale, a scaling system which is probably less adequate for the productions of adults than for those of children. The writing of letters was a task which some of the adults did not accept readily, but which would be more generally acceptable if a satisfactory pretext for the letter were arranged.

The distribution of scores on the *Stanford Achievement Arithmetic Computation and Reasoning Tests,* which are given in Tables XVIII and XIX, are both good. The tests are apparently quite adequate for adults of the status of those examined.

TABLE XIX

DISTRIBUTION OF SCORES ON THE STANFORD ACHIEVEMENT ARITHMETIC REASONING TEST

SCORE	MALE	FEMALE	TOTAL
36–39	1		1
32–35	3	2	5
28–31	6		6
24–27	6	1	7
20–23	12	5	17
16–19	6	7	13
12–15	3	1	4
8–11		1	1
Number	37	17	54
Mean	23.86	20.47	22.80
Sigma	5.97	6.05	5.39

3. NON-LANGUAGE TESTS

The large majority of the tests of the *Pintner Paterson Performance Scale* are far too easy for adults from average or near-average levels of the population. A few of them, the Mare and Foal (time score), the Seguin, and the Substitution, were moderately discriminative for the adults examined; and the ranges of scores, means, and sigmas for these have been presented in the third section of Table XI. None of these tests is particularly valuable in studying the normal adult nor interesting to him, but several of them are valuable in studying low-grade or deteriorated cases.

TABLE XX

DISTRIBUTION OF SCORES ON THE PINTNER NON-LANGUAGE MENTAL TEST

SCORE	MALE	FEMALE	TOTAL
525–575	1		1
475–524	7	1	8
425–474	4	3	7
375–424	5	3	8
325–374	8	5	13
275–324	3	5	8
225–274	7		7
175–224	1		1
125–174	1		1
75–124		2	2
Number	37	19	56
Mean	365.78	337.68	356.43
Sigma	110.67	98.26	107.80

TABLE XXI

DISTRIBUTION OF SCORES ON THE PORTEUS MAZE TEST

SCORE	MALE	FEMALE	TOTAL
216–227	1		1
204–215	3	4	7
192–203	2	1	3
180–191	6	3	9
168–179	7	2	9
156–167	3	4	7
144–155	1	2	3
132–143		2	2
120–131		1	1
108–119	1		1
96–107			
84–95		1	1
72–83			
60–71			
48–59		1	1
Number	24	21	45
Mean	177.75	160.00	169.47
Sigma	21.53	39.28	34.60

The *Pintner Non-Language Test* is an interesting one; it presents material and requires performances not involved in the majority of other tests, but any exact determination of the abilities required by the various sub-tests for adult subjects or the reliability of the estimates is impossible. The distributions of total scores are given in Table XX. The distributions of scores on all the sub-tests but the Easy Learning were fairly satisfactory. Experience with the adult subjects showed that they often had difficulty understanding the problem of the Reversed Drawing Test.

The *Cancellation Test* was included as a check on the results of the more complex tests. The distributions were positively skewed.

The scores for the *Kuhlmann Test of Immediate Recall* are based on both time and errors, with the higher scores indicating the poorer performances. The distributions were good. Attempts to determine the patient's method on this test were comparatively fruitless. Observations of behavior and introspections indicated at least three types of attack: visual, kinaesthetic, and verbal. The first two appeared most frequently. Apparently more than one of the three types was often used by the same subject.

The test of *Drawing a Chair* from a model was moderately discriminative, but it is not a test which adults attempt readily, and a few refused to try it at all.

The adults examined often showed a reluctance to attempt the *Good-enough Drawing Test* also. The scores distributed well, but it is highly probable that they are not at all comparable to similar scores obtained by children.

The *Porteus Maze Test,* unlike the drawing tests, was well received by practically all subjects; it is apparently as interesting to adults as to children. The distributions are shown in Table XXI.[12]

[12] It is to be noted that this test, which was added to the battery about a year after the beginning of the study, happened to be given to a relatively smaller proportion of the men and a relatively larger proportion of the women than any other test. This situation does not account for the fairly large sex difference in favor of the men, however, because the men omitted were mostly from the Orthopedic Hospital, that is, the group which tended to give the highest scores.

V

COMPARATIVE FINDINGS FOR MEN AND WOMEN

1. Mean Scores
2. Variabilities

1. MEAN SCORES

THE group of men, containing 47 cases, is about twice the size of the group of women, and this proportion usually holds for the scores on any test. The two groups are similar in age and in educational and occupational status. The mean age for the men is 35 years and the mean for the women 38. Each group falls at the eighth grade in educational level. The mean occupational rating on the Barr Scale for the men is 8.01, for the women 8.77.[1]

The most striking fact evident from a comparison of scores for men and women is the closeness of the means for the two groups on the large majority of tests. As inspection of Table XXII will show, some of the mean scores are practically identical. The slight disparities shown on others probably result from fluctuations of sampling. It is noteworthy, however, that the means for the men are superior to those for the women on all but three of the tests included in the table. None of the differences is statistically significant. Those most nearly approaching significance appear for the Drawing Completion Test of the Pintner Non-Language Scale, the Stanford Binet, the Kuhlmann Immediate Recall, and the Drawing of a Chair, with differences divided by the sigmas of the differences of 2.43, 1.72, 1.49, and 1.49 respectively. It is particularly interesting that no significant sex differences in means were found on either verbal or numerical tests, for such differences have often been reported in the literature, especially in studies of children and adolescents.[2]

[1] The slightly higher mean for women on the Barr Scale is not an altogether satisfactory criterion because the figure may be artificially raised by the rating of 8.37 given the housewives. There is no reason to suppose, however, that the mean educational level for women is not so accurate as that for men, and the equality of the two groups in this regard justifies direct comparison.

[2] It is true, however, that the three tests in which the mean score for women was slightly superior to that for men—the Oral Analogies, the Thorndike Word Knowledge, and the Letters—were all verbal tests, and that four of the five sub-tests of the Stanford Binet in which the women excelled the men—Reading and Free Association at Year X, Dissected Sentences at Year XII, and Differences between Abstract Words at Year XVI—were also verbal.

2. VARIABILITIES

Because of the small groups there is little decisive evidence on the question of differences in variability. The sigmas shown in Table XXII, like the means, are in most cases close. Among the few fairly large differences in absolute variability those appearing for the Substitution, the Kuhlmann Immediate Recall, and the Porteus Maze Test each depend on an extreme case, but that for the Stanford Dictation Test reflects the more platykurtic

TABLE XXII

MEAN SCORES AND SIGMAS FOR MALES AND FEMALES

TEST	MEAN		SIGMA	
	Male	*Female*	*Male*	*Female*
Printed Opposites: Time	71.0	79.6	39.3	22.0
Oral Analogies	20.0	21.2	6.7	6.3
Printed Analogies	21.4	20.3	7.0	6.1
Completion Beta	7.9	7.6	1.3	1.3
Stanford Binet Vocabulary	55.3	52.8	13.1	15.1
Thorndike Word Knowledge	61.5	62.2	21.4	20.5
Stanford Binet Scale	167.9	155.8	26.0	27.9
Gray Oral Reading, Set IV	72.7	75.6	19.0	23.0
Thorndike McCall Reading	24.9	24.2	4.8	4.4
Chapman Reading-Comprehension	16.8	15.8	6.3	7.2
Gates Oral Spelling	26.7	25.1	7.3	6.0
Stanford Dictation	91.0	89.8	23.9	19.5
Writing a Letter	5.1	5.6	1.2	1.0
Arithmetic Computation	26.9	25.3	7.0	7.4
Arithmetic Reasoning	23.9	20.5	6.0	6.1
Seguin	15.3	16.0	4.4	3.3
Substitution	99.9	114.7	29.1	62.9
Pintner Non-Language	365.8	337.7	110.7	98.3
Imitation	6.8	5.5	2.9	2.2
Hard Learning	34.1	32.7	9.3	10.9
Drawing Completion	14.1	12.0	3.0	3.2
Reversed Drawing	5.7	5.1	2.0	2.0
Picture Reconstruction	13.3	11.7	3.8	7.5
Kuhlmann Immediate Recall	159.8	196.0	58.3	83.6
Drawing a Chair	16.8	14.0	5.5	4.3
Goodenough Drawing	28.2	28.2	10.4	11.0
Porteus Maze	177.8	160.0	21.5	39.3

distribution of scores for the men as compared with those for the women. Expressed in terms of Pearson's Coefficient of Variation, the relative variabilities for the remaining tests showing large differences in absolute variability are as follows:

	Males	*Females*
Printed Opposites: Time	56.1	28.9
Gray Oral Reading, Set IV	26.1	30.4
Pintner Non-Language	30.3	29.1
Picture Reconstruction	28.6	64.1

Thus it is evident that the men are relatively much more variable than the women in speed on the Printed Opposites Test, and slightly more variable on the Stanford Dictation Test, while the women are much more variable than the men on the Picture Reconstruction Test.

Findings for men and women with regard to inequalities in the performance level in any individual are reported in Chapter VII. For the sake of completeness, it is to be noted here that there is no appreciable difference between the group of men and the group of women in mean or sigma of the index taken to represent individual inequalities in performance level.

VI

DIFFERENCES IN TEST LEVELS IN DIFFERENT AGE GROUPS

1. FINDINGS FOR THE AGE DECADES TWENTY THROUGH FIFTY

THE material furnishes an interesting basis for the study of the relationship between age and mental test performance; for while the number of cases is small and the results can be no more than tentative, the selection of cases at each age is satisfactory and the range of test performances covered is wide. The cases examined are so distributed according to age that 10 to 20 scores are available on most of the tests for each of the decade groups from the twenties through the fifties. The seventeen tests showing the best distributions and taken by at least 40 subjects were chosen for the survey and the mean scores for the group in each decade calculated. The significance of the differences between the means for successive decades from twenty on, and between the means for the twenties and forties, the thirties and fifties, and the twenties and fifties were computed by Fisher's formula.[1]

The means for the decade groups together with the number of cases on which each is calculated and the correlation of each variable with age are shown in Table XXIII. The first two rows give the mean school grade and the mean Barr occupational rating for each decade. None of the differences between these means is significant.[2] That the mean school grade is slightly higher in the twenties than in any other decade probably results, not from a more selected group, but from changing conditions in education. The mean Barr rating is a little lower in the twenties than in the thirties, forties, or fifties, a condition which probably reflects the shorter occupational experi-

[1] R. A. Fisher, *Statistical Methods for Research Workers*, 105.

[2] Fisher considers a P of .05 or less evidence of a significant difference, but most authorities draw the line at .02, the criterion which has been accepted here. None of the differences in occupational rating and only one of the differences in educational level would be significant by Fisher's criterion: the difference between the mean school grade for the twenties and forties has a P value of .05 to .02.

TABLE XXIII

MEAN SCORES FOR THE AGE DECADE GROUPS AND THE CORRELATION OF EACH VARIABLE WITH AGE

	AGE 20–29		AGE 30–39		AGE 40–49		AGE 50–59		CORRELATION
	N	*Mean*	*N*	*Mean*	*N*	*Mean*	*N*	*Mean*	WITH AGE
School grade reached	14	*9.1*	19	*7.8*	12	*7.3*	12	*7.5*	−.22
Barr Occupational Rating	12	*7.9*	21	*8.5*	14	*8.3*	13	*8.6*	+.16
Tests									
Oral Analogies	10	*23*	9	*24*	9	*18*	10	*18*	−.37
Printed Analogies	14	*24*	16	*23*	11	*19*	12	*18*	−.32
Completion Beta	14	*7.9*	15	*8.1*	12	*7.8*	10	*7.2*	−.19
Stanford Binet Vocabulary	15	*52*	21	*57*	13	*56*	13	*56*	+.16
Thorndike Word Knowledge	14	*53*	14	*68*	12	*64*	8	*73*	+.33
Stanford Binet Scale	14	*166*	20	*173*	12	*160*	12	*151*	−.18
Thorndike McCall Reading	13	*25*	17	*26*	11	*25*	11	*23*	−.07
Gates Oral Spelling	13	*25*	13	*29*	11	*26*	10	*26*	−.05
Stanford Dictation	11	*85*	17	*96*	13	*88*	11	*91*	+.01
Arithmetic Computation	13	*27*	18	*27*	11	*25*	7	*27*	−.05
Arithmetic Reasoning	13	*23*	17	*25*	10	*22*	7	*22*	−.03
Mare and Foal	9	*25*	17	*30*	11	*31*	7	*34*	−.42
Seguin	9	*15*	14	*15*	10	*16*	7	*19*	−.38
Substitution	9	*91*	15	*100*	10	*101*	5	*169*	−.43
Pintner Non-Language	14	*391*	16	*356*	10	*324*	9	*280*	−.53
Kuhlmann Immediate Recall	14	*175*	11	*154*	8	*169*	5	*201*	−.25
Porteus Maze	12	*178*	10	*174*	8	*162*	11	*156*	−.35

ence of the subjects in the twenties. The close similarity of all four decades from the point of view of educational and occupational status indicates that they may justly be compared for test performances, and that differences in test performances may be attributed to differences in age.[3]

Calculation of the P values for the differences between the mean test scores reported in Table XXIII shows no significant differences between means at any one-decade or two-decade interval. For the three-decade interval from twenty to fifty, however, there is a significant rise in the case of the Thorndike Test of Word Knowledge and there are significant declines in the case of the Printed Analogies and the Seguin Formboard.

From inspection of Table XXIII it will be evident that, while the differences in means are in most cases comparatively slight, there is a very interesting division of the tests according to their course with age. All the language tests, with the exception of the Printed Analogies, show higher means in the thirties than in the twenties; while the non-language tests, with the exception of the Kuhlmann Immediate Recall, show no such rise.[4] Their means tend rather to decline steadily from the twenties through the fifties, and the Printed Analogies is like the non-language tests in this respect. The means for the language tests all decline from the thirties to the forties, but there is not always a further decline from the forties to the fifties, and in the case of the Thorndike Word Knowledge and the Stanford Dictation Tests there are rises from the forties to the fifties. On the Thorndike Test of Word Knowledge the mean score in the fifties is higher than that in any other decade.[5]

[3] It is assumed that for the tests analyzed here sex differences are negligible not only for the group as a whole but for the decade groups. The male and female decade groups are too small to be treated separately, but the means for the men alone, which are generally derived from the larger number of cases, show approximately the same course with age as those for the total group. This is an important point because the proportion of women differs somewhat in certain decades, and if sex differences for the decade groups were not negligible, the means of all four decades would not be directly comparable. On most of the tests analyzed there are a few more scores for women than for men in the fifties, one or two less in the twenties, and approximately a third the number of scores for women as for men in the thirties and forties.

[4] Medians for the separate decades reveal almost exactly the same relationships as means in the case of all but two tests: the Stanford Binet and the Completion Beta. Medians for both of these are highest in the twenties and decline slightly in each successive decade. The fact that the highest medians fall in the twenties and the highest means in the thirties results from differences in the distributions in these small groups; on each test there are a larger number of cases reaching high scores in the thirties than in the twenties, and these high scores have, of course, a greater effect on the means than on the medians.

[5] In the study of a large group of adults tested in the course of an educational program, Grace obtained data on the Thorndike Test of Word Knowledge which may be cited in comparison with those reported here. (A. G. Grace, *Individual Differences in Adults,* J. Educ.

2. FINDINGS FOR THE AGE GROUPS TWENTY THROUGH THIRTY-FOUR AND FORTY-FIVE THROUGH FIFTY-NINE

When two larger age groups are selected for comparison, the one composed of all subjects aged twenty through thirty-four and the other of all subjects aged forty-five through fifty-nine, the mean score for the older group is significantly lower than that for the younger in the case of four tests: the Printed Analogies, the Seguin, the Substitution, and the Pintner Non-Language. The comparative findings are shown in Table XXIV. In no case is the mean score for the older group significantly higher than that for the younger. On both vocabulary tests, however, the means for the older group are slightly higher than those for the younger, and on four other language tests—the Completion Beta, the Thorndike McCall Reading, the Gates Oral Spelling, and the Stanford Dictation—and on both arithmetic tests, the means for the older fall only slightly below those for the younger.

Differences in variability between the younger and older groups are on the whole small. In 10 of 17 cases the older group is apparently the more variable. The quartile deviations for the two groups on the tests of the preceding analysis are as follows:

	Younger group	*Older group*
Oral Analogies	3.0	6.3
Printed Analogies	4.5	5.3
Completion Beta	.8	1.1
Stanford Binet Vocabulary	7.8	11.8
Thorndike Word Knowledge	15.8	13.0
Stanford Binet	12.8	22.3
Thorndike McCall Reading	4.5	3.3
Gates Oral Spelling	3.3	5.8
Stanford Dictation	20.1	16.8
Arithmetic Computation	5.5	4.5
Arithmetic Reasoning	5.8	3.8
Mare and Foal	4.3	4.5

Psychol., 1932, 23, 179–186.) If the scores from his table are grouped in decade periods, there are 1,030 for ages under twenty, 889 for the twenties, 224 for the thirties, 71 for the forties, and 10 for the fifties The distributions seem to indicate that there were different types of sample at decades forty and fifty. Calculation of the mean scores for the three younger age groups shows them to be 65.1, 71 2, and 60.1 respectively. The differences in mean score from below twenty to the twenties and from the twenties to the thirties are both significant, with differences divided by the sigmas of the differences of 7 4 and 5.8. Thus Grace's findings show an early and significant rise in average test score followed by a decline which is also significant and which comes at the twenty-thirty interval.

	Younger group	*Older group*
Seguin	2.5	2.0
Substitution	14.5	23.5
Pintner Non-Language	51.0	62.3
Kuhlmann Immediate Recall	47.8	41.5
Porteus Maze	13.5	14.3

The problem of age and test performance will be taken up again in Chapter IX, after the correlations with age have been discussed, but in this connection it may be noted, first, that differences in mean score on the test performances analyzed are comparatively slight between the third and the sixth decade. Performances involved in the language tests, with the exception of the analogies, are all relatively well maintained with age. In the case

TABLE XXIV

MEAN SCORES FOR AGES TWENTY THROUGH THIRTY-FOUR AND FORTY-FIVE THROUGH FIFTY-NINE, AND VALUES FOR THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE MEANS

	AGE 20–34		AGE 45–59		
	N	*Mean*	*N*	*Mean*	P VALUES
School grade reached	26	*9.0*	22	*7.6*	.6–.5
Barr Occupational Rating	25	*8.4*	25	*8.7*	.6–.5
Tests					
Oral Analogies	16	*22*	18	*19*	.1–.05
Printed Analogies	24	*24*	22	*18*	0–.01
Completion Beta	21	*8.0*	22	*7.5*	.3–.2
Stanford Binet Vocabulary	28	*55*	24	*56*	.9–.8
Thorndike Word Knowledge	24	*59*	19	*68*	.2–.1
Stanford Binet Scale	26	*172*	23	*155*	.1–.05
Thorndike McCall Reading	25	*25*	21	*24*	.6–.5
Gates Oral Spelling	22	*27*	19	*25*	.8–.7
Stanford Dictation	23	*93*	22	*92*	.9–.8
Arithmetic Computation	25	*28*	17	*26*	.4–.3
Arithmetic Reasoning	24	*24*	16	*22*	.8–.7
Mare and Foal	20	*26*	17	*33*	.1–.05
Seguin	18	*14*	16	*18*	0–.01
Substitution	19	*86*	14	*127*	0–.01
Pintner Non-Language	25	*387*	18	*303*	0–.01
Kuhlmann Immediate Recall	23	*167*	13	*204*	.2–.1
Porteus Maze	17	*177*	19	*158*	.2–.1

of the vocabulary tests there are even slight increases in the mean scores of the older groups, and on the Thorndike Word Knowledge the mean score for the fifties is significantly higher than that for the twenties. The second important point to be noted is that the most marked changes with increasing age are the declines shown on the analogies and on the non-language tests. On the Printed Analogies and on the Seguin Formboard the decreases in mean score between the twenties and the fifties are significant. When the larger age groups twenty through thirty-four and forty-five through fifty-nine are compared, there are again significant differences in mean score on these tests and in addition significant decreases in mean score on the Substitution and on the Pintner Non-Language Tests. These declines cannot in every case be attributed to the more generally recognized characteristics of behavior changes with age, for example, to diminishing speed or manual dexterity. Neither of these factors enters into the Printed Analogies Test, for example. It is possible that older persons have increasing difficulty with the analogies or the non-language tests because they become less able to adapt to the particular situations involved; it is also possible that the peaks in the development of the performances involved are actually reached early in life. In all probability the latter is the better explanation, for observations of the patients and their ready adaptation to most of the test situations in question, as well as the character of some of these situations, would seem to rule out the first possibility.

