

LB

1131

C3

UC-NRLF



\$B 93 419

Psychological Examinations of College Students

BY

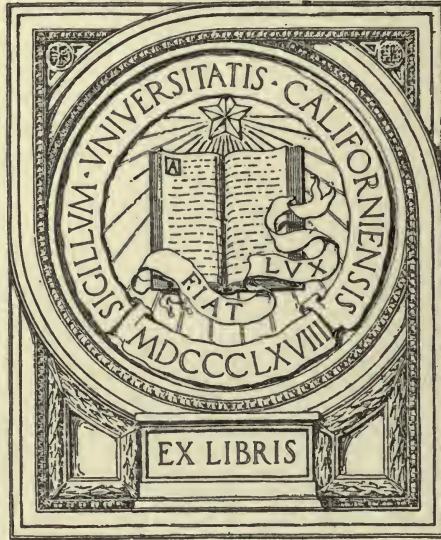
F. EDITH CAROTHERS

*Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy, in the Faculty of
Philosophy, Columbia University*

NEW YORK

1922

EXCHANGE



EX LIBRIS

UNIVERSITY OF CALIFORNIA
Psychological Examinations of
College Students

BY
F. EDITH CAROTHERS

*Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy, in the Faculty of
Philosophy, Columbia University*

NEW YORK

1922

517566

THE
ABSOLUTE

LB1131
C3

Exchange

G. W. H.

UNIVERSITY OF
CALIFORNIA

ACKNOWLEDGMENT

The writer is indebted to Professor H. L. Hollingworth for the original plan of this investigation and the opportunity to conduct the tests in his laboratory; for his constant interest and his valuable suggestions in the preparation of the results. For helpful criticism regarding the arrangement and statistical treatment of the data she is also indebted to Professor R. S. Woodworth, Professor A. T. Poffenberger, and Professor J. McKeen Cattell.

C -

1000
1000
1000
1000

SECTION I

HISTORY OF FRESHMAN TESTS

The recommendation in 1882 by Galton¹ of the establishment of anthropometric and medico-metric laboratories for the examination of individuals represents the first definite recognition of the need of examining individuals in order to give them vocational guidance. Galton saw the importance both to science and to individuals of collecting complete life-histories of people which should include photographs, anthropometric measurements, and medical facts. To meet this need he established his now famous laboratory in the South Kensington Museum, London. There, by payment of a small fee, individuals could go and have certain physical measurements made and undergo tests for keenness of vision and hearing, dynamometer pressure, reaction time, etc.

Several years later, at the World's Columbia Exposition in 1893,² Professor Joseph Jastrow arranged a laboratory devoted to tests of a strictly psychological nature. Prior to Jastrow's work, however, Cattell proposed³ and tried out a series of ten mental tests and measurements on students in the psychological laboratory of the University of Pennsylvania. In devising his series of tests Cattell followed Galton in combining physical measurements with psychophysical and strictly mental tests. He went a step farther, however, by emphasizing the necessity of standardizing methods of procedure in administering tests so that results secured by different experimenters might be comparable. In addition to the Pennsylvania students, tests were also given to the students of Cambridge University and Bryn Mawr College.

Galton's work stimulated other investigators to devise tests for measuring the capacities of individuals. Of particular interest is the list of ten fundamental traits or properties proposed by Kraepelin⁴ as the basic factors to be considered in examining both normal individuals and the "mentally sick." These so-called fundamental dispositions include: the mental capacity to do work, the ability to

¹ Fortnightly Review, 1882, p. 332.

² Cattell and Farrand, L. Physical and Mental Measurements of the Students of Columbia University.

³ "Mental Tests and Measurements," J. McK. Cattell with appendix by Francis Galton, Mind, 1890.

⁴ Der Psychologische Versuch in der Psychiatrie; Emil Kraepelin, Psychologische Arbeiten, 1895.

be influenced by practice, strength of practice or general memory, special memory ability, susceptibility, fatigability, the ability to recuperate, the depth of sleep, the intensity of distraction and adaptability. To each one of these fundamental traits Kraepelin arbitrarily assigned a certain test, assuming that excellence of performance in the assigned test, say adding, would indicate excellence in the corresponding quality, say the capacity to do work. Although his assumption, without statistical proof, that certain tests would measure certain functions rendered his results inaccurate, from the modern standpoint, his work is interesting in that it is representative of a distinct stage in the use of tests for diagnostic purposes.

With the accumulation of data and the gradually increasing clearness of conception of the meaning of tests, methods of administering them were revised. In 1896⁵ appeared the first report of the results of mental and physical tests made on freshmen only. It concerned the work done by Professor Cattell and Dr. Farrand on one hundred Columbia University students in 1894-5 and 1895-6. At this time there was conceived the plan of testing Columbia students during their freshman and senior years. Their tests comprised ten records and twenty-six measurements. Such physical measurements were taken as the color of hair and eyes, height and weight, breathing capacity, sensation areas, and strength of right and left hands. Other measures were of a sensory character, while certain simple tests of a mental character were taken, such as the rate of perception and the perception of space and time. In addition, a personal record-blank was filled out by the student and a record of the impressions made upon him by the subject was filled in by the experimenter both before and after testing. The tests were given individually, the investigators and several assistants acting as experimenters, and required from forty minutes to one hour for their completion. The underlying purpose in giving these tests is clearly stated in this statement by Cattell and Farrand:⁶

"When used with freshmen on entering college the record is of interest to the man and may be of real value to him. It is well for him to know how his physical development, his senses, his movements, and his mental processes compare with those of his fellows. He may be able to correct defects and develop aptitudes. Then when the tests are repeated later in the college course and in subsequent life the record of progress or regression may prove of substantial importance to the individual."