VII

THE RELATIONSHIP BETWEEN VARIOUS PERFORMANCE LEVELS IN THE INDIVIDUAL CASE

1. METHOD OF ANALYSIS

WITH records for each case on a large and varied battery of tests, the material was particularly well suited to a study of the equalities or inequalities in performance levels in the individual. For adults studies of this sort have never been carried far because of the brief examinations usually employed. Yet the problems which may be attacked by such an analysis are of great interest. It is important to know how much variation in performance level the individual shows on a variety of tests of intelligence and educational achievement; what the relationship is between this variation and sex, age, education, and the grade of the performance as indicated by test score; and finally what facts the individual case records will yield as to characteristic constellations in mental test performance, that is, groupings which appear when certain test scores are high in comparison to other test scores.

To make a survey of different test performances it is, of course, necessary to convert the various test scores into measures which will be directly comparable. Standard measures were computed from the constants of the distributions of the total populations.[1] Since the groups were small and the distributions only approximately normal, the technique of using standard measures is not altogether justifiable, but it furnishes a satisfactory indication of the trend of the results. Only tests giving fairly good distributions were considered for the analysis, and from among these ten were chosen which would cover a number of different types of activity: the Kelley Trabue Completion Beta, Stanford Binet Vocabulary, Printed Analogies,

[1] The formula used was:

$$\text{Standard measure} = \frac{\text{gross measure} - \text{mean}}{\text{sigma}}$$

Kelley, *Comparable Measures*, J. Educ. Psychol., 1914, 5, 589–595.

Thorndike McCall Reading, Gates Oral Spelling, Letters, Arithmetic Computation, Pintner Non-Language, Kuhlmann Immediate Recall, and Porteus Maze Tests.

Twenty of the women, excluding the 3 who took fewest tests, and 20 of the men selected at random were chosen as the first groups. When certain marked differences between the men and the women were found in the correlations between the index of individual variation and other variables, a second group of men was selected, including the 20 of the 27 remaining male subjects who had taken the greatest number of tests. The standard measures for each of the tests selected, the mean of these measures, and their average deviation from the mean standard measure were determined for each individual. This average deviation (A.D.) was taken as a measure of the variations in performance levels in the individual case, and was correlated with mean standard measure, age, and education. When the statistical analyses had been completed, the individual case records were plotted on graphs showing the relative position of the standard measures for the ten tests analyzed, and these graphs were then compared.

2. RESULTS

The mean standard measures and the means and sigmas of the average deviations for the three groups are as follows:

	Mean standard measure	*Mean A.D.*	*Sigma A.D.*
Females	−.10	.49	.17
Males, Group A	.00	.44	.17
Males, Group B	+.13	.50	.24

There is a close correspondence between the three groups in the means of the average deviations, but the women are relatively low in mean standard measure and the men of the second group relatively high. The men of this Group B are also slightly more variable than the subjects of the other groups.

The correlations for each group and for the two groups of men combined are given in Table XXV. For such small numbers the st ndard errors are, of course, high.

According to Fisher's criterion of the significance of a correlation, only the correlation of −.53 between A.D. and education in Group A, which

has a P value of .02, can be taken to indicate a definite relationship.[2] It seems probable that the higher correlations throughout this group depend on some factor in its selection. The findings as a whole indicate that there is little, if any, relationship between inequality in test performance as measured by average deviation and general level of test performance, age, or education.

Studies of the individual case records are interesting. Thirty-two of the 60 cases analyzed (23 of the men and 9 of the women) make standard scores on all tests within plus or minus one sigma of their mean standard measures. In other words, slightly more than half the group and an almost equivalent proportion of men and women show no great individual variation in the ten test performances. It must be noted that these include not

TABLE XXV

CORRELATIONS OF THE AVERAGE DEVIATION OF THE STANDARD MEASURES WITH OTHER VARIABLES

	FEMALES	MALES		
		Group A	*Group B*	*Groups A and B*
Mean standard measure	$-.001 \pm .23$	$-.49 \pm .17$	$-.04 \pm .23$	$-.21 \pm .15$
Age	$+.22 \pm .21$	$+.34 \pm .20$	$+.12 \pm .22$	$+.18 \pm .15$
School grade reached .	$-.004 \pm .23$	$-.53 \pm .16$	$-.01 \pm .23$	$-.17 \pm .15$

NOTE: Each correlation is followed by its standard error.

only tests manifesting a fairly high degree of relationship, as reported in Chapter VIII, but also tests such as the Porteus Maze and the Kuhlmann Immediate Recall which are much less closely related to the majority of the verbal tests. Comparison of these cases with those giving one or more test performances which fall more than one sigma from the mean standard measures shows slight differences in age and education; as one would expect, the subjects with the smaller variations are on the average younger (three years in the case of the men and two in the case of the women) and slightly higher in educational level (half a year for the men and a year for the women) than the subjects with one or more performances deviating beyond one sigma from their respective mean standard measures. The situa-

[2] *Op. cit.*, Table V, A.

tion has been suggested by the correlations with average deviation reported in Table XXV, but only one of these was found to be significant.

Among the 28 cases whose performances do not fall within plus or minus one sigma of the mean standard measure, 11 deviate more widely than this in only one performance. These single and relatively large deviations appear more often for the Porteus Maze than for any other test (4 cases).

Inspection of the graphs showing the mean standard measures for each case reveals no characteristic constellations among the test performances analyzed. There are a few cases showing relatively high or relatively low non-language performances, as indicated by the three non-language tests, but for the group at large there are no clear-cut subdivisions in the pattern of the variations. Nor is there any indication of generally high or generally low performances on the so-called educational achievement tests in cases in which the educational level is considerably above or considerably below the average.[3]

The absence of large deviations in the majority of cases is a demonstration of the fact that adults who do well on some of the tests in question usually do well on the others, and conversely that those who have difficulty with some tests are likely to have difficulty with others. The further evidence from the analysis of individual case records shows that particular constellations of abilities, as indicated, for example, by high scores on non-language tests with low scores on language, are the exception rather than the rule; for the most part, the adult subjects show relatively small deviations from their respective mean standard measures, and these have an irregular character, or at least one which cannot be ordered according to a logical analysis of the nature of the tests.

[3] The relative performance levels on educational achievement and intelligence tests will be analyzed in the section on Education and Test Performance in Chapter IX, where a number of graphs are presented showing various performance levels in individual cases.

VIII

CORRELATIONAL FINDINGS

1. Correlations with Age
2. Correlations with Education
3. Correlations with Occupation
   a. Chi-Square determinations, Taussig ratings
   b. Correlations with Barr Scale ratings
4. Intercorrelations between the Most Important Tests
   a. Correlations with the Stanford Binet Scale
   b. Correlations with the Kelley Trabue Completion Exercise Beta
   c. Correlations with the Vocabulary Tests
   d. Correlations with the Thorndike McCall Reading Scale
   e. Correlations with the Gates Oral Spelling and the Stanford Achievement Dictation Tests
   f. Correlations with the Printed Analogies Test
   g. Correlations with the Stanford Achievement Arithmetic Tests
   h. Correlations with the Pintner Non-Language Test
   i. Correlations with the Goodenough Drawing Test
   j. Correlations with the Porteus Maze Test

1. CORRELATIONS WITH AGE

IN general the correlations between age and non-verbal tests are negative and of relatively high degree, ranging up to $-.53$, while those between age and verbal tests are of relatively low degree, but still negative. As determined by Fisher's criterion, the only reliable positive correlation between age and a verbal test appeared for the Thorndike Word Knowledge.[1] Reliable negative correlations appeared for the Oral and Printed Analogies, the time score on the Mare and Foal, the Seguin, the Substitution, the Pintner Non-Language, and the Porteus Maze Tests. Correlations between age and thirteen of the more important tests of the battery are presented in Table XXVI.

Apparently the wide range of the group had little effect on the degree of

[1] Since the numbers were in each case comparatively small, the standard errors were high. Each correlation was tested for significance by Fisher's formula (*op. cit.*, Table V, A) and was taken to indicate the presence of an actual relationship between the two variables only if it had a P value of .02 or less for a group of the given size.

TABLE XXVI

CORRELATIONS OF TEST RESULTS WITH AGE, SCHOOL GRADE REACHED, AND OCCUPATIONAL RATING ON THE BARR SCALE

TEST	AGE		SCHOOL GRADE REACHED		BARR SCALE	
	N	r	N	r	N	r
Stanford Binet Scale	65	$-.18\pm.13$	61	$+.62\pm.08$	55	$+.44\pm.11$
Completion Beta	57	$-.19\pm.13$	53	$+.60\pm.09$	51	$+.07\pm.14$
Stanford Binet Vocabulary	69	$+.16\pm.12$	64	$+.49\pm.10$	61	$+.52\pm.09$
Thorndike Word Knowledge	54	$+.33\pm.12$	50	$+.46\pm.12$	47	$+.59\pm.10$
Thorndike McCall Reading	59	$-.07\pm.13$	56	$+.63\pm.08$	51	$+.52\pm.10$
Gates Oral Spelling	53	$-.05\pm.14$	50	$+.59\pm.09$	46	$+.48\pm.12$
Stanford Dictation	58	$+.01\pm.13$	50	$+.62\pm.09$	52	$+.39\pm.12$
Printed Analogies	60	$-.32\pm.11$	56	$+.58\pm.08$	52	$+.38\pm.12$
Arithmetic Computation	55	$-.05\pm.14$	49	$+.67\pm.08$	48	$+.49\pm.11$
Arithmetic Reasoning	54	$-.03\pm.14$	51	$+.64\pm.08$	46	$+.61\pm.09$
Pintner Non-Language	55	$-.53\pm.10$	50	$+.56\pm.10$	47	$+.33\pm.13$
Goodenough Drawing	53	$-.26\pm.13$	48	$+.30\pm.13$	48	$+.21\pm.14$
Porteus Maze	45	$-.35\pm.13$	41	$+.57\pm.11$	41	$+.12\pm.16$

NOTE: Each correlation is followed by its standard error. The correlations were calculated from a raw score formula; those in italics are reliable.

the test intercorrelations, for when age was partialed out of a number of those for which the scatters were definitely linear, changes in the degrees of the coefficients were slight.[2] Typical examples are as follows:[3]

Thorndike McCall—Arithmetic Computation . . +.66, age out +.67
Thorndike McCall—Arithmetic Reasoning . . +.66, age out +.67
Chapman Reading—Arithmetic Reasoning . . +.76, age out +.77
Printed Analogies—Arithmetic Reasoning . . . +.68, age out +.68
Printed Analogies—Porteus Maze . . . . . +.62, age out +.58
Pintner Non-Language—Completion Beta . . +.56, age out +.58
Pintner Non-Language—Porteus Maze . . . . +.58, age out +.47

2. CORRELATIONS WITH EDUCATION

The correlation between education and age for the 64 cases is $-.22 \pm .12$; it is unreliable.

[2] The scatter diagrams for the age correlations show only one which is definitely curvilinear, that with the Stanford Binet Scale. Three others may be curvilinear: those with the Stanford Binet Vocabulary, the Completion Beta, and the Arithmetic Computation Tests. Eta was not computed.

[3] See also intercorrelations for the sub-tests of the Pintner Non-Language Scale with age partialed out, Table XXX, page 99.

Correlations between education and test score were calculated for nineteen of the tests of the battery. In all cases the correlations are positive and in all but two cases they are reliable; the two exceptions are those for the Kuhlmann Immediate Recall and the Goodenough Drawing Tests. Thirteen of these correlations are given in the second column of Table XXVI.

TABLE XXVII

PARTIAL CORRELATIONS INDICATING THE RELATIVE IMPORTANCE OF THE FACTORS OF SCHOOL GRADE REACHED AND TEST INTELLIGENCE IN VOCABULARY AND EDUCATIONAL ACHIEVEMENT TEST SCORES

TEST	EDUCATION		STANFORD BINET	
	First order	*Second order**	*First order*	*Second order†*
Thorndike Word Knowledge	.46±.11	.12±.14	.62±.09	.48±.12
Gray Oral Reading	.36±.13	.18±.14	.39±.13	.22±.14
Thorndike McCall Reading	.63±.08	.38±.12	.75±.06	.62±.09
Chapman Reading-Comprehension	.63±.10	.36±.14	.81±.06	.46±.08
Gates Oral Spelling	.59±.09	.20±.13	.68±.08	.57±.12
Stanford Dictation	.60±.09	.23±.13	.74±.06	.57±.10
Arithmetic Computation	.67±.08	.49±.11	.67±.08	.44±.12
Arithmetic Reasoning	.59±.09	.26±.13	.81±.05	.72±.07

NOTE: Each correlation is followed by its standard error.

* Stanford Binet out. † Education out.

The relatively high correlations between scores on the tests and school grade reached do not necessarily indicate that success in the tests is a function of education, for education may as well be considered the dependent as the independent variable and some of the variations in test score with education may depend on the relationship existing between intelligence and the amount of education. If the Stanford Binet score be considered an index of intelligence, it is possible to determine by the partial correlation technique the degree of relationship between education and test score when the effect upon these of the variation in intelligence is discounted. It seemed worth while to calculate these partial correlations for the Thorndike Test of Word Knowledge and for the various educational achievement tests, success in all of which might be expected to depend largely on the amount of

formal education. The results are presented in the first two columns of Table XXVII.

The correlations with the Thorndike Test of Word Knowledge show that when the variability in word knowledge resulting from differences in test intelligence is ruled out, no definite relationship remains between word knowledge and education. The correlations for four other tests—the Gray Oral Reading, the Gates Oral Spelling, the Stanford Dictation, and the Arithmetic Reasoning—are also reduced to the point where little relationship remains. The Thorndike McCall Reading is reduced by 25 points, the Chapman Reading-Comprehension by 27, and the Arithmetic Computation by 18. Expressed in other terms, it may be said that for this group the variation in scores resulting from the variation of education is:

Thorndike Word Knowledge	00.8 per cent
Gray Oral Reading	01.6 per cent
Thorndike McCall Reading	07.4 per cent
Chapman Reading-Comprehension	06.7 per cent
Gates Oral Spelling	02.1 per cent
Stanford Dictation	02.7 per cent
Arithmetic Computation	12.9 per cent
Arithmetic Reasoning	03.5 per cent

The Arithmetic Computation Test apparently depends most greatly on school training, while the two reading comprehension tests show the next closest dependence.

The facts as a whole indicate definitely that the influence of intelligence outweighs that of education.[4] The situation may be represented in another way by using the partial correlation technique to rule out the variations in test score resulting from differences in amount of education. These partial correlations are presented in the third and fourth columns of Table XXVII. The second order correlations with education partialed out show much less reduction than the corresponding correlations with Stanford Binet score partialed out.

For these tests the variations resulting from differences in intelligence level as indicated by the Stanford Binet scores, are:

Thorndike Word Knowledge	12.4 per cent
Gray Oral Reading	02.5 per cent

[4] The examiners had no reason to suspect that the more intelligent patients reported the number of years in school correctly while the less intelligent falsified the data.

Thorndike McCall Reading	21.2 per cent
Chapman Reading-Comprehension	30.4 per cent
Gates Oral Spelling	11.0 per cent
Stanford Dictation	17.6 per cent
Arithmetic Computation	10.1 per cent
Arithmetic Reasoning	30.5 per cent

A comparison of these figures with those given above for the variation in test score with education shows that for every test except the Arithmetic Computation, the percentage of variation with the Stanford Binet score is higher than that with the school grade completed. For most of the tests, including the Thorndike Test of Word Knowledge, the Thorndike and the Chapman reading comprehension tests, the Stanford Dictation, and the Arithmetic Reasoning, the differences are marked.

3. CORRELATIONS WITH OCCUPATION

a. *Chi-Square determinations, Taussig ratings*

The degree of association between location in the Taussig classification and score on some of the tests was studied by means of the Chi-Square technique.[5] Group I, Professional, and Group II, Semi-professional and Business, were combined for this purpose, as were Groups IV and V, Semi-skilled and Unskilled labor. Scores on the tests were thrown into two classifications, above and below the mean. The tests used and the P values for the Chi-Square of each are given below:

	N	*Chi-Square*	*P values*
Taussig—Age	52	7.15	.03
Taussig—Education	47	3.85	.15
Taussig—Stanford Binet	48	7 66	.05–.02
Taussig—Completion Beta	43	16.74	0–.01
Taussig—Stanford Binet Vocabulary	52	10.49	0–.01
Taussig—Thorndike McCall Reading	44	11.38	0–.01
Taussig—Gates Spelling	38	6.33	.10–.05
Taussig—Printed Analogies	44	3.22	.20
Taussig—Arithmetic Computation	44	7.48	.10–.05
Taussig—Seguin	36	2.76	.3 –.2
Taussig—Substitution	36	1.28	.7 –.5
Taussig—Pintner Non-Language	41	.70	.9 –.8
Taussig—Porteus Maze	33	.78	.7 –.5

[5] Housewives are not classified in any of the Taussig groups, so that the subjects in this analysis are mostly men.

The results indicate a division between the verbal and the non-verbal tests; the Printed Analogies falls with the non-verbal group. There is a significant association between Taussig classification and scores on the Completion Beta, the Thorndike McCall Reading, and the Stanford Binet Vocabulary Tests, and probably a significant relation between the Taussig and the Stanford Binet Scale as a whole. Among the verbal tests, only the Gates Spelling and the Printed Analogies fail to show a real association with the Taussig classification. Of the non-verbal tests, not one is associated with this classification of occupations.

The presence or absence of association is strikingly evident from the contingency tables, three of which are reproduced in Table XXVIII.

TABLE XXVIII

CONTINGENCY TABLES FOR TAUSSIG RATINGS AND VARIOUS TEST SCORES

	TAUSSIG GROUPS			
	I and II	*III*	*IV and V*	*Total*
Stanford Binet scores below 160 . . .	3	6	13	22
Stanford Binet scores above 160 . . .	13	6	7	26
Total	*16*	*12*	*20*	*48*
Completion Beta scores below 7.80 . .	1	5	16	22
Completion Beta scores above 7.80 . .	12	5	4	21
Total	*13*	*10*	*20*	*43*
Pintner Non-Language scores below 350	6	5	10	21
Pintner Non-Language scores above 350	7	6	7	20
Total	*13*	*11*	*17*	*41*

b. *Correlations with Barr Scale ratings*

The correlation between Barr ratings and age is $+.16 \pm .13$, and that between Barr ratings and education $+.48 \pm .11$. The latter is reliable.

Correlations with Barr ratings were computed for one or more of the tests in each group, vocabulary, reading, and so on. They are all positive and range from .07 $\pm .14$ with the Completion Beta to .71 $\pm .09$ with the Stanford Arithmetic Reasoning. It is interesting to find that the next to the lowest of those correlations calculated is $+.12 \pm .16$ with the Porteus Maze Test. Thirteen of the correlations with Barr Scale ratings are presented in the third column of Table XXVI.

4. INTERCORRELATIONS BETWEEN THE MOST IMPORTANT TESTS

Of the 666 intercorrelations possible for the 37 tests reported in this monograph, 172 were calculated. In addition to the intercorrelations of the most important tests of the battery, this number included correlations calculated for some particular purpose, for example, to find the degree of correspondence between oral and printed forms of a test, or to determine the interrelationship of sub-tests in a scale.[6] For economy in presentation only the intercorrelations of the most important tests will be reported. These are shown in Table XXIX. The first eight tests would generally be considered predominantly verbal, and their intercorrelations are relatively high. It is to be noted, however, that the correlations with the Printed Analogies are considerably lower than those with the other tests of this group. They are lower than the correlations between the language and the arithmetic tests, though higher than the correlations between the language and the non-language tests.

a. *Correlations with the Stanford Binet Scale*

Reference to Table XXIX will show that the Stanford Binet, the composite test of so-called general intelligence, does indeed show the highest inter-test correlations, on the average, of any test reported.[7] The correlations with the Printed Analogies and the two arithmetic tests are about as high as those with tests of the language block. The scatter diagram for the correlation between the Stanford Binet and the Printed Analogies Tests, which is presented in Figure 1, shows the relationship to be fairly close; most of the scores cluster near the regression line. The same clustering is apparent in the diagram for the Stanford Binet correlation with the Arithmetic Computation, and to an even greater extent in that for the correlation with the Arithmetic Reasoning Test. The correlation between the Stanford Binet and the Pintner Non-Language Test, $+.69 \pm .07$, is high in comparison to the findings of other authors on children. Morganthau reported a

[6] A few correlations with the Cancellation Test were calculated in order to determine what relationships existed between this comparatively simple test and the more complex intelligence tests. The correlation with the Stanford Binet Scale was $+.23 \pm .14$ for 46 cases, with the Completion Beta $+.40 \pm .13$ for 43 cases, and with the Pintner Non-Language Test $-.06 \pm .16$ for 40 cases.

[7] The highest correlation with the Stanford Binet is that of the Chapman Reading-Comprehension Test: $+.83 \pm .05$ for 42 cases.

correlation of $+.44 \pm .06$ for 97 children, aged ten to sixteen.[8] The difference might result from the wider range in the adult group, but this explanation seems unlikely in view of the fact that the Stanford Binet shows a low negative relation to age and the Pintner Non-Language Test a fairly high

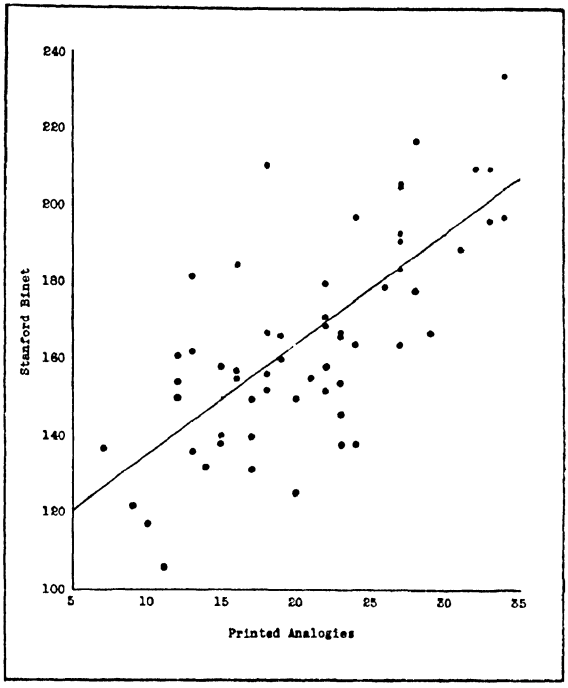

Figure 1. Correlation between the Stanford Binet and the Printed Analogies Tests:
$r = +.72$; $b = +2.90$

[8] D. Morganthau, *Some Well-Known Mental Tests Evaluated and Compared,* Arch. Psychol., 1922, 7, No. 52, 38.

Walters found an even lower correlation, +.13, for a group of 165 children in grades VI B and VII A. The Stanford Binet correlations with the sub-tests of the Pintner Non-Language were —.04 with Imitation, — 11 with Hard Learning, +.23 with Drawing Completion, +.32 with Reversed Drawing, and + 26 with Picture Reconstruction. It is to be noted, however, that while nearly all of these children were born in the United States, more than two-thirds of them came from families of foreign extraction, so that there may have been inequalities in facility with the English language. (F. C. Walters, *A Statistical Study of Certain Aspects of the Time Factor in Intelligence,* Teach Coll. Contri. to Educ, 1927, No. 248, 34–35.)

negative relation. It is impossible to check by partialing because of the curvilinear relationship between age and Stanford Binet score.

The correlation between the Stanford Binet and the Goodenough Test of Drawing a Man, on the contrary, is much lower than that usually found for

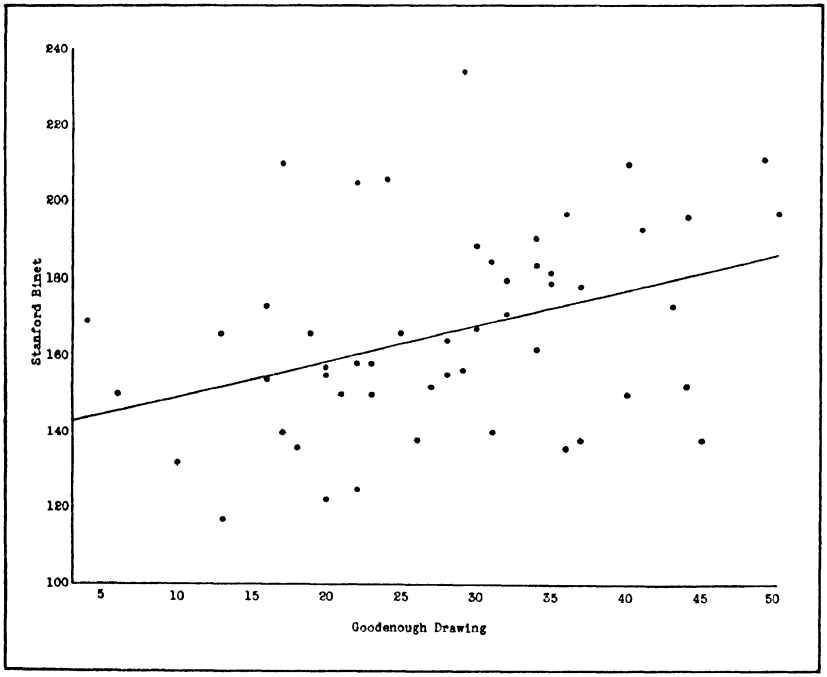

Figure 2. Correlation between the Stanford Binet and the Goodenough Drawing Tests: r = +.39; b = +.87

children. It is 17 to 46 points lower than the correlations reported by Goodenough for children aged four to ten.[9] The scatter diagram, reproduced in Figure 2, shows how slight the relationship between the two test performances is at the adult level.

The correlation between the Stanford Binet and the Porteus Maze Test, +.59 ±.10, is about the same or slightly lower than the correlation reported by other authors for children. Morganthau found a correlation of

[9] F. L. Goodenough, *Measurement of Intelligence by Drawings*, 51.

+.54 ±.05 for her 110 boys and girls;[10] Gaw a correlation with Stanford Binet I.Q. of +.52 for 52 boys and one of +.29 for 48 girls;[11] and Worthington a correlation of +.75.[12]

The scatter diagram for the 43 adults shows a rather extensive spread around the regression line, with some fanning above a score of about 144 on each test. The cases are too few to make a definite conclusion possible, but the tendency is in line with the statements Porteus has made about the closer correspondence of Stanford Binet and Porteus scores in the case of low-grade than of high-grade subjects.[13]

In view of the fact that some of the tests in this research were more satisfactory than the Stanford Binet for the adult group from the point of view of their acceptability to the subject and their discriminative value, several batteries requiring about as much time as the Binet were set up and multiple correlations calculated. The groups are, of course, too small for the results to be more than suggestive. One battery contained three tests which were each satisfactory individually: the Printed Analogies (X_2), the Completion Beta (X_3), and the Thorndike Word Knowledge (X_1). The findings for the battery were:

$$R_{1\,234} = +.860$$
$$X_1 = 1.32X_2 + 7.83X_3 + .33X_4 + 57.14$$
$$\sigma_{1\,234} = 11.73$$

The amount of variation in Stanford Binet score resulting from variation in the other three tests was 49 per cent.

Since a close relationship with the subject is valuable in many cases, either because it affords an opportunity for study of his reactions or because it ensures better cooperation, an oral battery was tried. The tests included were the Stanford Binet Vocabulary (X_2), the Oral Analogies (X_3), and the Oral Absurdities (X_4). The findings were as follows:

$$R_{1\,234} = +.914$$
$$X_1 = 1.08X_2 + .94X_3 + 3.18X_4 + 34$$
$$\sigma_{1\,234} = 12.38$$

Sixty-two per cent of the variation in Stanford Binet score resulted from variation in the three tests of the battery.

As a criterion against which to check other tests or batteries of tests, the Stanford Binet is far from perfect. The difficulties are particularly serious in work

[10] *Op. cit.*, 38.

[11] F. Gaw, *A Study of Performance Tests*, Brit J. Psychol., Gen. Sec., 1924–25, 15, 385–386.

[12] M. R. Worthington, *A Study of Some Commonly Used Performance Tests*, J. Appl. Psychol., 1926, 10, 221.

[13] S. D. Porteus, *Maze Tests and Mental Differences*, 84.

with adults. It is probable that some of the sub-tests show improvement with increasing age after maturity while others show little change and still others decline. If so, the same total score would not have the same significance at different ages. This difficulty is, of course, present in any composite score for a battery of tests such as those discussed above, and for this reason, whenever the battery is not homogeneous, it is undoubtedly better to report the test results separately than to employ a composite score.

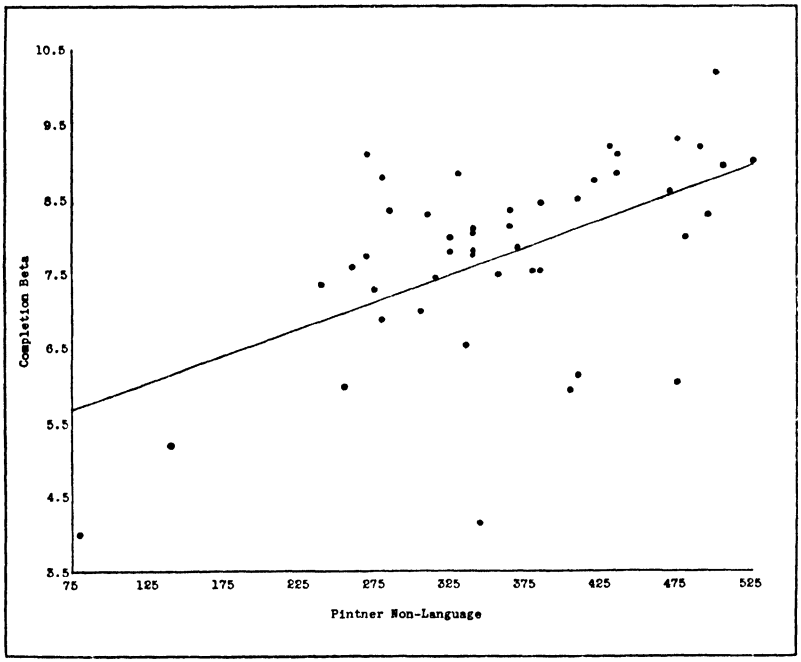

Figure 3. Correlation between the Kelley Trabue Completion Beta and the Pintner Non-Language Tests: r = +.56; b = +.007

With results for the three tests reported separately, it is probable that the first battery would make a satisfactory brief examination of intelligence. The Completion Beta is the most valuable language test of the battery as a whole: it interests adult subjects and it has more discriminative power at the upper levels than any other. The Thorndike Test of Word Knowledge is discriminative for adults from the middle levels of the population and, while it probably includes a certain number of doubtful items, it is on the whole a good test. In relation to these two, both of which demand activities generally considered as excellent indices of verbal intelligence, the Printed Analogies is more or less an unknown

quantity. It probably introduces factors not present in the other test performances, but the nature of these factors cannot as yet be defined.

For old or uncooperative subjects or any others requiring individual work, the oral battery would probably be useful. It would undoubtedly be improved greatly, however, by an extended Oral Absurdities Test. All three tests of this battery are easily administered and well received, and in cases showing resistance to an intelligence test the battery may be introduced as a language test.

b. *Correlations with the Kelley Trabue Completion Exercise Beta*

The correlations between the Completion Beta and the language tests, including the Printed Analogies, are fairly high.[14] Those with the arithmetic tests are also fairly high. The relationships shown by the scatter diagrams are consistent throughout the ranges. All nine of these correlations are reliable.

The correlation of $+.56 \pm .10$ with the Pintner Non-Language Test is also reliable.[15] The scatter diagram, which is shown in Figure 3, brings out an interesting situation: the relationship is more marked when the Completion Beta score is over 6.50; lower Beta scores may be associated with high scores on the Non-Language Test.

In view of the relatively high correlation which both the Completion Beta and the Porteus Maze show with the Stanford Binet, and also with the Printed Analogies Test, the Beta-Porteus correlation of only $+.10 \pm .16$ is of considerable interest. It is not easily interpreted. The diagram, presented in Figure 4, shows a diffuse rectilinear scatter, making the lack of relationship between the two variables quite apparent.

c. *Correlations with the Vocabulary Tests*

The first six correlations with each of the vocabulary tests are reliable and fairly high.[16] The correlation of $+.88 \pm .03$ between the Stanford Binet Vocabulary and the Thorndike Test of Word Knowledge is the second highest in the table.

The correlation between the Stanford Binet Vocabulary and the Printed

[14] The highest of the correlations calculated for the Completion Beta was that with the Chapman Reading-Comprehension Test: $+.85 \pm .05$ for 38 cases

[15] The correlation between the Completion Beta and the Drawing Completion Test of the Pintner Non-Language is $+.47 \pm .12$ for 46 cases.

[16] The highest correlation with the combined time and error scores on the Gray Oral Reading Test, Set IV, was that of the Thorndike Word Knowledge: $+.69 \pm .08$ for 44 cases. The correlation between the Gray Oral Reading and the Stanford Binet Vocabulary was lower: $+.57 \pm .10$ for 49 cases.

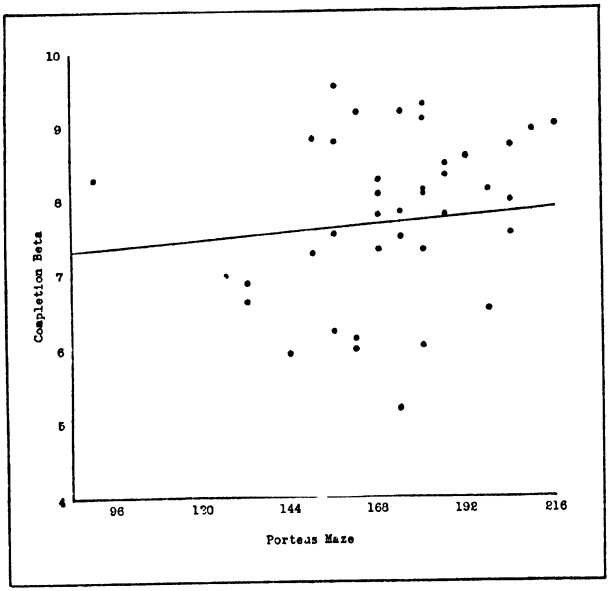

Figure 4. Correlation between the Kelley Trabue Completion Beta and the Porteus Maze Tests:
r = +.10; b = +.004

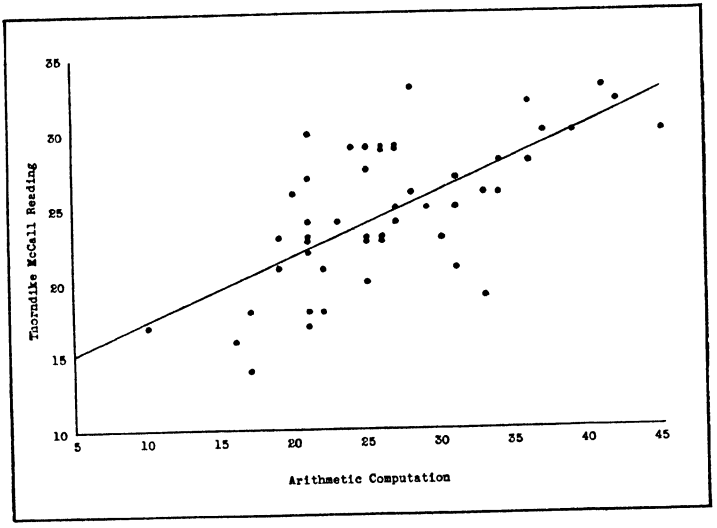

Figure 5. Correlation between the Thorndike McCall Reading Scale and the Arithmetic Computation Test:
r = +.66; b = +.43

Analogies Test is moderately high, but the diagram shows a wide scatter from the regression line. The diagram for the surprisingly low correlation of $+.31 \pm .13$ between the Thorndike Word Knowledge and this Analogies Test reveals a curious scatter, with a fairly close clustering of the scores around the regression line below a score of 50 on the Word Knowledge Test, and an increasing fanning of the scores above that (see Figure 6).

Correlations with the arithmetic tests are lower than with the first six language tests of the table, but they are reliable. The diagrams show fairly wide general scatters from the regression line, except in the case of that for the correlation of the Word Knowledge and Arithmetic Computation Tests where the relationship is relatively close at the lower end of the scales.

The diagrams for the remaining correlations with the vocabulary tests show no peculiarities.

d. *Correlations with the Thorndike McCall Reading Scale*

All twelve of the correlations calculated between the Thorndike McCall Scale and other language tests are reliable, the highest being the correlation of $+.83 \pm .04$ with the Completion Beta.[17] The scatter diagrams for the correlations of $+.66$ with each of the arithmetic tests are similar. That for the Arithmetic Computation is reproduced in Figure 5 to show how closely most of the cases cluster around the regression line; yet paragraph reading and arithmetic computation would hardly be said, on superficial analysis at least, to have a great deal in common.[18]

In the case of the other tests correlated with the Thorndike McCall the scatters are fairly wide and without pattern.

e. *Correlations with the Gates Oral Spelling and the Stanford Achievement Dictation Tests*

The correlation of $+.90 \pm .03$ between the Gates Oral Spelling and the Stanford Dictation is the highest inter-test correlation found. In general, the series of correlations with the two tests are similar, but in the case of the language tests, with the exception of the analogies, and in the case of

[17] A similar statement holds for the corresponding correlations for the Chapman Reading-Comprehension Test, but the correlations of this series were in general a few points higher than those with the Thorndike McCall.

[18] The correlation between the Thorndike McCall and the Arithmetic Computation Tests is not materially changed when age is partialed out ($+.67 \pm .08$), it is lower when education is partialed out ($+.46 \pm .12$).

the arithmetic tests, the relationships to the Stanford Dictation are closer than those to the Oral Spelling.

The scatter diagrams show nothing unusual, except in the case of that for the correlation between the Gates Spelling and the Pintner Non-Language, where the relationship is closer above a score of 20 on the former.

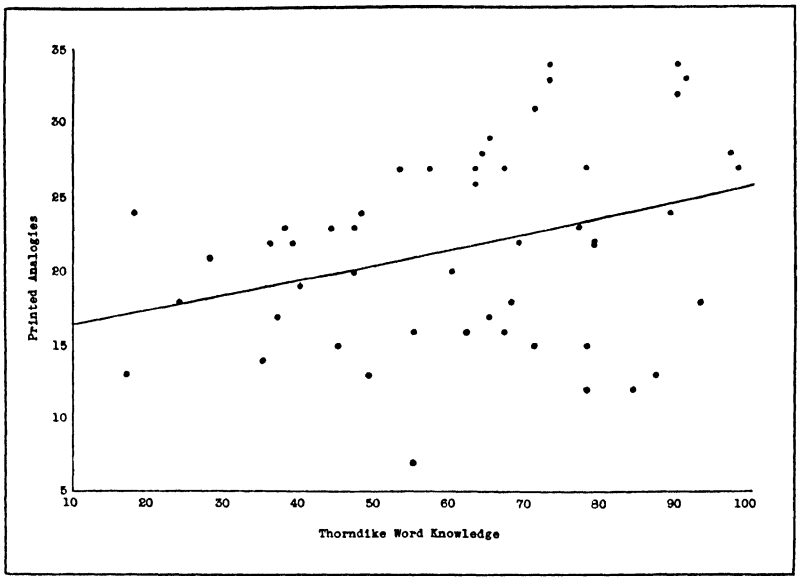

Figure 6. Correlation between the Printed Analogies Test and the Thorndike Test of Word Knowledge: r = +.31; b = +.10

f. *Correlations with the Printed Analogies Test*

The highest correlations found for the Printed Analogies are with the Stanford Binet Scale and the two arithmetic tests.[19]

There is a large difference between the degree of relationship shown for the Printed Analogies and the Stanford Binet Vocabulary and that shown for the Printed Analogies and the Thorndike Test of Word Knowledge, the former being 21 points higher than the latter. The scatter about the re-

[19] The highest correlations found for the Oral Analogies are with the Completion Beta, the Oral Absurdities, and the Stanford Binet Scale, but on the whole the two analogies tests, which correlate to the extent of +.83 ±.05, differ comparatively little in their relationships to other tests. The pairs of scatter diagrams, as well as the pairs of correlation coefficients, are similar.

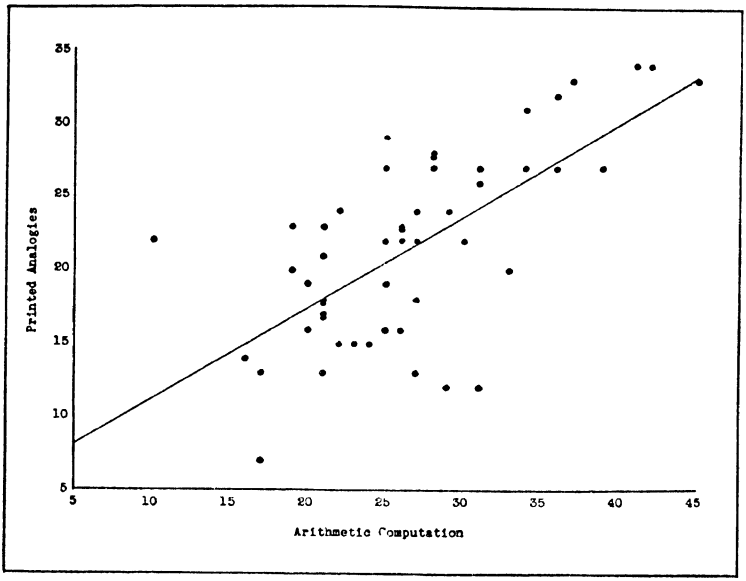

Figure 7. Correlation between the Printed Analogies and the Arithmetic Computation Tests: r = +.67; b = +.62

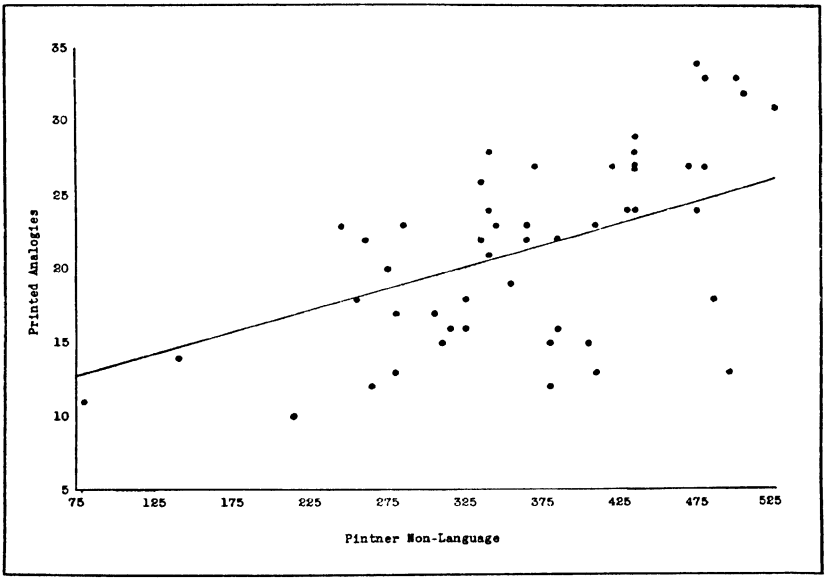

Figure 8. Correlation between the Printed Analogies and the Pintner Non-Language Tests: r = +.55; b = +.03

gression lines is wide in both instances, but it is narrower at the lower levels in the diagram for the correlation between the Printed Analogies and Word Knowledge Tests (Figure 6).

It is interesting to find that the Printed Analogies Test correlates fairly highly with the two arithmetic tests. The scatter diagrams both show closer

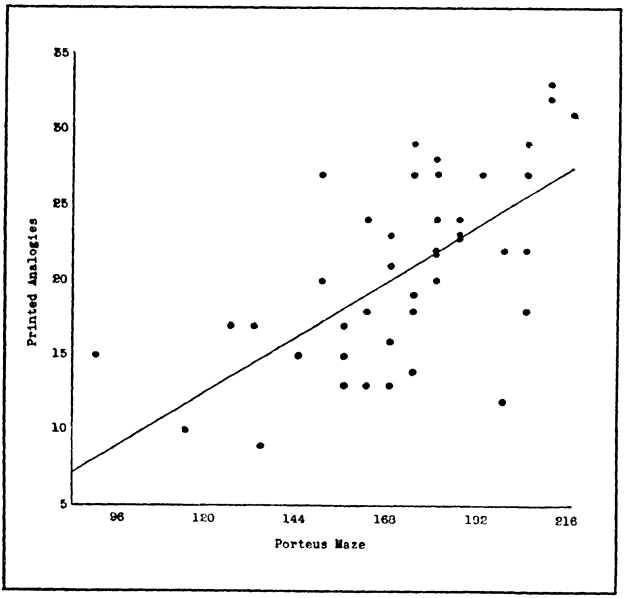

Figure 9. Correlation between the Printed Analogies and the Porteus Maze Tests: r = +.62; b = +.14

relationship above the means; that for the Printed Analogies and the Arithmetic Computation is reproduced in Figure 7.

The correlation of +.55 ±.10 with the Pintner Non-Language Test is higher than most of the other correlations with this test. The diagram, presented in Figure 8, shows an unusual triangular distribution of scores with the apex in the lower left corner.

The correlation between the Printed Analogies and the Porteus Maze is one of the most interesting in the table. With the exception of one slightly aberrant case, the scores cluster fairly closely around the regression line throughout the entire range, and the diagram, which is repro-

duced in Figure 9, resembles that for Printed Analogies and the Completion Beta. The correlation is higher than those with most of the language tests; and this fact together with the high correlations with the arithmetic tests and the Pintner Non-Language certainly suggests the presence of a non-verbal factor in this test.

TABLE XXX

CORRELATIONS WITH AGE AND INTER-TEST CORRELATIONS OF THE SCORES OF SUB-TESTS OF THE PINTNER NON-LANGUAGE (56 CASES)

	IMITATION	HARD LEARNING	DRAWING COMPLETION	REVERSED DRAWING	PICTURE RECONSTRUCTION
Correlations with age	−.52± 10	−.46±.11	−.42±.11	−.50±.10	−.28±.13
INTER-TEST CORRELATIONS					
Imitation		+ 67± 07	+.69± 07	+ 66± 08	+ 56±.09
Hard Learning			+.55±.09	+.52±.10	+.39±.12
Drawing Completion				+.57±.09	+.60±.09
Reversed Drawing					+.47±.10
INTER-TEST PARTIAL CORRELATIONS (AGE OUT)					
Imitation		+ 53± 10	+ 60± 09	+ 58±.09	+.52±.10
Hard Learning			+.54± 09	+.45±.11	+.38±.12
Drawing Completion				+.53± 10	+.57±.10
Reversed Drawing					+.48±.11

NOTE. Each correlation is followed by its standard error.

g. *Correlations with the Stanford Achievement Arithmetic Tests*

Inspection of the columns of correlation coefficients for the two arithmetic tests in Table XXIX shows that the Arithmetic Computation bears about the same degree of relationship to most of the other tests of the series, verbal as well as non-verbal, as the Arithmetic Reasoning. The only two exceptions are the correlations with the Stanford Binet and Pintner Non-Language Tests. With the Stanford Binet the Arithmetic Reasoning gives a higher correlation than the Computation, and with the Pintner Non-Language Test the situation is reversed. The scatter diagrams for the correlations between the two arithmetic tests and the Stanford Binet show similar distributions of the scores, but a closer clustering around the regression line in the case of the correlation with Arithmetic Reasoning. The scatter diagram for the correlation of +.65 ±.08 between the Arithmetic Com-

putation and the Pintner Non-Language Tests shows a wide spread around the regression line and it is evident that the relationship is not close.

h. *Correlations with the Pintner Non-Language Test*

In general the correlations with the Pintner Non-Language Test assume an intermediate position; they are, on one hand, lower than most of the correlations with language tests, including the analogies, or with the arithmetic tests; and, on the other hand, higher than the correlations with the remaining non-verbal tests. All but one of the correlations with the Pintner Non-Language Test are significant, the exception being that for the Thorndike Word Knowledge.

The two highest correlations in the group are with the Stanford Binet and Arithmetic Computation Tests; the next highest with the Porteus Maze,[20] the Completion Beta, and the Printed Analogies. The remaining correlations indicate only slight positive relationships between the variables; the scatter diagrams are all more or less diffuse.

Intercorrelations of the Pintner Non-Language sub-tests, with the omission of Sub-test 2, the Easy Learning, may be read from Table XXX.[21] They range from +.69 for Imitation and Drawing Completion to +.39 for Hard Learning and Picture Reconstruction. These correlations are much higher than those reported by Morganthau on children of ten to sixteen years.[22] When those for Sub-test 2 are omitted, the correlations obtained by this author range from +.04 for Reversed Drawing and Drawing Completion to +.46 for Reversed Drawing and Picture Reconstruction. It seemed possible that these correlations resulted in part from the larger age range among the adults. With age partialed out, however, the second order correlations, which are shown in the final section of Table XXX, are still higher than those reported by Morganthau.

i. *Correlations with the Goodenough Drawing Test*

The correlations with the Goodenough Drawing Test are almost all low. Only three were of a degree higher than might have been obtained by

[20] Morganthau found a correlation of +.36 between the Pintner Non-Language and the Porteus Maze. (*Op. cit.*, 45)

[21] The total score on the Pintner Non-Language Test, whenever given, includes the scores on all sub-tests. Sub-test 2 was not studied in relation to the other tests, however, because it gave a skewed distribution which showed that it was inadequate at the adult level.

[22] *Op. cit.*, 33.

chance: those with the Pintner Non-Language, the Stanford Binet, and the Thorndike McCall Reading Scale.[23] The diagrams are all similar, with a diffuse scatter around the regression line throughout the entire range.

j. *Correlations with the Porteus Maze Test*

From the point of view of the correlational analysis, the Porteus Maze Test was one of the most interesting. The relationships indicated by the coefficients of correlation are puzzling. Four of these coefficients are reliable: the highest is $+.62 \pm .10$ with the Printed Analogies, and the two closely following are $+.59 \pm .10$ with the Stanford Binet and $+.58 \pm .11$ with the Pintner Non-Language Test; the fourth reliable correlation is that of $.41 \pm .14$ with the Gates Oral Spelling Test. The very low correlation with the Completion Beta has already been noted (see Figure 4). There is a slight but unreliable negative correlation with the Thorndike Test of Word Knowledge.

[23] The correlation of $+.47 \pm .12$ between the Goodenough Drawing and the Drawing Completion Test of the Pintner Non-Language is also reliable

IX

SURVEY OF THE FINDINGS AND THEIR SIGNIFICANCE

1. Adult Performance Levels
2. Comparative Findings for Men and Women
3. Age and Test Performance
4. Relationships between Various Performance Levels in the Individual Case
5. Education and Test Performance
6. Occupation and Test Performance
7. Relationships between Various Test Performances

1. ADULT PERFORMANCE LEVELS

THE scores reported in Table XI are too few in number to establish standards for the various test performances, but they are probably good indications of the standards for the middle levels of the adult urban population. The facts which justify this statement are the characteristics of the sample in relation to data from the 1930 census report and from the results of the Army testing, and also the consistency shown in the findings themselves. As reported in Chapter II, there is a close correspondence between the Philadelphia population and the test group with regard to the proportions in different types of occupation. On the question of the educational and occupational levels of the test group, and indirectly of its intellectual status, the Army findings offer more direct evidence. The mean school grade reached by the males of the test group is 8.0 and that reached by the white draft in the Army 6.9. It has been noted that this difference probably reflects changing conditions in education rather than a definite superiority in the test group. In terms of occupational rating on the Barr Scale there is a very close correspondence: the mean for the males of the test group is 8.0 and that for the group of occupations for which the median Army Alpha scores was nearest the mean score for the total draft is 8.1.

Interesting evidence of the consistency of the results which has not yet been presented is to be found in comparisons of the mean scores for the three separate hospital groups. It is known that these differ slightly in the educational and occupational status of their ward patients and that a ranking on either of these criteria would place the Orthopedic Hospital first,

the Graduate second, and the Philadelphia General third. Consequently, it would be expected that, if the samples at the separate hospitals were adequate to give trustworthy results, the mean test scores would indicate the same ranking. Analysis shows that they do; mean scores for all but three

TABLE XXXI

MEAN SCORES ON THE MOST IMPORTANT TESTS FOR PATIENTS DISTRIBUTED BY HOSPITAL SOURCE

	PHILADELPHIA GENERAL HOSPITAL		GRADUATE HOSPITAL		ORTHOPEDIC HOSPITAL	
	N	*Score*	*N*	*Score*	*N*	*Score*
Stanford Binet Scale	20	157.8	13	154.5	32	171.0
Completion Beta	16	7.2	13	7.5	28	8.3
Stanford Binet Vocabulary	22	52.2	15	51.5	32	57.4
Thorndike Word Knowledge	14	56.4	12	57.4	28	66.3
Thorndike McCall Reading	17	23.8	12	24.8	30	25.1
Chapman Reading-Comprehension	15	15.3	11	15.6	16	18.0
Gray Oral Reading	12	80.3	11	78.1	26	68.9
Gates Oral Spelling	15	22.1	13	25.1	25	29.0
Stanford Dictation	18	82.4	14	84.3	26	99.6
Oral Analogies	18	18.7	13	22.0	10	22.0
Printed Analogies	20	19.4	13	20.1	27	22.7
Arithmetic Computation	16	24.6	11	26.4	28	27.5
Arithmetic Reasoning	15	21.5	11	22.4	28	23.7
Pintner Non-Language	18	327.1	11	368.4	27	371.1
Kuhlmann Immediate Recall	8	201.9	11	172.9	25	163.8
Goodenough Drawing	16	25.1	12	25.2	25	31.6
Porteus Maze	18	162.3	12	174.0	15	174.4

tests are highest at the Orthopedic, second highest at the Graduate, and lowest at the Philadelphia General. As shown in Table XXXI, two of these three exceptions are for the Stanford Binet Scale and the Stanford Binet Vocabulary, on each of which the Philadelphia General group slightly excels that at the Graduate, and the third is for the Oral Analogies, on which the groups at the Graduate and Orthopedic make the same mean scores. The remarkable consistency which this analysis reveals would not be surprising if the three groups were far apart, but they are not; the mean

scores for most tests are very close and yet the ranking according to the educational and occupational criteria holds. The findings offer evidence not only on the character of the test group, but also on the discriminative value of the tests in the study of adults.[1]

The findings for these adults have not been expressed in terms of mental age and should not be so expressed. It is apparently true that most of the average scores for adults from the middle levels of the population are numerically about the same as the standards established for the child of fourteen or thereabouts. This fact has been shown by other studies and is evident again in the findings in this research. The large majority of the average scores on the tests of this battery would be assigned to a level within six months above or below fourteen years; this group includes scores on the Stanford Binet Scale, the Stanford Binet Vocabulary, the Thorndike Word Knowledge, the Thorndike McCall Reading, the Chapman Reading-Comprehension, the Stanford Arithmetic Reasoning, speed on the Mare and Foal Test, and the Porteus Maze Test.[2] It is, of course, possible to say that average scores for adults on these tests are equivalent to the scores of the typical fourteen-year-old only if the standards for children and adolescents have been established on large and unselected groups. Quite aside from this point, however, the statement that the adult score is equivalent to the child's score at a certain age meets with further objections. In all probability there are differences between the adult and the child in the method of work or type of difficulty encountered in many specific performances, for example, in defining words or in solving arithmetic problems. It is fairly certain that the adult differs from the child in his attitude toward the test. All these factors are difficult to evaluate but they are probably of sufficient importance to lead to qualitative differences between the adult and the child in performances for which the scores are numerically equivalent.

In addition, it is evident that whenever a number of different types of performance are sampled, the average scores for adults spread over a wide range of mental and educational levels. The findings reveal this condition clearly; while the majority of the tests fall at levels just above or below fourteen, one, the Kelley Trabue Completion Beta, rates as high as fifteen years and four months,[3] and another, the Goodenough Drawing, as low as

[1] This evidence is important in view of the fact that it was impossible to make determinations of reliability for adults in the case of many of the tests.

[2] It is interesting to note that the average score on the Porteus Maze (14–1) would fall six months above that on the Stanford Binet Scale (13–7).

[3] It is true that a few of the subjects to whom the Stanford Binet and the majority of the

ten years. The Completion Beta is the only test showing an average score much above fourteen, but there are a number showing average scores which fall below thirteen, notably the Pintner Non-Language (12–6), the Kuhlmann Immediate Recall (between 12 and 13), the Substitution (12–0), the Arithmetic Computation (11–11), and the Seguin (11–0). Incidentally it is interesting to note that one of the tests in this low-mental-age group is the arithmetic computation, and the other tests all non-verbal. In short, there is definite evidence that scores for adults on a test which includes different types of performance or composite scores on a variety of tests can never be justly expressed in terms of mental or educational ages for children; furthermore, even when the comparison is limited to one particular performance for which standards have been determined on a large and unselected group of children, scores for the adult in terms of mental ages are still misleading because they imply a qualitative similarity in performances which are simply rated as equivalent numerically.

Another confusion which the mental-age method of expressing test results is likely to introduce is that between the average level for a group of adults of a wide age range and the age at which the highest average score is reached. Because the average scores for the adults of this group with ages up to sixty were equivalent to fifteen years and four months on the Completion Beta and twelve years and six months on the Pintner Non-Language Test, for example, it is not to be supposed that the peak of development is reached a little after fifteen on the former and between twelve and thirteen on the latter. The Completion Beta showed a correlation of only $-.19 \pm .13$ with age, the Pintner Non-Language a correlation of $-.53 \pm .10$. In general, the larger the number of older subjects in any group examined, the lower the average score and the mental-age equivalent on either test. The larger proportion of older subjects, however, would bring the greater reduction in the case of the Pintner Non-Language, though the peak of development on this test might come at fifteen, twenty, or even later.

A further point which must be noted is that the tests differ greatly in discriminative power for adults. In reporting the findings it was stated that certain of the tests were decidedly inadequate in this regard, notably the Morrison McCall Spelling Scale and a number of the non-language tests—the Easy Learning, the Two-Figure and the Casuist Formboards, the Ship

other tests were given had too great difficulty in reading to attempt the Completion Beta; but if these cases were added at the lower end of the distribution for the Completion Beta and a median calculated, the median would still fall at a mental age above 15. In short, the higher level on the Completion Beta is not to be explained on the basis of a selected group.

Test, and the Healy Picture Completion II. Even among the other tests of the battery there are differences in discriminative value. In many cases, of course, a piling up of scores at the upper end of the scale means simply that the test is too easy for adults. This situation is particularly obvious in the case of the performance tests. The problems they present are so easy that the scores for many adults must come close to the physiological speed limit for the activities involved. For other types of test different explanations for the negatively skewed distributions may well be necessary. It has been suggested that spelling is an activity which the majority of adults, including those of only moderate education and ability, have fairly well in command and that no process of increasing the difficulty of a test, short of introducing catch words, would yield anything but negatively skewed distributions. Spelling or any other activity which shows negatively skewed distributions may be studied at adult levels for a knowledge of the particular performance, but for a study of mental growth or individual differences between adults the more satisfactory indices are, of course, the tests with the higher discriminative value. The choice among these tests is a further problem, the solution of which will depend on the specific purposes of the study; in some cases it may be desirable to test performances like those tested by the Printed Analogies which show a definite decline in average scores after early maturity; in other cases it may be better to test performances such as vocabulary which are relatively well maintained through the adult years, and incidentally it may be remarked that there are probably many other performances of this type for which satisfactory tests have yet to be devised.

2. COMPARATIVE FINDINGS FOR MEN AND WOMEN

In studying sex differences the small size of the group was a disadvantage, for comparisons had to be made between all males and all females rather than between the groups at short age periods. The diversity of the tests on which data had been obtained was valuable, however, since it permitted comparisons on many performances for which sex differences are in dispute.

Analysis of the material not only fails to indicate male superiority on numerical or spatial tests and female superiority on verbal or "memory" tests, but shows no significant sex differences for any of the tests of the battery. In most cases mean scores for the two sexes are very close, but it is noteworthy that the men are slightly superior to the women on all but three of

the important tests. Comparisons of sigmas reveal no general differences in variability in the two groups.

In the relatively small size of the differences the results are on the whole in accord with recent work, which has indicated that sex differences are less marked at adult than at childhood and adolescent levels. Most of the studies of adults, however, have shown greater differences between men and women than were evident in this research.

3. AGE AND TEST PERFORMANCE

Two approaches to the important problem of age and test performance have been made: the comparison of mean test scores for different age groups and the correlations of the various tests with age. The findings indicate that differences in age from twenty to sixty have comparatively little effect on most test scores. For the small decade groups there are no significant differences between mean scores at any one-decade interval or at any two-decade interval, and the majority of the tests show only slight and insignificant correlations, usually negative, with age. It would not be adequate to conclude, however, that the adult at any age from twenty to sixty stood an equally good chance of making the same score on any of the tests of the battery. Certain tests show significant differences in mean score between the twenties and the fifties, and others show significant differences between the larger age groups twenty through thirty-four and forty-five through fifty-nine. In general, the evidences of an increase in score with age appear only for the vocabulary tests, notably the Thorndike Word Knowledge, while the evidences of a decrease with age are most marked in the case of the analogies and the non-language tests. The Thorndike Word Knowledge is the only test showing a reliable positive correlation with age; both analogies tests, the Mare and Foal, the Seguin, the Substitution, the Pintner Non-Language, and the Porteus Maze show reliable negative correlations with age.

That high points in development apparently occur at different ages and that the course with age varies in different test performances are conditions of great importance. The fact that scores on most of the language tests, and also on the arithmetic tests, are well maintained in the thirties suggests that development in these fields continues not only beyond fourteen, sixteen, or twenty, but that it may still be evident up to thirty or beyond. It is to be noted, however, that while the means for the small decade groups were

higher in the thirties than in the twenties in the case of nine of the eleven language or arithmetic tests, the peaks of development for the much larger groups studied by Willoughby and by Jones and Conrad almost all fell in the twenties or before.

The vocabulary, reading, spelling, and arithmetic tests, on which the scores are not only relatively high in the thirties but remain relatively high at later ages, may be said to involve activities fairly well within the experience of the older as well as the younger adult. It seems natural that vocabulary should be well maintained and should even improve with age. Interestingly enough, arithmetic computation, which apparently depends to a considerable degree on school training, also remains on a rather even level between the twenties and the fifties. There is obvious truth in the suggestion that these performances are all well maintained in adult years because they are within the natural realm of experience for the older person, but the possibility that they involve abilities which mature late or decline slowly must not be overlooked.

The early peaks followed by steady and rather rapid decline which appear for the analogies and non-language tests, must also depend on several factors. It has long been believed that motor skills requiring rapid movements were best accomplished by the younger person, and recent investigators have contributed more precise information on this point. Miles has shown that motor and perceptual abilities, especially those which depend on speed rather than diligence, manifest a steady decline after early maturity, although the declines are not so sharp as those for certain sensory capacities and memory and learning abilities.[4] The findings for the patients in this research on the Seguin, the Substitution, the Hard Learning, and the Mare and Foal may be considered as supporting the evidence of a steady drop in relatively simple activities where perceptual and motor abilities and speed are at a premium.

The findings, however, like those of Willoughby and Jones and Conrad, also show steady and rather marked declines in complex non-language tests, in which speed and skill in certain relatively simple motor and perceptual abilities are not so important, and also in the analogies tests.[5] The drop on

[4] W. R. Miles, *Age and Human Ability*, Psychol. Rev., 1933, 40, 112–116.

[5] The interpretation of the declines with age on some of these tests may be questioned on the ground that the older subjects are handicapped by the loss of certain secondary abilities, such as satisfactory understanding of directions or ready adaptation to unusual situations. Data on these points from the Stanford Maturity Study show that about the same negative correlation exists between age and score on an oral directions test (Price, *op. cit.*, 589) as be-

certain of the sub-tests of the Pintner Non-Language probably could not be explained on the basis of decreasing speed or facility in simple motor or perceptual operations. The decline on the analogies tests could not be attributed to any secondary factor, such as decreasing speed or visual efficiency or failure to adapt to a difficult situation. For none of these tests have the performances been sufficiently well analyzed to provide an explanation, although the possibility of a spatial factor immediately comes to mind. In this research the Printed Analogies repeatedly showed characteristics of some of the non-language tests, or close relationship to them, but it was not possible to determine whether these could well be accounted for by the assumption of a spatial factor. The problem is of considerable importance and one which requires further work on the characteristics of normal mental functioning before the many recent studies of pathological disturbances in "spatial thinking" can be rightly interpreted.

In a cross-section study of this sort it is, of course, impossible to determine whether the more intelligent man continues to develop longer than the less intelligent; the only statements which can be made are for the general tendencies of average or near-average individuals at different ages. On the whole, it is clear that the greatest development in "test intelligence" occurs before the age of twenty, and that the gains or losses which appear from that age to sixty are comparatively very slight. They do occur, however, and the losses in some activities, notably in the non-language field, are of sufficient magnitude to require the establishment of "normal" levels for different age groups in the adult years.

4. RELATIONSHIPS BETWEEN VARIOUS PERFORMANCE LEVELS IN THE INDIVIDUAL CASE

The extent of the test battery made it possible to study the relationships between a large number of performance levels in each individual case—a

tween age and score on the Otis Test, but that adaptability to new situations (Marsh, *op. cit.*, 589), as determined by score on a number of performance tests, shows only a slight tendency to decrease with age. Apparently, then, it is safe to assume that declines on non-language tests, such as were found in this research, depended on increasing age rather than on a progressively poorer adaptation among the older subjects. Certainly observations of the patients' reactions would support such a conclusion. In view of the considerable drop with age in the understanding of oral directions, however, it is apparent that age curves for tests requiring complex oral directions must be interpreted with caution, and particularly in the case of grade-school subjects, who showed a more marked decline than college subjects in the Stanford Study.

problem which has never been pursued far because of the difficulty in administering a long series of tests to each subject. When the scores for three separate test groups of 20 subjects each are expressed in standard measures, the means of the average deviations of these measures range between .44 and .50. There is little relationship between the inequalities in test performance in the individual case, as indicated by the average deviation, and mean test level, age, or education. On the whole, however, the younger subjects and those of a higher educational level tend to show slightly less variation. The small size of the differences is of great interest: it seems clear that if school training or any other factor tends to equalize performance levels in early maturity the effects persist in adult life; and that, at least within the limits of the tests studied, there is little tendency toward greater specialization with advancing years.

An analysis of the tests or groups of tests falling low or high in the individual cases gives negative results. Particular constellations of abilities, as indicated by high scores in certain groups of language or in non-language tests, are exceptional, and on the whole it is evident that adults who do well in one type of performance usually do well in the others, and vice versa.

5. EDUCATION AND TEST PERFORMANCE

The amount of education which the patients had received, as estimated by the school grade completed, shows a significant and moderately high degree of relationship to almost all the test performances studied; the only exceptions are a few of the non-verbal tests. Correlations with the so-called intelligence tests are almost as high as those with the educational achievement tests, that is, the tests of activities supposed to be directly school trained. This situation, together with the impressions which developed as studies were made of dull patients with considerable formal schooling and intelligent patients with little, seemed to throw some doubt on the assumption that education in itself raised the test performances.

The results of analyses by the partial correlation technique, showing the relationship between education and various test performances with Stanford Binet score partialed out and the relation between Stanford Binet and other test performances with education partialed out, are consistent and extremely interesting. For many performances the influence of test intelligence clearly outweighs that of education. This situation might be expected for the Thorndike Test of Word Knowledge, for vocabulary has usually

been regarded as one of the best indices of intelligence. Nevertheless, the extent to which success on this vocabulary test apparently reflects a man's intelligence rather than his education is surprising; only .8 per cent of the variation in scores results from the variation of education while 12.4 per cent results from the variation of test intelligence.

Among tests of specific school subjects, the reading comprehension and the arithmetic reasoning tests might also be expected to show a closer rela-

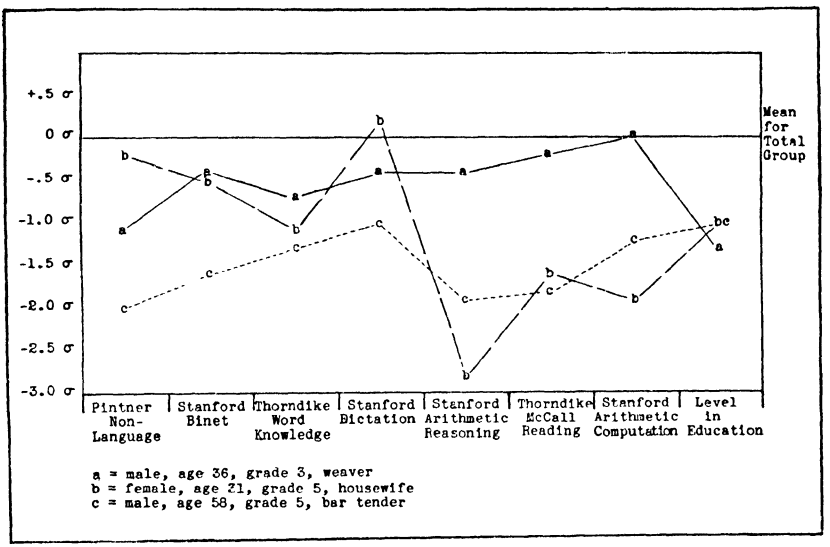

Figure 10. Relations between educational and various test levels. One case at third grade level and two at fifth

tion to intelligence than to education, for while they demand particular skills learned in school, they also demand complex mental functioning which characterizes the intelligent person and cannot be instilled into the dull. The results confirm this supposition, and show that the performances required by the reading comprehension and the arithmetic reasoning tests are indeed more closely related to test intelligence than to education.

The activities involved in the other educational achievements analyzed—oral reading, spelling, and arithmetic computation—would generally be considered more closely dependent on school training. The most striking fact of the analysis is that the spelling tests, the Gates Oral Spelling and the

Stanford Dictation, are more closely related to test intelligence than to education,[6] while the Gray Oral Reading shows little difference one way or the other, and only the Arithmetic Computation reveals a slightly closer relation to education than to intelligence. The evidence that the school grade reached is a comparatively unimportant factor for success in spelling

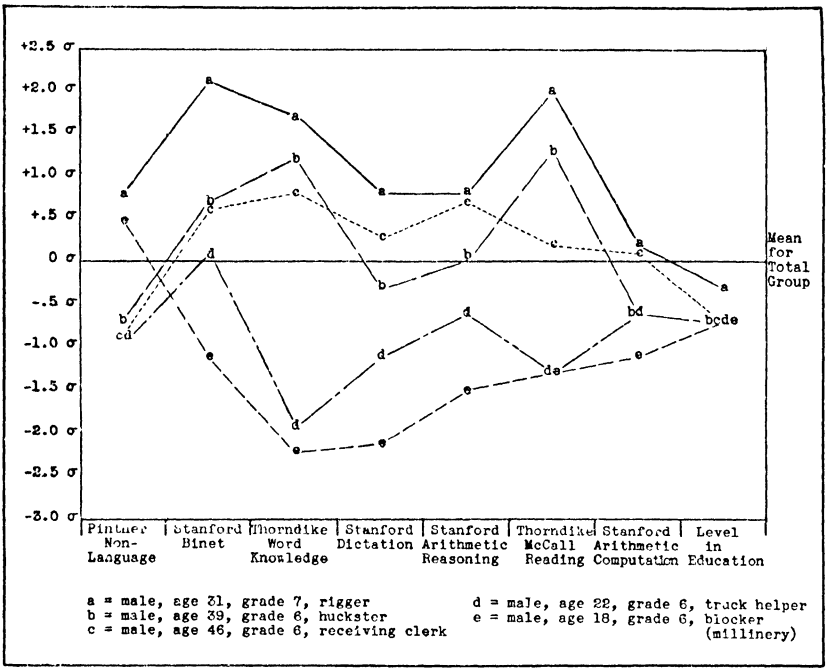

Figure 11. Relations between educational and various test levels. Four cases at sixth grade level and one at seventh

is clear: with Stanford Binet scores partialed out, correlations for the two spelling tests with education are reduced from +.59 and +.60 to +.20 and +.23; with education partialed out, correlations with Stanford Binet are reduced only from +.68 and +.74 to +.46 and +.57.

The correlational findings were extremely interesting, but the groups on which the partial correlations had been calculated were so small that it

[6] It is worth noting in this connection that neither spelling nor writing to dictation are component parts of the Stanford Binet Scale as vocabulary and reading comprehension are.

seemed worth while to make an analysis of the relation between education and test performance in the individual cases. Seven tests were selected as a basis for the study: the Pintner Non-Language, the Stanford Binet, the Thorndike Word Knowledge, the Stanford Dictation, the Thorndike McCall Reading, and the Stanford Arithmetic Computation and Reason-

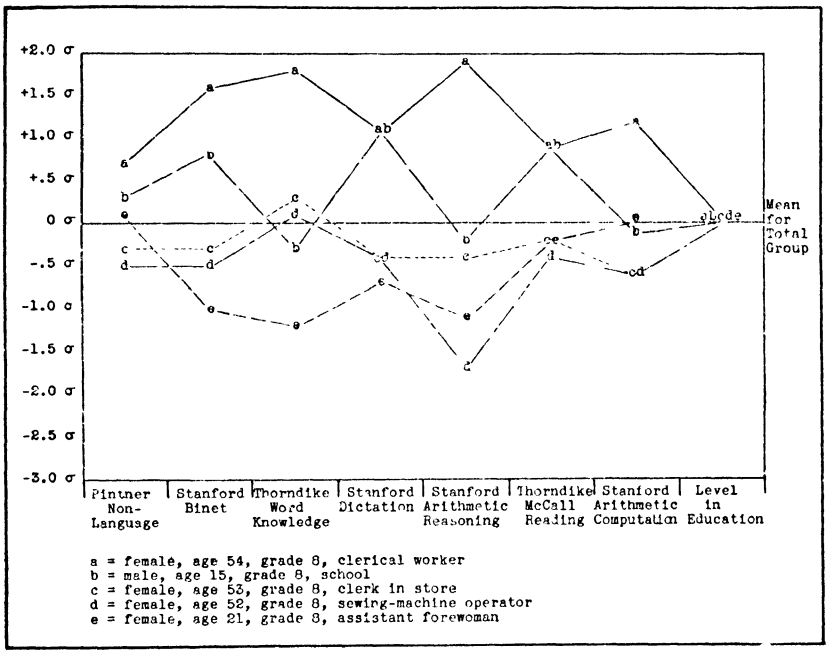

Figure 12. Relations between educational and various test levels. Five cases at eighth grade level

ing Tests. For each of the 35 subjects who had taken all of these, the sigma deviation of each score from the mean for the total group was determined and the individual case records were plotted on graphs such as those shown in Figures 10 to 14. Sixteen cases showed a difference of less than .3 sigma[7] between the deviation from the mean in educational level and in Stanford Binet score, and for simplicity in presentation these have been omitted from the final graphs. The remaining 19 subjects, with differences of .3

[7] This point was selected arbitrarily as the one which gave the most satisfactory distribution of cases for clarity in the graphs.

sigma or more between the levels in education and in Stanford Binet, are the interesting cases from the point of view of the relationship between education and test performance.[8] Nine of them made Stanford Binet scores .3 sigma or more above their respective educational levels, while the other 10 made scores .3 sigma or more below.

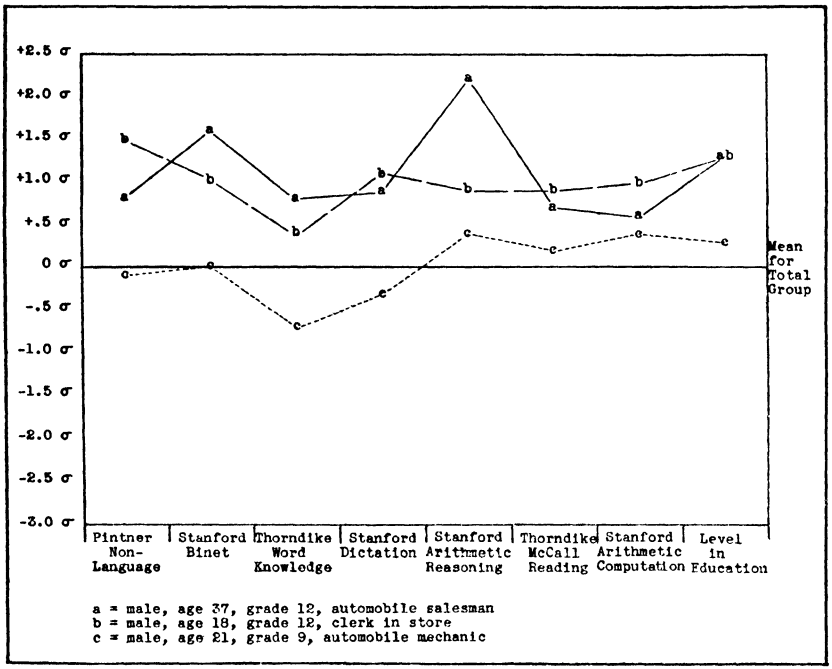

Figure 13. Relations between educational and various test levels. One case at ninth grade level and two at twelfth

A study of the graphs will show that the individual case lines follow the level set by the Stanford Binet score far more often than the level set by the school grade completed. Among the 9 cases with Stanford Binet scores which are high in relation to the educational levels, 5 or possibly 6 show test performances on the Thorndike Word Knowledge, the Stanford Dictation, the Arithmetic Reasoning, the Thorndike McCall Reading, and the

[8] The educational level is shown at the extreme right of each graph. Note that cases from more than one school grade appear in Figures 10, 11, 13, and 14.

Arithmetic Computation which are closer to the Stanford Binet than to the educational level.[9] These are cases *a* in Figure 10, *a*, *b*, and *c* in Figure 11, and *a* and possibly *b* in Figure 12. These subjects were apparently able to compensate for a relatively poorer education than their intelligence probably warranted. Cases *b* in Figure 10, *d* in Figure 11, and *a* in Figure

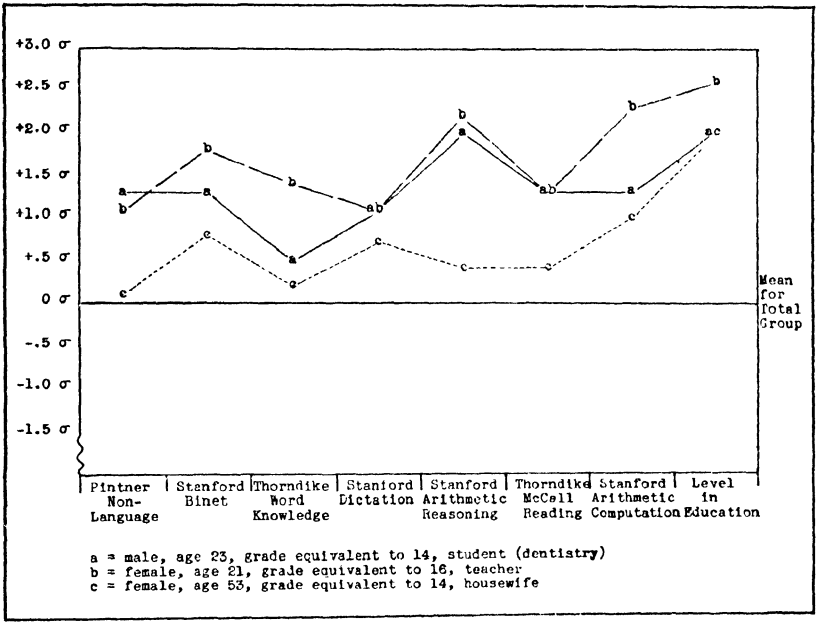

Figure 14. Relations between educational and various test levels. Two cases with equivalent of fourteen years' schooling and one with equivalent of sixteen

13, on the contrary, each with the Stanford Binet slightly higher than the educational level, not only failed to bring up other test scores to match the

[9] The Pintner Non-Language Test was included in the analysis to give an additional index of intelligence. Its correlation with education, +.56 ±.10, is almost as high as that of the Stanford Binet, +.62 ±.08. In 7 of the 19 cases the levels on the Binet and the Pintner are less than .5 sigma apart. Among the remaining cases there is one where the various tests fall nearer the Pintner than the Binet and 7 others where some of the tests fall nearer the Pintner and some nearer the Binet. To a certain degree, therefore, the level on the Pintner may possibly indicate abilities which account for the irregularity in cases where the various performances do not approximate the level of the Stanford Binet; but so many different factors are, of course, involved that it is difficult to estimate the effect of one or two in isolation.

Stanford Binet, but for the most part obtained scores relatively inferior to the educational level in terms of sigma deviation.

All 10 subjects in the group having a Stanford Binet score lower than the educational level showed performance levels on the vocabulary and educational achievement tests which were also as a general rule inferior to the educational level. These cases must, however, be considered in relation to the part of the scale of ability in which they fall. Cases *c* in Figure 10, *e* in Figure 11, *c, d,* and *e* in Figure 12, *b* and *c* in Figure 13, and *c* in Figure 14 are apparently subjects who were either unable to profit by the extent of the educational training they experienced or deteriorated after leaving school. These subjects stand in direct opposition to those of the first group who seem to have been able to compensate for a relatively poor education. Cases *a* and *b* in Figure 14, while they appear from the graphs to be subjects whose test performances fall considerably short of their educational levels, are not to be classed among those who failed to profit by school training or deteriorated after school. The situation in these two cases is more correctly understood as the result of inadequacies in the tests employed; both subjects came close to the upper limits on most of the tests and would undoubtedly have shown less discrepancy between sigma deviation in test performance and sigma deviation in education if the tests had been more difficult.[10]

It goes without saying that the many factors involved make the significance of the findings somewhat problematical. For a decisive conclusion the analyses would, of course, have to be carried out on larger groups in which the factors of age, type of occupation, general environment, and even marked differences in personality were controlled as far as possible.[11]

[10] The scores for these two subjects, one with fourteen and one with sixteen years' schooling, were as follows:

	Case a	*Case b*
Pintner Non-Language	500	480
Stanford Binet	196	210
Thorndike Word Knowledge	73	91
Stanford Dictation	115	114
Arithmetic Reasoning	34	35
Thorndike McCall Reading	30	30
Arithmetic Computation	37	45

It must be noted that the scores of subjects like these, for whom the tests were not adequate, would decrease the correlations between education and test performance. There were few such subjects, however, and for the group as a whole the tests were fairly satisfactory from the point of view of discriminative power, so that the correlations are probably little affected.

[11] There is little difference in type or level of occupation between the intelligent subjects with poor education and relatively high test performances and the duller subjects with good

Nevertheless the findings strongly suggest that those adults who had progressed further in school than their intelligence probably warranted had never acquired or had forgotten much of what they were taught in fundamental school subjects, while those who had not continued in school so far as their intelligence warranted yet had by some means acquired an amount of formal knowledge commensurate with their respective intelligence levels.

The tests used in this research are to a certain extent unsatisfactory for a complete study of the problem, not only because of their somewhat limited discriminative power but also because they do not cover a sufficiently wide range of performances. It may be said that the individuals who remain in school longer, regardless of intellectual level, acquire more extensive knowledge than those who leave early, for example, in the fields of history, science, or literature. It would be surprising, however, if those who remained in school and profited by these studies did not continue to improve in subjects tested in this research, notably in vocabulary and the understanding of reading. In other words, tests of these subjects are at least partial indices of educational development beyond the stage of the three R's, and the results indicate clearly that success on these tests has a much closer relation to intelligence than to education. There is, of course, still the possibility that those who remain in school longer, the dull as well as the bright, gain in subtler values which are not demonstrable by tests, but impressions gained from close work with the subjects did not relate such values to length of school training.

The trend of the results gives one pause, particularly in view of the sums spent to provide eight or more years of formal school training for each child. It seems clear, as educators and psychologists have long contended, that the adequate determination of the child's abilities is an essential part of any educational program, and that the nature and extent of the training should be adapted to these abilities. In this research there were clear examples of persons who were intellectually unable to profit by more than a certain minimum of the ordinary school curriculum, and yet had been

education but relatively poor test performances The subjects of the first class included a weaver, a rigger, a huckster, a receiving clerk, and a clerical worker, the of the second class a bartender, a hat blocker, two clerks, a sewing-machine operator, an assistant forewoman, and an automobile mechanic. In other words, within the limits of the group analyzed here, there is no indication that differences in test performances in adult years depend on the character of the daily occupation. Occupation, like education, is undoubtedly a factor and there are, of course, other factors as well, but in all probability none of them is so important as intelligence.

exposed to this sort of training for as much as eight years. They had not yet enjoyed and probably never would enjoy any but the simplest skills in reading, writing, and arithmetic, and there is little doubt that a good part of their school years should have been spent on other types of training, vocational or avocational. The individuals in the other class, who had received less formal schooling than they could have absorbed, may leave one content with a *laissez faire* doctrine, but it must be noted that these adults who did well in spite of a short school training would probably have done still better if they had been carried through an educational experience more nearly in accord with their intellectual ability.

6. OCCUPATION AND TEST PERFORMANCE

When the groups of the Taussig Scale are combined to form three—the first, professional, semi-professional, and business; the second, skilled labor; and the third, semi-skilled and unskilled labor—Chi-Square determinations show a definite association between success on three of the tests and the professional and business group. These three tests are all verbal: the Completion Beta, the Stanford Binet Vocabulary, and the Thorndike McCall Reading Scale. The contingency table for the Completion Beta is particularly interesting: only one case from the professional or business group made a score below the mean while 16 of the 20 cases of semi-skilled and unskilled labor made scores below the mean. None of the non-language tests was found to have any definite association with any of the Taussig groups. On the basis of the Chi-Square determinations, then, it is possible to say that members of the professional and business group are definitely superior to the other two groups in the abilities involved in the Completion Beta, Stanford Binet Vocabulary, and Thorndike McCall Reading Scale. They are not superior in the abilities sampled by the non-language tests, and at the same time it is noteworthy that the skilled laborers do not excel the other groups here.

The correlations between test scores and Barr Scale ratings are rather low in view of the fact that the ratings were assigned on the basis of the intelligence supposedly necessary to carry on the given occupation. The two highest of those calculated—the correlations with the Arithmetic Reasoning and the Thorndike Word Knowledge—were about +.60, while that for the Stanford Binet was only +.44, and those for the Pintner Non-Language and Porteus Maze were lower and unreliable.

7. RELATIONSHIPS BETWEEN VARIOUS TEST PERFORMANCES

Statements as to the relationship between various test performances must necessarily be based on the intercorrelations, for the comparatively small number of subjects taking all the more important tests precluded any further statistical analysis. It is impossible therefore to make any definite statements as to the mental functions involved or their relative prominence. The days have long gone by when each investigator could suit his own fancy in stating what his test tested, but it is to be noted that even now his statistical analysis of mental functions leaves him considerable latitude in naming them, and that there is still room for speculations on the basis of such material as is available in this study.

All but one of the test performances for which calculations were made were found to be positively correlated, the exception being the slight negative correlation between the Thorndike Test of Word Knowledge and the Porteus Maze. While some of the correlations with the more specific non-verbal tests were low and unreliable, the majority of the correlations were moderately high. The fairly close relationships thus indicated apparently depended rather on the functions involved than on the wide age range of the group, for when age was partialed out in a number of crucial cases the correlations remained about the same. When educational level was partialed out, however, some of them dropped about ten to twenty points.

The intercorrelations of the tests of the language group discussed in Chapter VIII are in general the highest of the lot, and are of sufficient magnitude, even if educational level is partialed out, to show clearly that there is a fairly close relationship between the functions tested by the Stanford Binet, the Completion Beta, the Stanford Binet Vocabulary, the Thorndike Word Knowledge, the Thorndike McCall Reading, the Gates Oral Spelling, and the Stanford Dictation Tests.[12] Among the intercorrelations in this group the highest appear between tests supposedly sampling the same type of activity, that is, between the two spelling tests and the two vocabulary tests.[13] On this point, at least, the correlations support the

[12] For simplicity in presentation the statements made in this and the following paragraphs are based on the most important tests of the battery. In general, however, they apply to all of the tests reported. For example, the Oral Absurdities shows the characteristics of the other language tests reported here from the point of view of the correlational findings. Similarly, the Seguin and the test of Drawing a Chair show characteristics like those of the other rather specific non-verbal tests, for example, the Goodenough Drawing.

[13] High correlations also appear between the Thorndike McCall Reading and the Chapman Reading-Comprehension, and between the Oral and Printed Analogies Tests.

logical deductions which one would make as to the nature of the abilities involved. In order of degree the next correlations are those between the reading comprehension tests and certain of the so-called language intelligence tests. These are of sufficient magnitude to indicate that the functions involved are definitely related. It seems clear that for the adults of this group there is little distinction between the language intelligence tests and the reading comprehension tests, either in their intercorrelations or in their correlations with other tests. While there is evidence of similarity in the function or functions involved in all these tests, the definition of these functions is, of course, another matter. The most obvious guess is that all these performances involve verbal ability and that the major part of the relationship is probably to be accounted for on this basis.

When the tests of the battery were first grouped the analogies were put with the "language intelligence" tests, for they obviously involved word knowledge and facility in the production of linguistic symbols under certain controlling conditions. All the evidence showed, however, that these tests stood between the verbal and the non-verbal. They bore moderately close relationships to the tests of the language group, although not so close as the interrelationships among these tests, and at the same time they were fairly closely related to the arithmetic and the non-verbal tests. They showed particularly low correlations with the Thorndike Word Knowledge, a fact which seems to indicate that the analogies test is not so much a test of vocabulary as has sometimes been supposed. What the functions are which set the analogies a little apart from the other language tests is, of course, not known; but it may possibly be, as already suggested, that they involve a spatial factor and that this is the explanation for their relationship to some of the non-verbal tests, notably the Porteus Maze.

A further consideration of the correlations between the tests in the language group and those outside it suggests that there are definite relationships between the functions tested by the language and by the arithmetic tests, and that these are closer in the case of the Arithmetic Reasoning Test. As to the Pintner Non-Language and the Porteus Maze, there are still definite relationships with the Stanford Binet but, in general, low and often insignificant relationships with the other tests of the language group. It is evident that these two tests involve functions considerably different from those tested by the language tests, and also by the arithmetic tests, and the irregularity in the series of correlations with the Pintner Non-Language

and with the Porteus Maze would suggest the presence of several different factors.

The Goodenough Drawing, as well as other non-verbal tests of the battery, apparently involves different and probably fairly specific factors, bearing little relationship to the language or the arithmetic tests. These non-verbal tests show certain relationships to the Pintner Non-Language and some of its sub-tests, and certain interrelationships among themselves, particularly in the case of tests of apparently similar processes such as the Hard Learning and the Substitution. On the whole, however, these tests clearly involve more specific functions than do the verbal, and this is probably the most that can be said about them at present, for it is impossible to make any immediate determinations of the functions actually involved.

X

CONCLUSIONS

THIS research has tapped a new and valuable source for the study of adult intelligence. Too little advantage has yet been taken of the opportunity the hospital offers for detailed psychological studies of adults from the middle levels of the population. Within the hospital group there are undoubtedly many types and degrees of abnormality in mental functioning, but there are also patients who may safely be said to be "normal" mentally. Among them there are cases requiring hospitalization for two weeks or longer, and such patients are practically unique among adult subjects in that their time is comparatively unlimited and most of them are willing to spend a large amount of it on the examinations. They may be given not one but many tests, with the result that a much more complete knowledge of the individual's mental functioning can be obtained than is usually the case. Furthermore, they may be given repeated examinations without fear that their interest will flag; when the studies are properly conducted, interest in the examinations as well as friendliness with the examiner usually increases steadily throughout the work. It is true that individual studies are costly, but they are rewarded by the certainty that the tests have served their purpose fully and that they have yielded not only a score value for the particular performance, but an opportunity for observing how that performance was carried through and what difficulties stood in its way.

The hospital group as a whole is the only single group which comprises all social and economic strata. The ward patients, who formed the subjects in this research, constitute a more limited sample, but a sample which has every indication of being truly representative of the middle levels of the population. It has been shown to have approximately the same occupational distribution as the population of Philadelphia; it seems to be close to the average of the Army population with regard to the grade of intelligence required by the occupations represented; and it shows a distribution for education similar to the Army distribution.

The correlation between intelligence and the predisposition to disease or environmental conditions favorable for disease is still an unknown quantity. If it is negative the consequence would naturally be that hospital patients, because of the disease conditions, tended toward inferiority in intelligence.

It is probable that such a tendency, if it exists, is more marked in certain diseases than in others, and notably in those diseases in which neglect or poor living conditions are etiological factors. Again the problem resolves itself into proper choice among the hospital patients; and, granted this, it seems clear that any slight selective factors which may exist because of a relationship between intelligence and disease are of little weight in comparison to the advantage the hospital group offers as a random sample of the middle levels of the adult population.

Because the purpose of this research was a study of normal adults as a basis for work with aphasic and non-aphasic cases of cerebral lesion, every effort was made to select only patients who would be generally accepted as mentally normal. Many were rejected whose performances on the tests might well have been normal, for example, the diabetics, but with so little knowledge of the characteristics of mental functioning in various diseases, risks had to be cut to a minimum. It goes without saying that there is a large field for work on the problem of intelligence and disease, and that beginnings should be made in accumulating cases which are clear-cut from the diagnostic point of view.

When one considers the problems of adult intelligence and the various studies contributing to the knowledge on the subject, it seems clear that the average level of so-called general intelligence and the characteristics of its course with age have been fairly well established. Within recent years there have also been reports on specific types of test performances, but many of the activities which are important for a knowledge of mental functioning in the adult years have as yet been little studied. In this research a certain number of these activities have been investigated. The groups examined were comparatively small so that the actual norms obtained are no more than tentative. It is believed, however, that they are highly trustworthy: the individual nature of the work ensured maximum effort from all ages and types of subject, the group is apparently typical of the middle levels of the population, and the findings themselves are consistent.

With regard to the particular tests of the battery and their use with adults, the most important conclusions may be listed as follows:

1. The sentence completion test is undoubtedly one of the most important types of so-called language test for adults. The one used in this research, the Kelley Trabue Completion Exercise Beta, proved satisfactory from every point of view and would probably be difficult enough for a group slightly superior to this.

2. Vocabulary tests are well suited for work with adult subjects and readily accepted by them. In use with the normal subjects in this research there was little to choose between the two vocabulary tests from the point of view of their discrimination within the group or their relation to other tests. The Stanford Binet Vocabulary, however, naturally had the advantage over the Thorndike Test of Word Knowledge in providing an opportunity for study of the patient's reactions and attitudes.

3. The mixed analogies test in the forms arranged in this research differentiates fairly well among "average" adults, is readily accepted by them, and because of its apparent relationships with the non-verbal tests and its marked decline with age is one of the most interesting theoretically.

4. The absurdities test is also well adapted for use with adults, but is not satisfactory in the forms employed with these subjects and needs further development.

5. Many of the educational achievement tests employed were satisfactory in their existing forms. Both tests of reading comprehension, the Thorndike McCall and the Chapman, were adequate from the point of view of their discriminative power, but the former suffered a little in use with adults because of too childish content. Both the arithmetic tests, the Stanford Achievement Computation and the Reasoning, were found to be very satisfactory in discriminating among these adults. The Gray Oral Reading Test, the Gates Oral Spelling, and the Stanford Dictation each showed a piling up of scores at the upper end of the scale, probably not so much because the tests were too easy as because the activities involved are fairly well acquired by adult subjects, even by those of only moderate education and ability. The Morrison McCall Spelling Scale was obviously too simple for adults from the middle levels of the population, but it is well adapted for work with low-grade or pathological cases.

6. Many of the non-verbal or performance tests at present available fail to discriminate among normal adults. Most of the tests of the Pintner Paterson Performance Scale fall in this class. The data on these tests will not be useful in connection with further studies of normal adults, but have a definite value in studies of deteriorated adults whose level is low enough to require comparatively simple tests and whose deterioration may be estimated roughly by reference to the median and quartiles for normal adults.[1]

7. Other non-verbal tests, notably those of the Pintner Non-Language,

[1] Weisenburg and McBride, *op. cit.*, 577–578.

are more satisfactory in differentiating among "average" adults and contribute data on abilities which are important at adult as at childhood levels. There are not enough of these tests, however, or rather they do not cover all the types of performance which should be studied, for example, form or picture analogies.

8. The Goodenough Drawing Test presented an assignment unwelcome to many adults, and was undoubtedly a different type of problem for them than for the child, with success on the test much less closely related to intelligence than among children. Experience with the test of Drawing a Chair indicated that this too was a difficult problem, and suggested that studies of drawing in adults run up against very variable factors, emotional as well as intellectual, and must be interpreted with caution.

The analysis of the results for their bearing on the question of age and test performance is one of the most interesting sections of the work. The first conclusion is that by far the greatest extent of mental development as indicated by these test performances has occurred before the twenties, and that from this decade through the fifties there is little further gain and comparatively little decline. While the magnitude of the development before twenty far outweighs any change between twenty and sixty, smaller changes do occur in the adult period. As they appear in specific types of mental activities, these changes are comparatively little known, but are none the less important, for example, in estimating the relative value of the worker in a given occupation at twenty and at fifty, or the value of educational projects for the "average man" in earlier or later decades. The results of this study show clearly that the peak of development and the course with age are different for different performances. A survey of the findings suggests two conclusions which may be briefly stated as follows:

1. The abilities sampled by most of the so-called language tests are well maintained through the thirties and in the majority of cases show only slight declines thereafter. The abilities sampled by the vocabulary tests show little or no decline through the fifties.

2. The abilities sampled by another group of tests show an early peak, probably falling either in the twenties or before, with successive declines in the thirties, forties, and fifties. These declines are not significant at any one-decade or two-decade interval, but most of them are significant for the interval between twenty and fifty or for the larger age groups twenty

through thirty-four and forty-five through fifty-nine. A point of great interest is that all but one of these tests are of the non-language type, while the tests which do not show a drop in mean score before the thirties are language or arithmetic tests. The only "language" test in the group characterized by steady declines with age is the Printed Analogies, which repeatedly manifested a close relationship to the non-language tests. Declines with age on some of the non-language tests probably reflected the generally accepted declines in simpler motor and perceptual abilities and particularly in those in which speed was a dominant factor. Declines in others, notably the Porteus Maze and the more complex tests of the Pintner Non-Language Scale, and also in the Printed Analogies, could not be explained on this basis. Furthermore, they were apparently not to be attributed to a greater difficulty among the older subjects in adapting to the test situations, for such difficulty was not observed in the situation itself. It is suggested rather that these tests probably involve spatial factors and that the declines are to be understood on this basis.

From the point of view of the relation of education and test performance, the findings indicated that the amount of formal schooling was probably not such an important factor as it is generally supposed to be. An analysis of individual cases revealed that the intelligent men with poor school training had effected some compensation by adult life, while the dull individuals who had received more training in the fundamental school subjects than their intelligence warranted either had never profited by it or had forgotten most of what they once knew. It would not be wise to lay too great stress on this point in view of the small size of the group, and also in view of the fact that at least one performance, arithmetic computation, showed a slightly closer relation to education than to test intelligence. It is important to note, however, that scores on all the other "educational achievement" tests analyzed bore a closer relation to the test intelligence than to the school grade completed, and that this relationship held even for the spelling tests, the skills involved in which are definitely school trained. The results furnish additional proof, though such is hardly necessary, of the wisdom which lies in studying the mental abilities of the school child and making the nature and extent of his training fit his potentialities for development.

A study of test performance in relation to the occupational classes of the Taussig Scale indicated a condition which is easily accepted on the basis of

common observation, namely that the professional and business groups excel the others in the abilities sampled by the Sentence Completion, the Stanford Binet Vocabulary, and the Thorndike McCall Reading Scale. It is noteworthy that these tests are all verbal, and that none of the non-verbal tests showed a relationship to any of the groups of the Taussig Scale. It would seem that the abilities involved in these non-language tests either are not those in which the skilled laborer excels the professional or business man, or that the abilities shown by the skilled laborer in different trades are in themselves so diverse that group analysis obscures the general trend.

Survey of the results led to the conclusion that sex differences on the test performances studied were negligible. No significant difference was found for any test performance, and the slight inequalities appearing for a few probably resulted from fluctuations of sampling.

Analysis of the relationship between various performance levels in the individual case showed that adults who did well with some test performances usually did well with others, while those who had difficulty with some also had difficulty with others. This condition represents only the general trend, however, and there were large variations in some cases. Within the limits of the ten language and non-language tests studied, these variations revealed only a few cases of marked constellations in mental abilities, with some groups of tests falling at very different levels from others.

There is no doubt that one of the most important problems in the study of adult intelligence, and indeed in all studies of intelligence, is the analysis of mental abilities and their relationship. With groups as small as those in this research, statistical analyses such as those developed by Spearman, Kelley, or Thurstone promised little, and have not been carried through, so that it is impossible to arrive at any conclusions on this basis as to the particular mental abilities in the adult or their interrelationship. Without throwing caution to the winds entirely, however, it is possible to discuss certain interesting leads which appear from the correlational and other data for the normal group and from the findings for the abnormal subjects examined, notably the cases of cerebral lesion with and without aphasia.

The study of the normal shows positive correlations between almost all tests, but much higher correlations in general between the so-called language tests than between language and non-language or within the non-language group. Naturally rigid lines of division do not hold: one striking instance of a crossing of the line is the Printed Analogies Test, which ob-

viously involved verbal material, but is apparently as closely related to the arithmetic and non-verbal as to the verbal tests. The Arithmetic Computation Test also occupies a sort of mid-way position.

The evidence for a distinction between different groups of mental abilities gains additional support from the findings for the pathological groups studied by the same tests.[2] Cases of right-sided cerebral lesion without aphasia or history of it were found to resemble the normal most closely in sentence dictation, oral spelling, and oral vocabulary, and to be significantly inferior to the normal only on the arithmetic and some of the non-language tests. These non-aphasic cerebral cases represent a sample inferior to the normal, but this condition does not explain the marked inequalities in their performance levels. Language abilities are apparently well maintained while the abilities required by the arithmetic and most non-language tests show evidences of deterioration.

The distinction between language and non-language is more pronounced in typical cases of aphasia; and here the opposite condition is to be found, that is, the language activities are more seriously affected than the non-language. Interestingly enough, the arithmetic activities usually stand in between. Some few patients whose language processes are extremely limited or confused so that they speak and understand only a few words correctly nevertheless do better than the average of the normal group on non-language tests, and may also do moderately well in arithmetic. Others, the great majority, show some deterioration in non-language activities and in arithmetic, but far less than in language. Only a few aphasic patients have anything like as great difficulty in non-language as in language work. In short, aphasia is a deterioration which usually extends beyond language processes and involves activities which do not require overt verbal responses, but the changes in non-language performances are less marked than those in language and in some cases the differentiation is clear enough for one to say that the aphasic patient has lost one form of intelligent response, the verbal, and retained other forms which do not require language.

Uncertain as reasoning from the pathological to the normal must be, it seems probable that the characteristics of cases of cerebral lesion, with and without aphasia, give certain new cues for an understanding of normal mental functioning. Despite the positive correlations found for the normal group between almost all the many activities studied and despite the fact

[2] *Ibid.*, chaps. XVI and XVIII.

that analyses of individual cases revealed few marked constellations of mental test performances, there are in all probability more or less independent groups of mental abilities. The typical normal adult does not show great differences in the development of these different abilities; in so far as performances on tests of intelligence and educational achievement may be taken as indices, the various abilities all fall at fairly similar levels. In cases of brain disease, however, or at least in cases of localized brain disease, they may be affected unequally. As to the nature of these groups of mental abilities and their interrelationships, there is still little precise knowledge and it may well be that groups as yet undefined will have to be postulated to explain results such as those found in this research. For the present it seems clear from this study and others, however, that there is probably a verbal ability of a rather extensive nature. Numerical and spatial abilities have also been indicated, particularly in Kelley's work, and the necessity for some such groups as these is suggested by the findings for both normal and pathological subjects.

APPENDIX

DESCRIPTION OF THE TESTS

1. Language Intelligence Tests
   a. Controlled association tests
   b. Kelley Trabue Completion Exercise Beta
   c. Vocabulary tests
   d. Stanford Binet Scale
   e. Miscellaneous tests
2. Educational Achievement Tests
   a. Reading tests
   b. Spelling tests
   c. Writing a Letter
   d. Arithmetic tests
3. Non-Language Tests
   a. Tests from the Pintner Paterson Performance Scale
   b. Pintner Non-Language Mental Test
   c. Cancellation
   d. Immediate Recall: Kuhlmann Binet
   e. Drawing a Chair
   f. Goodenough Drawing
   g. Porteus Maze Test

1. LANGUAGE INTELLIGENCE TESTS

a. *Controlled association tests*

OPPOSITES TESTS. The Opposites Tests used were chosen for the aphasia study and were too easy for normal adults.[1]

MIXED ANALOGIES TESTS. Two forms of the Van Wagenen Graded Analogies Tests were selected for the Oral and Printed Analogies Tests, B for the Oral and A for the Printed.[2]

The instructions were approximately the same as those outlined by Van Wagenen, except that the subject was asked to give his response orally.

I am going to say three words. I want you to tell me a fourth word that has

[1] For lists of stimulus words and procedure see Weisenburg and McBride, *op. cit.*, 581–582.

[2] M. J. Van Wagenen, *Graded Opposites and Analogies Tests,* J. Educ. Psychol., 1920, 11, 241–263.

the same relation to the third word I say as the second has to the first. Here's an example: *color* is to *red* as *name* is to —— *John*. Here's another: *page* is to *book* as *handle* is to —— *knife*. See if you can tell me the fourth word for this one: *chew* is to *teeth* as *smell* is to ——. (If the subject could not tell the fourth word, it was told him, and the directions were given again.)

The successive items of the Oral Test were then read aloud, and the subject gave his responses orally. If both Oral and Printed Tests were used, the latter was given second and with no further instructions except that the subject was asked to read the words instead of listening to them. He responded orally. If the Printed Analogies were used alone, printed directions were given and for convenience these were placed at the top of the test sheet.

Both tests were timed, but the score was taken as the total number of correct responses.

As given, each test included 50 items[3] but some of these were of doubtful value, either because subjects sometimes misunderstood them or because of difficulty in assigning a plus or minus score. When doubtful items had been eliminated the Oral Test contained 36 analogies, the Printed, 40. Analysis of the difficulty value of each item showed some overlapping, that is, two or more items falling at the same difficulty value. One item of such a group was cut from the Oral Test, and five from five different groups from the Printed. Each scale then contained 35 items. The two lists, each ranked in order of difficulty, are given below, together with the accepted response or responses to each analogy.

Oral Analogies Test

1. horn is to blow as bell is to —— (ring, clamor)
2. rain is to summer as snow is to —— (winter)
3. vinegar is to sour as sugar is to —— (sweet)
4. work is to day as sleep is to —— (night)
5. baker is to bread as bees are to —— (honey)
6. rug is to floor as pictures are to —— (wall)
7. iron is to heavy as aluminum is to —— (light)
8. year is to month as week is to —— (day)
9. high is to low as near is to —— (far)
10. stove is to heat as lamp is to —— (light)
11. raise is to lower as open is to —— (shut, close)
12. eat is to food as wear is to —— (clothes)
13. country is to road as city is to —— (street)
14. under is to over as down is to —— (up)
15. navy is to sailors as army is to —— (soldiers)
16. foot is to leg as hand is to —— (arm)

[3] One change on the Oral Form was made before the test was given: Item "blade : knife :: tine : ——" was changed to "knife : blade :: fork : ——" so as to simplify the vocabulary problem. "Prong" and "points" were accepted as satisfactory answers.

17. sweet is to taste as red is to —— (color)
18. silver is to tarnish as iron is to —— (rust)
19. front is to back as top is to —— (bottom)
20. hard is to soft as rough is to —— (smooth)
21. triangle is to three as square is to —— (four)
22. fair is to cloudy as sunshine is to —— (rain)
23. fruit is to basket as water is to —— (pail, bucket, bottle, pitcher)
24. box is to wood as bottle is to —— (glass)
25. wall is to paper as floor is to —— (carpet, rug, paint, varnish, stain)
26. coal is to black as gold is to —— (yellow)
27. foot is to ankle as hand is to —— (wrist)
28. man is to legs as carriage is to —— (wheels)
29. leg is to knee as arm is to —— (elbow)
30. kettle is to utensil as chair is to —— (furniture)
31. park is to gate as house is to —— (door)
32. knife is to blade as fork is to —— (tine, prong, points)
33. number is to figures as word is to —— (letters)
34. grain is to wheat as fruit is to —— (any kind of fruit)
35. house is to door as field is to —— (gate)

Printed Analogies Test

Write a fourth word that has the same relation to the third word in the row that the second word has to the first word in the row:

color red name JOHN
page book handle KNIFE
chew teeth smell ——

1. birds fly fish —— (swim)
2. summer hot winter —— (cold)
3. light day dark —— (night)
4. sun shines wind —— (blows)
5. father son mother —— (daughter)
6. hat head shoe —— (foot, feet)
7. barn hay library —— (books)
8. cup saucer knife —— (fork)
9. cats scratch bees —— (sting)
10. food eat books —— (read)
11. hair black eyes —— (blue, brown, black, etc.)
12. man husband woman —— (wife)
13. thermometer temperature clock —— (time)
14. clothes tear dishes —— (break)
15. water drink air —— (breathe)
16. lamp oil stove —— (coal, fuel, wood)
17. happy laugh sad —— (cry, weep)
18. potato vegetable veal —— (meat)

19. far near there —— (here)
20. sparrow bird mosquito —— (insect, bug)
21. bread flour candy —— (sugar, molasses)
22. city mayor state —— (governor)
23. water glass coffee —— (cup)
24. sight blind hearing —— (deaf)
25. July month Friday —— (day)
26. slipper shoe cap —— (hat)
27. cheap many costly —— (few)
28. air birds water —— (fish)
29. complex difficult simple —— (easy)
30. month week day —— (hour)
31. present known future —— (unknown)
32. mail write telephone —— (talk)
33. picture frame field —— (fence, wall)
34. victory defeat success —— (failure)
35. work problems play —— (games)

b. *Kelley Trabue Completion Exercise Beta*

The Kelley Trabue Completion Exercise Beta was given and scored according to Kelley's revision.[4] It was suitable for almost all the subjects of the normal group, with the exception of a few who had great difficulty in reading, and for these it was not used.

c. *Vocabulary tests*

STANFORD BINET VOCABULARY TEST. This test was given and scored in the regular way.[5] No evidence that subjects had had experience with this test before was ever found, although it has been published in popular magazines and newspapers.

THORNDIKE TEST OF WORD KNOWLEDGE. This test, a multiple choice type of vocabulary test, was given and scored according to Thorndike's directions.[6] It was not used in the case of those few subjects who had great difficulty in reading.

d. *Stanford Binet Scale*

The Stanford Binet Scale was given and scored according to Terman's direc-

[4] Kelley, *Individual Testing with Completion Test Exercises,* Teach. Coll. Rec., 1917, 18, 371–382.

When the study was begun the Completion Exercises could be obtained from the Teachers College Bureau of Publications. When they went out of print shortly afterward, the tests were multigraphed, but they may be had from C. H. Stoelting and Company, Chicago.

[5] Terman, *The Measurement of Intelligence,* 1916, 224–231.

[6] Published by Bureau of Publications, Teachers College, Columbia University. A sheet of directions for giving and scoring the test comes with the test blanks.

tions.[7] A complete test was always made but the Code Test, Year XVI, as noted in Chapter IV, was usually omitted because it was not well received by the adult subjects. All tests were given from the X-year level through the level at which no test was passed; in cases of one or more failures at X years, the earlier tests were given down to the age at which all tests were passed. In no case did the testing have to be extended below the V-year level.

e. *Miscellaneous tests*

ABSURDITIES TESTS. These were composed of absurd statements, chiefly from among those suggested by Ballard.[8] They were divided into two groups by selecting alternate items; one group was presented orally and the other in print. After a few trials the numbers in each group were cut to 12, with the elimination of those items which caused difficulty. When all the tests had been made, the responses were scored by the three psychologists working together. Two more items were cut, one from each series; both of them were often misleading to good as well as poor subjects.

The following directions were used in giving the Oral Test:

> I am going to read you a paragraph that has something foolish about it, something absurd. When I've finished, I want you to tell me what is foolish about it. What is foolish about this sentence? "A soldier in the march complained that every man was out of step except himself." (If the subject could not indicate the absurdity, it was explained to him.)

If the Printed Test were given later no further directions were added. For convenience in cases where the Printed Test alone was used, similar directions were added at the top of the printed page.

Oral Absurdities Test

1. A man asked a boy where Mr. Smith lived. He said: "The first house you come to is a barn and the next is a haystack—the next is Mr. Smith's!"
2. A gentleman fell from his carriage and broke his neck, but received no further damage.
3. The judge said to the prisoner: "You are to be hanged, and I hope this will be a warning to you."
4. When there is a collision the last car of the train is usually damaged most, so the guard thinks it would be best if the last car were always taken off before the train starts.
5. A boy wrote on his composition: "Soap smells nice, but it tastes horrid. It tastes worst of all when you get it in your eye."

[7] Terman, *op. cit.*, 189–348.

[8] P. B. Ballard, *The Limit of the Growth of Intelligence*, Brit. J. Psychol., Gen. Sec., 1921, 12, 125–141.

Ballard quotes one of Whipple's and one of Yerkes' tests, both of which were used.

6. An old gentleman complained that he could no longer walk around the park as he used to; he could now go only halfway round and back again.
7. A householder saw an advertisement: "Buy one of Simkin's stoves and save half your coal." He bought two in order to save all of it.
8. I am not conceited, but I don't think I'm half as clever as I really am.
9. You are thin and I am thin, but he is thinner than both of us put together.
10. A showman advertised for a giant and a dwarf. A man of ordinary height presented himself and offered to fill both parts. He claimed to be the smallest giant in the world and the biggest dwarf.
11. Every rule, even this one, has an exception.

Printed Absurdities Test

Read this sentence and decide what is foolish about it.

A soldier in the march complained that every man was out of step except himself.

What is foolish about each of these paragraphs?

1. I like end slices of bread. I gave the girl a whole loaf of bread and told her to bring me the two end slices. I afterward found that she had sliced the entire loaf. I asked her why she did this. She said: "How could I get the second end piece unless I did?"
2. I received a letter from a friend in which he said: "If you don't get this letter, just let me know and I'll write again."
3. I read in the paper that they fired two shots at a man. The first shot killed him but the second one didn't.
4. This morning I met a smart young man. He was walking down the street with his hands in pockets and twirling a brand new walking stick.
5. An Irishman called one day at the post office and asked if there was a letter waiting for him. "What is your name?" asked the postmaster. "Sure," said the man, "you will find my name on the envelope."
6. It is said that a certain town in Greece contains two relics of St. Paul; one his skull when he was a boy and the other his skull when he was a man.
7. A teacher said to his boys: "Tomorrow we will have an examination, attendance at which is voluntary; so if any boy is absent he'd better look out."
8. A man said to his shoemaker: "You blockhead! I told you to make one of the shoes larger than the other, and instead of that you have made one of them smaller than the other."
9. There is a tree in America so tall that it takes two men and a boy to see the top.
10. The three men laughed, then stopped suddenly as the eyes of each met those of the others across the table.
11. The horse obeys his master because his eyes magnify so that his master seems to the horse to be much larger than the horse himself.

The items were first scored by allowing two points for a correct response and one point for a partially correct response, but differentiations between the correct, the partially correct, and the zero responses were difficult, and a plus or minus scoring system was finally adopted. In general correct answers are those stating the absurdity, those correcting the absurdity, or commenting upon it in such a way as to show its clear recognition, or, less good, answers definitely implying the absurdity. Incorrect answers represent for the most part incorrect criticisms or interpretations, irrelevant comments, answers denying the absurdity, answers restating it, or complete failure to respond. Examples of answers considered correct are given below.

Scoring Guide for Oral Absurdities Test

1. How could a haystack be a house?
   Why didn't he say the first house was where Mr. Smith lived?
   The first "house" would not be a house; it would be a barn or a haystack.
2. He'd be dead if he broke his neck.
   That's damage enough.
3. What was the object of warning him if he was to be hanged?
   Too late to warn him.
   He didn't need a warning. He'd be a warning to others.
   That was his finish.
   He expected him to live, I suppose, after he was hanged!
4. There would always be a last car.
   If the last car is taken off, the car next to the last would be wrecked.
   You would have no cars on the train then.
5. Can't taste in your eye.
   It burns your eye.
6. That is the same as all the way round.
   He was still able to walk around just as far.
7. He was burning twice as much.
   He'd have to burn some coal if he wanted heat.
   He should have bought three stoves to get some back!
8. He is conceited when he says he thinks he's clever.
   He is contradicting himself there.
   He is conceited.
9. Two put together would be fatter not thinner.
   He couldn't compare the two of us; we would make one larger than he is.
   He would have to be fat then.
   Comparing the thinness of one person to two is idiotic.
10. Impossible for a man of normal stature to assume at will the proportions of a dwarf and a giant.
    If he wants a giant he must get a big man, and if he wants a dwarf he must get a small man.

He was neither.
He couldn't be both at the same time.

11. It would have to include that one; then every rule does not have an exception.
If every rule, then that one the same thing, and then it's not true.

Scoring for Printed Absurdities

1. She should have gotten both end pieces by slicing both ends off, not by cutting the whole loaf.
Foolish to cut the whole loaf.
All she had to do was to turn both ends around.
2. You would have to get the letter to let him know.
How could the person know if they didn't receive the letter?
3. He was already dead with the first shot.
He couldn't be killed twice.
The second one couldn't.
They would not have to shoot the second shot.
Not necessary for a second to kill him after one did.
4. Impossible to twirl a stick with his hands in his pockets.
5. The postmaster could not give the Irishman his letter if he did not know his name.
He would not know what name to look for.
He wouldn't be able to tell which envelope.
6. Two skulls for St. Paul; that's one too many.
There couldn't have been two skulls.
St. Paul lived to be a man. Couldn't get skull when he was a boy.
St. Paul had only one skull and the relic part is a fake.
7. If attendance is voluntary, why need the boys look out?
Voluntary is foolish; it was really compulsory.
That was a threat; it wasn't voluntary.
Voluntary means to do what you please, so if a boy is absent he would be doing what the teacher said.
8. That's the same thing.
Just made the job right.
He did what he told him to do already, if he did not tell him which shoe to make larger.
One would be smaller anyway.
9. If any of the three had ordinary eyesight, they could see the top alone.
It would not take but one person to see the top of the tree.
The boy could see as high as two men and a boy.
Why does it take three people to see the top? That's foolish.
How do they relay that seeing game?
10. Three men's eyes could not meet all at the same time.

How see two other pairs at once?
They would have to be crosseyed.
Their eyes could not meet.

11. Would magnify himself too.
Horse cannot see himself to compare with his master.

PINTNER-TOOPS REVISED DIRECTIONS TEST. This test was useful for certain abnormal cases as a quick index of reading ability.[9] Interpretation of the results was sometimes difficult, however, because for two items at least the subject could get enough cues from the non-verbal test situation to make the correct response. For normal subjects the test is not simply an index of reading ability.

2. EDUCATIONAL ACHIEVEMENT TESTS

a. *Reading tests*

GRAY ORAL READING PARAGRAPHS. Sets I through IV covered the necessary range of difficulty so well that the examiners used the particular set or sets best suited to the reading level of each subject.[10] This procedure made for efficient testing but the results were difficult to handle. As noted in Chapter IV, about 90 per cent of the group took Set IV, and the results for these subjects were presented, with the scores based on the total time for reading the 150 words of the test plus 1/50 of the total time for each error.[11]

THORNDIKE MCCALL READING SCALE. This test was given and scored according to the author's directions.[12] Form 1 was used for all the normal subjects. No time limit was set, but with very few exceptions the subjects finished all they could do in twenty to thirty minutes.

CHAPMAN UNSPEEDED READING-COMPREHENSION TEST. This was selected especially because the non-verbal response made it valuable in the aphasia study and it turned out to be well adapted in form and content to normal adult subjects.[13] No time limit was set, but the subjects usually finished within thirty minutes.

b. *Spelling tests*

GATES ORAL SPELLING TEST. This test, composed of 36 words, was scored in terms of the total number of words right on the first attempt or corrected spontaneously by the subject.[14]

[9] Published by C. H. Stoelting, Chicago.

[10] Published by the Public School Publishing Company, Bloomington, Illinois.

[11] A number of these also took Set III, but standard errors were too high to estimate the score on one set from that on another: 26.16 for IV estimated from III, and 9.16 for III estimated from IV.

[12] Published by Teachers College Bureau of Publications, Columbia University. A scoring sheet comes with the tests.

[13] Published by Educational Test Bureau, 3416 Walnut Street, Philadelphia.

[14] A. I. Gates, *The Improvement of Reading*, 1927, 387–388.

MORRISON MCCALL SPELLING SCALE. This test proved to be valuable for the abnormal cases but on the whole too easy for the normal.[15] List 1 was used for the normal subjects.

STANFORD ACHIEVEMENT DICTATION TEST. The test used was the New Form W.[16] The sentences were begun at a level well below the subject's ability, usually at the point where the sixth grade is supposed to begin, and continued until all three crucial words in two successive sentences had been failed. In the case of some few subjects this second condition was not obtained; for after repeated failures, although not on six words in succession, they refused to continue the test.

c. *Writing a Letter*

In the course of the examination, each subject was asked to write a letter, to whomever he pleased and on whatever subject he pleased. When all the letters had been collected, scale values for each were determined on the basis of the Nassau County Supplement to the Hillegas Scale.[17] Each scale value represented the value assigned by at least two of the three psychologists working independently, or in the comparatively rare cases in which three different values were assigned, the middle one of these.

d. *Arithmetic tests*

The Stanford Achievement Arithmetic Test was used in the old Form A; both sections, Computation and Reasoning, were given.[18] No time limit was set, but few subjects required more than fifteen to twenty minutes for each.

3. NON-LANGUAGE TESTS

a. *Tests from the Pintner Paterson Performance Scale*

Many of the ten tests of the Pintner Paterson Performance Scale which were used were poor in discriminating among the adults of the normal group, and only the best of them—the Mare and Foal, the Seguin, and the Substitution—have been reported.[19] All these were given and scored according to the Pintner-Paterson directions.

b. *Pintner Non-Language Mental Test*

The Pintner Non-Language Mental Test was given as an individual test,

[15] *Morrison McCall Spelling Scale*, Yonkers: World Book Company.

[16] A sample set of the New Stanford Achievement Dictation Test, containing the forms V, W, and X, may be obtained from the World Book Company, Yonkers.

[17] M. R. Trabue, *Supplementing the Hillegas Scale*, Teach. Coll. Rec., 1917, 18, 51–84.

[18] Tests and manual published by the World Book Company, Yonkers.

The old Form A was used because it had been given to a number of the abnormal cases before the new form was published.

[19] R. Pintner and D. G. Paterson, *A Scale of Performance Tests*, 1917.

with the necessary demonstrations on paper rather than on a blackboard, but otherwise exactly according to Pintner's directions.[20] Pintner's weighting system was followed in obtaining the total score, but the data for the various sub-tests in terms of raw score were also analyzed separately.

c. *Cancellation*

The Cancellation Test was one of a small group of relatively simple tests tried out with the abnormal cases, and the only one used with the normal subjects.[21] The score was taken as the number of *a*'s correctly crossed in one and a half minutes.

d. *Immediate Recall: Kuhlmann Binet*

The Test of Immediate Recall of Unfamiliar Forms was given and scored according to Kuhlmann's directions, although for purposes of comparison with the abnormal cases time and error scores were also calculated separately.[22]

e. *Drawing a Chair*

The test of Drawing a Chair was arranged for the abnormal cases, to study the ability to draw from a model. The chair was set at such an angle that the subject could see the left rear leg between the right front and rear legs. If he started to draw the chair as seen straight from the front, he was allowed to continue without comment. The type of chair always used was an upright chair without arms.

The scoring system devised was modeled after that Goodenough arranged for the drawing of a man in that points were given for the essential components of the drawing and for the skill and accuracy with which these were represented.[23]

Scoring for Drawing of a Chair

1a. Top line for back of chair
 b. Upright on one side of back
 c. Upright on the other side
 d. Middle slat, bar, or panel
2a. Seat of chair
 b. Attachment of seat to back
3a. Left rear leg

[20] Pintner, *A Non-Language Group Intelligence Test,* J. Appl. Psychol., 1919, 3, 199–214. Test manual and booklets published by College Book Company, Columbus, Ohio.

[21] Published by C. H. Stoelting Company, Chicago; Test No 27008
See A. Bronner, W. Healy, *et al., A Manual of Individual Mental Tests and Testing,* 1927, 138.

[22] F. Kuhlmann, *A Handbook of Mental Tests,* 1922, 134.

[23] Goodenough, *op. cit.,* 87–110.

b. Left front leg
c. Right front leg
d. Right rear leg
e. Correct position and angle of left rear leg
f. Correct position and angle of left front leg
g. Correct position and angle of right front leg
h. Correct position and angle of right rear leg
i. Stretchers
j. Correct number of stretchers
k. Correct position of stretchers

4a. Proportion: back narrower than high
b. Proportion: back longer than front legs
c. Proportion: seat area less than that of back
d. Proportion: legs roughly equivalent in length

5a. Lines "reasonably firm and mostly meeting"
b. Lines firm and meeting

6a. Perspective indicated by shape of seat
b. Angle of attachment of seat and back, indicating that seat and back are approximately at right angles
c. Legs showing no transparency, and correct superposition, i.e., of seat, legs, and stretchers

Each drawing was scored independently by two of the three psychologists; moot points were then discussed with the third psychologist and a final decision reached.

The 35 drawings were ranked by each of the three psychologists. The rank-difference correlation between the average rank so determined and the rank in order of score assigned was $+.88 \pm .03$.

f. *Goodenough Drawing*

The Goodenough Drawing Test was scored according to Goodenough's directions.[24] As for the drawing of a chair, each production was scored by two of the three psychologists, working independently; moot points were discussed with the third and a final decision reached. The 66 drawings were then rated, first by the three psychologists, who ranked them in order of their quality as representations of a man, and then by three artists. The rank-difference correlation between the two rankings was surprisingly high, $+.97$. The correlation between the ranking assigned by the three artists and the order in terms of the Goodenough Score was also high, $+.91$.

g. *Porteus Maze Test*

The Porteus Maze Test was given and scored according to Porteus' direc-

[24] *Ibid.*, 85–153.

tions.[25] After a number of tests had been made, it became evident that the recording of times would help in the analysis of the response. In addition to the total time, the number of seconds between the presentation of the maze and the beginning of the tracing was recorded, and also the duration of any marked pauses at particular points.

[25] Porteus, *Porteus Tests—The Vineland Revision,* 1919; *Studies in Mental Deviations,* 1922, 75–115.

BIBLIOGRAPHY

ALEXANDER, H. B. A Comparison of the Ranks of American States in Army Alpha and in Social-Economic Status. Sch. & Soc., 1922, 16, 388–392.

ANONYMOUS. The Experience of the German Army with the Defective and the Feeble-Minded. J. Psycho-Asthenics, 1911, 16, 68–76.

BABCOCK, H. An Experiment in the Measurement of Mental Deterioration. Arch. Psychol., 1930, 18, No. 117. 105 pp.

BALLARD, P. B. The Limit of the Growth of Intelligence. Brit. J. Psychol., Gen. Sec., 1921, 12, 125–141.

BEESON, M. F. Intelligence at Senescence. J. Appl. Psychol., 1920, 4, 219–234.

BINET, A., and HENRI, V. La psychologie individuelle. Année psychol., 1895, 2, 411–465.

BINET, A., and SIMON, TH. Sur la nécessité d'une méthode applicable au diagnostic des arriérés militaires. Annales médico-psychol., 1910, 11, 123–136.

BINGHAM, W. V., and DAVIS, W. T. Intelligence Test Scores and Business Success. J. Appl. Psychol., 1924, 8, 1–22.

BRIGHAM, C. C. A Study of American Intelligence. Princeton: Princeton Univ. Press, 1923. xxv + 210 pp.

BRIGHAM, C. C. A Study of Error. New York: College Entrance Examination Board, 1932. xiii + 384 pp.

BRIGHAM, C. C. Intelligence Tests of Immigrant Groups. Psychol. Rev., 1930, 37, 158–165.

BROLYER, C. R. Report of the Commission on the Scholastic Aptitude Test. Annual Report of the College Entrance Examination Board, 1930–1934, 5–9.

BRONNER, A. F., HEALY, W., LOWE, G., and SHIMBERG, M. A Manual of Individual Mental Tests and Testing. Boston: Little, Brown, 1927. x + 279 pp.

BROOKS, F. D. The Psychology of Adolescence. Boston: Houghton Mifflin, 1929. xxiii + 629 pp. Chapters IV, V, and VI.

BURR, E. Minimum Intellectual Levels of Accomplishment in Industry, J. Person. Res., 1924, 3, 207–212.

CALDWELL, H. H. Adult Tests of the Stanford Revision Applied to University Faculty Members. J. Exper. Psychol., 1922, 5, 247–262.

CAROTHERS, F. E. Psychological Examinations of College Students. Arch. Psychol., 1922, 6, No. 46. 82 pp.

CARTER, H. D. Family Resemblances in Verbal and Numerical Abilities. Genet. Psychol. Monog., 1932, 12, 1–104.

CATTELL, J. McK. Mental Tests and Measurements. Mind, 1890, 15, 373–381.

CATTELL, J. McK., and FARRAND, L. Physical and Mental Measurements of the Students of Columbia University. Psychol. Rev., 1896, 3, 618–648.

CATTELL, R. B. Occupational Norms of Intelligence and the Standardization of an Adult Intelligence Test. Brit. J. Psychol., Gen. Sec., 1934–35, 25, 1–28.

CHENG, C. Y. Psychological Tests Found in China during the Han and Wei Dynasties. Chinese J. Psychol., 1922, 1, No. 4, 7. Abstracted in Psychol. Abstracts, 1928, 2, 461.

COBB, M. V. The Limits Set to Educational Achievement by Limited Intelligence. J. Educ. Psychol., 1922, 13, 449–464; 546–555.

CONRAD, H. S. General-Information, Intelligence, and the Decline of Intelligence. J. Appl. Psychol., 1930, 14, 592–599.

CONRAD, H. S. The Measurement of Adult Intelligence, and the Requisites of a General Intelligence Test. J. Soc. Psychol., 1931, 2, 72–86.

CONRAD, H. S., and JONES, H. E. Psychological Studies of Motion Pictures. III: Fidelity of Report as a Measure of Adult Intelligence. Univ. Calif. Publ. Psychol., 1929, 3, 245–276.

CONRAD, H. S., and JONES, H. E. Psychological Studies of Motion Pictures. IV: The Technique of Mental-Test Surveys among Adults. Univ. Calif. Publ. Psychol., 1929, 3, 277–284.

CONRAD, H. S., and JONES, H. E. Psychological Studies of Motion Pictures. V: Adolescent and Adult Sex Differences in Immediate and Delayed Recall. J. Soc. Psychol., 1931, 2, 433–459.

CONRAD, H. S., JONES, H. E., and HSIAO, H. H. Sex Differences in Mental Growth and Decline. J. Educ. Psychol., 1933, 24, 161–169.

COWDERY, K. M. Measures of General Intelligence as Indices of Success in Trade Learning. J. Appl. Psychol., 1922, 6, 311–330.

DAWSON, S., and CONN, J. C. M. Intelligence and Disease. London: H. M. Stationery Office, 1931. 53 pp. (Medical Research Council, Special Report Series, No. 162.)

DOLL, E. A. The Average Mental Age of Adults. J. Appl. Psychol., 1919, 3, 317–328.

FERNALD, M. R., and ARLITT, A. H. A Psychological Study of a Group of Crippled Children of Various Types. Sch. & Soc., 1925, 21, 449–452.

FISHER, R. A. Statistical Methods for Research Workers. 3rd ed. Edinburgh: Oliver and Boyd, 1930. xiii + 283 pp.

FOSTER, J. C., and TAYLOR, G. A. The Applicability of Mental Tests to Persons over Fifty Years of Age. J. Appl. Psychol., 1920, 4, 39–58.

FREEMAN, F. N. Mental Tests. Their History, Principles and Applications. Boston: Houghton Mifflin, 1926. ix + 491 pp.

FRY, F. D. The Correlation of the Reverse Audito-Vocal Digit Memory Span with the General Intelligence and Other Mental Abilities of 308 Prisoners of the Eastern State Penitentiary of Pennsylvania. Psychol. Clin., 1931, 19, 156–164.

FRYER, D. Occupational-Intelligence Standards. Sch. & Soc., 1922, 16, 273–277.

GARRETT, H. E. Statistics in Psychology and Education. New York: Longmans, Green, 1926. xiii + 312 pp.

GARRETT, H. E., and SCHNECK, M. R. Psychological Tests, Methods, and Results. New York: Harper, 1933. x + 137 + 224 pp.

GARTH, T. R. A Review of Race Psychology. Psychol. Bull., 1930, 27, 329–356.

GATES, A. I. The Correlations of Achievement in School Subjects with Intelligence Tests and Other Variables. J. Educ. Psychol., 1922, 13, 129–139; 223–235; 277–285.

GATES, A. I. The Improvement of Reading. New York: Macmillan, 1927. xii + 440 pp.

CAW, F. A Study of Performance Tests. Brit. J. Psychol., Gen. Sec., 1924–25, 15, 374–392.

GOODENOUGH, F. L. Measurement of Intelligence by Drawings. Yonkers: World Book Co., 1926. xi + 173 pp.

GRACE, A. G. Individual Differences in Adults. J. Educ. Psychol., 1932, 23, 179–186.

HALL, G. S. Senescence: The Last Half of Life. New York: Appleton, 1922. xxvii + 518 pp.

HOLLINGWORTH, H. L. Mental Growth and Decline: A Survey of Developmental Psychology. New York: Appleton, 1927. xii + 388 pp.

HSIAO, H. H. Performance of Army Alpha as a Function of Age. Thesis presented in fulfilment of requirements for M.A. degree, Columbia University, 1927. Unpublished.

JONES, H. E., and CONRAD, H. S. The Growth and Decline of Intelligence: A Study of a Homogeneous Group between the Ages of Ten and Sixty. Genet. Psychol. Monog., 1933, 13, 223–298.

KELLEY, T. L. Comparable Measures. J. Educ. Psychol., 1914, 5, 589–595.

KELLEY, T. L. Crossroads in the Mind of Man. A Study of Differentiable Mental Abilities. Stanford University: Stanford Univ. Press, 1928. vii + 236 pp.

KELLEY, T. L. Individual Testing with Completion Test Exercises. Teach. Coll. Rec., 1917, 18, 371–382.

KING, I., and GOLD, H. A Tentative Standardization of Certain "Opposites Tests." J. Educ. Psychol., 1916, 7, 459–482.

KLINEBERG, O. An Experimental Study of Speed and Other Factors in "Racial" Differences. Arch. Psychol., 1928, 15, No. 93. 111 pp.

KNIGHT, F. B. The Significance of Unwillingness to be Tested. J. Appl. Psychol., 1922, 6, 211–212.

KOLLARITS, J. Ueber Sprach- und Schreibstörungen im allgemeinen und als "kleine Zeichen der Geistesschwächung" bei Tuberkulose und im Alter. Das Verhältnis zwischen innerer Sprache und Schreibfehler. Arch. f. Psychiat. u. Nervenkh., 1933, 99, 109–196.

KORNHAUSER, A. W., and KINGSBURY, F. A. Psychological Tests in Business. Chicago: University of Chicago Press, 1924. ix + 187 pp.

KUHLMANN, F. A Handbook of Mental Tests. Baltimore: Warwick and York, 1922. 208 pp.

LAMBETH, M., and LANIER, L. H. Race Differences in Speed of Reaction. J. Genet. Psychol., 1933, 42, 255–297.

LEHMAN, H. C., and WITTY, P. A. Sex Differences: Some Sources of Confusion and Error. Amer. J. Psychol., 1930, 42, 140–147.

LINCOLN, E. A. The Mental Age of Adults. J. Educ. Res., 1922, 6, 133–144.

LIPPMANN, W. The Mental Age of Americans. New Republic, 1922, 32, 213–215, 246–248, 275–277, 297–298, 328–330; 1923, 33, 9–10.

LOGIE, H. B., ed. Standard Classified Nomenclature of Disease. New York: The Commonwealth Fund, 1933. xviii + 702 pp.

MARSH, C. J. Human Adaptability as Related to Age. Psychol. Bull., 1933, 30, 589.

MARSH, C. J. Reliability of Performance Tests with Adult Subjects. Psychol. Bull., 1934, 31, 594.

MEILI, R. Recherches sur les formes d'intelligence. Arch. de psychol., 1930, 22, 201–284.

MERRILL, M. A. Intelligence of Policemen. J. Person. Res., 1927, 5, 511–515.

MILES, C. C. Age and Certain Personality Traits of Adults. Psychol. Bull., 1933, 30, 570.

MILES, C. C. Influence of Speed and Age on Intelligence Scores of Adults. J. Gen. Psychol., 1934, 10, 208–210.

MILES, C. C., and MILES, W. R. The Correlation of Intelligence Scores and Chronological Age from Early to Late Maturity. Amer. J. Psychol., 1932, 44, 44–78.

MILES, W. R. Correlation of Reaction and Coordination Speed with Age in Adults. Amer. J. Psychol., 1931, 43, 377–391.

MILES, W. R. Movement Interpretations of the Silhouette of a Revolving Fan. Amer. J. Psychol., 1931, 43, 392–405.

MILES, W. R. Abilities of Older Men. Person. J., 1933, 1, 352–357.

MILES, W. R. The Maintenance of our Mental Abilities. Scient. Mo., 1933, 37, 549–552.

MILES, W. R. Age and Human Ability. Psychol. Rev., 1933, 40, 99–123.

MORGANTHAU, D. R. Some Well-Known Mental Tests Evaluated and Compared. Arch. Psychol., 1922, 7, No. 52. 54 pp.

MURSELL, G. R. Decrease in Intelligence with Increase in Age among Inmates of Penal Institutions. J. Juv. Res., 1929, 13, 197–203.

ODOM, C. L. A Study of the Mental Growth Curve with Special Reference to the Results of Group Intelligence Tests. J. Educ. Psychol., 1929, 20, 401–416.

PAULSEN, A. E. The Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency. Arch. Psychol., 1924, 11, No. 69. 45 pp.

PETERSON, J. Early Conceptions and Tests of Intelligence. Yonkers: World Book Co., 1925. xiv + 309 pp.

PETERSON, J., and LANIER, L. H. Studies in the Comparative Abilities of Whites and Negroes. Baltimore: Williams & Wilkins, 1929. vi + 156 pp. (Mental Measurement Monographs, vol. 2, serial No. 5.)

PIÉRON, H. Theoretical and Practical Aspects of Intelligence. Brit. J. Psychol., Gen. Sec., 1932, 22, 353–358.

PINTNER, R. A Non-Language Group Intelligence Test. J. Appl. Psychol., 1919, 3, 199–214.

PINTNER, R., and PATERSON, D. G. A Scale of Performance Tests. New York: Appleton, 1917. x + 213 pp.

PINTNER, R., and TOOPS, H. A. A Mental Survey of the Population of a Workhouse. J. of Delinq., 1917, 2, 278–287.

PORTEUS, S. D. Porteus Tests—The Vineland Revision. Vineland, N. J.: The Training School, 1919. 44 pp.

PORTEUS, S. D. Studies in Mental Deviations. Vineland, N. J.: The Training School, 1922. xi + 276 pp. Pages 75–115.

PORTEUS, S. D. Maze Tests and Mental Differences. Vineland, N. J.: Smith Printing and Publishing House, 1933. ix + 212 pp.

PRESSEY, L. W. Sex Differences Shown by 2,544 School Children on a Group

Scale of Intelligence, with Special Reference to Variability. J. Appl. Psychol., 1918, 2, 323–340.

PRESSEY, S. L., and COLE, L. W. Are the Present Psychological Scales Reliable for the Examination of Adults? J. Abnorm. Psychol., 1918–19, 13, 314–323.

PRICE, B. The Grasping of Spoken Directions as an Age Function in Adults. Psychol. Bull., 1933, 30, 588–589.

RICHARDSON, C. A. The Growth and Variability of Intelligence. Brit. J. Psychol., Monog. Suppl., 1933, 18, 1–23.

RICHARDSON, C. A., and STOKES, C. W. The Growth and Variability of Intelligence. Brit. J. Psychol., Monog. Suppl., 1933, 18, 24–83.

ROGERS, A. L. The Growth of Intelligence at the College Level. Sch. & Soc., 1930, 31, 693–699.

ROGERS, M. C. Adenoids and Diseased Tonsils; their Effect on General Intelligence. Arch. Psychol., 1922, 7, No. 50. 70 pp.

RUCH, F. L. Adult Learning. Psychol. Bull., 1933, 30, 387–414.

SCHWESINGER, G. C. Heredity and Environment. New York: Macmillan, 1933. viii + 472 pp.

SHARP, S. E. Individual Psychology: A Study in Psychological Method. Amer. J. Psychol., 1898–99, 10, 329–391.

SIMPSON, B. R. Correlations of Mental Abilities. New York: Teachers College, Columbia University, 1912. 122 pp. (Contributions to Education, No. 53.)

SORENSON, H. Mental Ability over a Wide Range of Adult Ages. J. Appl. Psychol., 1933, 17, 729–741.

SPEARMAN, C. The Abilities of Man, their Nature and Measurement. New York: Macmillan, 1927. xxiii + 415 pp.

SYMONDS, P. M. A Second Approximation to the Curve of the Distribution of Intelligence of the Population of the United States, with a Note on the Standardization of the Binet-Simon Scale. J. Educ. Psychol., 1923, 14, 65–81.

TEAGARDEN, F. M. A Study of the Upper Limits of the Development of Intelligence. New York: Teachers College, Columbia University, 1924. 112 pp. (Contributions to Education, No. 156.)

TERMAN, L. M. The Measurement of Intelligence. Boston: Houghton Mifflin, 1916. xviii + 358 pp.

TERMAN, L. M. Mental Growth and the I.Q. J. Educ. Psychol., 1921, 12, 325–341.

TERMAN, L. M. The Great Conspiracy. New Republic, 1923, 33, 116–120.

TERMAN, L. M., *et al.* A Trial of Mental and Pedagogical Tests in a Civil Service Examination for Policemen and Firemen. J. Appl. Psychol., 1917, 1, 17–29.

THOMPSON, H. B. The Mental Traits of Sex. Chicago: University of Chicago Press, 1903. vii + 188 pp.

THORNDIKE, E. L. On the Improvement in Intelligence Scores from Fourteen to Eighteen. J. Educ. Psychol., 1923, 14, 513–516.

THORNDIKE, E. L. On the Improvement in Intelligence Scores from Thirteen to Nineteen. J. Educ. Psychol., 1926, 17, 73–76.

THORNDIKE, E. L., *et al.* The Measurement of Intelligence. New York: Teachers College, Columbia University. xxvi + 604 pp.

THORNDIKE, E. L., *et al.* Adult Learning. New York: Macmillan, 1928. x + 323 pp.

THURSTONE, L. L. The Intelligence of Policemen. J. Person. Res., 1922, 1, 64–74.

THURSTONE, L. L. Intelligence Tests in Civil Service. J. Person. Res., 1924, 2, 431–441.

THURSTONE, L. L. A Method of Scaling Psychological and Educational Tests. J. Educ. Psychol., 1925, 16, 433–451.

THURSTONE, L. L. The Vectors of Mind. Psychol. Rev., 1934, 41, 1–32.

THURSTONE, L. L., and ACKERSON, L. The Mental Growth Curve for the Binet Tests. J. Educ. Psychol., 1929, 20, 569–583.

TRABUE, M. R. Supplementing the Hillegas Scale. Teach. Coll. Rec., 1917, 18, 51–84.

VAN WAGENEN, M. J. Graded Opposites and Analogies Tests. J. Educ. Psychol., 1920, 11, 241–263.

VITELES, M. S. Industrial Psychology. New York: Norton, 1932. xviii + 633 pp.

WALTERS, F. C. A Statistical Study of Certain Aspects of the Time Factor in Intelligence. New York: Teachers College, Columbia University, 1927. viii + 84 pp. (Contributions to Education, No. 248.)

WEISENBURG, T., and MCBRIDE, K. E. Aphasia. A Clinical and Psychological Study. New York: The Commonwealth Fund, 1935. xvi + 623 pp.

WELLS, F. L. Army Alpha, Revised. Person. J., 1932, 10, 411–417.

WHIPPLE, G. M. Manual of Mental and Physical Tests. Part II, Complex Processes. 3rd ed. Baltimore: Warwick and York, 1921. 336 pp.

WILLOUGHBY, R. R. Family Similarities in Mental-Test Abilities. Genet. Psychol. Monog., 1927, 2, 237–277.

WILSON, W. R. Information as a Measure of Intelligence and Maturity. J. Educ. Psychol., 1924, 15, 309–312.

WISSLER, C. The Correlation of Mental and Physical Tests. Psychol. Rev., Monog. Suppl., 1901, 3, No. 16. 62 pp.

WOOD, B. D. Measurement in Higher Education. Yonkers: World Book Co., 1923. xi + 334 pp. Chapter II.

WOODWORTH, R. S. Computing the Standard Deviation and Probable Error of a Binet Mental Age. Psychol. Bull., 1928, 25, 167.

WORTHINGTON, M. R. A Study of Some Commonly Used Performance Tests. J. Appl. Psychol., 1926, 10, 216–227.

YERKES, R. M., ed. Psychological Examining in the United States Army. Washington: Government Printing Office, 1921. vi + 890 pp. (Memoirs of the National Academy of Sciences, vol. 15.)

INDEX