⁵ Cattell, J. McK., and Farrand, L. *Physical and Mental Measurements of the Students of Columbia University*, *Psychological Review*, 1896, III, 618-647.

⁶ Above reference.

These Columbia freshman tests continued to be given each year under Professor Cattell's direction. In 1901⁷ an account and discussion of the results was published by Wissler. He discusses the changes and additions made in the tests and considers the records of 250 freshmen, a small number of seniors, and some Barnard girls. The tests employed were: length and breadth of head, strength of hands, fatigue, eyesight, color-vision, hearing, perception of pitch, perception of weight, sensation areas, sensitiveness to pain, perception of size, color preference, reaction time, rate of perception, naming colors, rate of movement, accuracy of movement, perception of time, association, imagery, memory, (auditory, visual, logical, and retrospective). Records of stature, weight, etc., together with data concerning parentage, personal habits, and health, the physical measurements taken in the gymnasium, and academic marks were also secured. From the similarity of the results of freshmen tested each year, Wissler concluded that freshmen entering Columbia from year to year are a homogeneous group and represent a type. His general conclusions are:

1. That the laboratory mental tests show little intercorrelation in the case of college students. Correlations range from $-.28$ (accuracy and speed in marking out A's), to $+.39$ (auditory and visual memory—correctly placed).

2. That the physical tests show a general tendency to correlate among themselves, but only to a very slight degree with the mental tests.

3. That the markings of students in college classes correlate with themselves to a considerable degree. Correlations run from $+.11$, (mathematics and logical memory) to $+.75$ (Latin and Greek).

These early Columbia tests and measurements were principally motor and sensory in character, and the few tests that might be considered to have an intellectual quality were so simple that they proved of little value for determining the mental status of the college freshman. They are, however, significant in that they represent the first definite attempt to establish standards of performance for freshmen and to show students how their standing in various tests compared with the average standing of their class.

Subsequent to the establishment of the practice of testing the Columbia students in their freshman and senior years, committees were appointed by the American Psychological Association in 1896

⁷ Wissler, Clark; *The Correlation of Mental and Physical Tests*; *Psychological Review*, Monograph Suppl., Vol. III, No. 1901, p. 62.

and 1907, respectively, to consider the possibility of accumulating mental and physical statistics through cooperation on the part of various psychological laboratories and to devise a standard series of group and individual tests. In 1896 the committee drew up a series of physical and mental tests appropriate for college students tested in a psychological laboratory.

Various other proposals were made for the scientific study of the college student. In 1899 President Harper of Chicago recommended that special study be made of the college student's character, intellectual capacity, and tastes, by the questionnaire method. In 1906 Thorndike⁸ called attention to the fact that the entrance examinations given by the College Entrance Board of the Association of Colleges and Preparatory Schools of the Middle States and Maryland did not measure at all accurately the candidate's capacity and emphasized the need of the scientific study of this matter. Williams⁹ also stressed the importance of studying the college student. Like President Harper, he recommended the questionnaire method for ascertaining facts concerning the student's personality, and suggested the use of Whipple's information test for obtaining a knowledge of the student's range of information. He also pointed out the need of vocational advisors for freshmen.

Calfee¹⁰ in 1913 reported the results of four general intelligence tests on 103 freshmen (51 boys and 52 girls) of the University of Texas. The tests used were card-dealing, card-sorting, alphabet-sorting, the mirror test, and the spirometer test for vital capacity. She finds inter-test correlations for the boys and girls combined ranging all the way from $+ .50$ to $.00$. The correlations between the tests and college grades range from $+ .32$ (card sorting and grades) to $+ .16$ (mirror test and grades). The correlation between the lung test and grades is $- .11$. Considering the girls' records alone, the inter-test correlations range from $+ .45$ to $+ .19$, and the correlations with college grades from $+ .28$ to $+ .13$, and with the lung test the correlation is $.00$.

No further attempt to measure the performance of college freshmen in tests is reported until December, 1915, when Dr. Karl T. Waugh presented a paper on "A New Mental Diagnosis of the College Student" before the American Psychological Associa-

⁸ Thorndike, E. L. An Empirical Study of College Entrance Examinations. *Science, N.S.*, 1906, 23, 839-845.

⁹ Williams, C. W. Scientific Study of the College Student.

¹⁰ Calfee, M. College Freshmen and Four General Intelligence Tests, *Journ. of Educ. Psychol.*, 1913, 4, 223-231.

tion.¹¹ In 1912 he applied seven tests¹² individually to freshmen in Beloit College, and three years later, in 1915, he gave the same tests to thirty-nine of the same subjects. Waugh's inter-test correlations range from $-.43$ to $+.54$, and he finds some improvement in the tests from freshman to senior year.

During the year 1913-14 Bingham¹³ gave nine tests to 200 Dartmouth freshmen, seven of them being given individually. As a number of psychology students, unpracticed experimenters, assisted Professor Bingham in his testing, the results of his investigation are somewhat inaccurate. He gives norms for the nine tests, (median, standard deviation and coefficient of variability) and the range from the poorest to the best. As no correlations are reported we have no information as to the relationships between the tests. Bingham's chief contribution consists in his use of the method of ogive percentile graphs. The data in seven of his tests are presented in this form, thus serving as a scale. Given the score made by any individual, the experimenter by reference to the chart can readily assign him a rank among his classmates. The speed with which a student may be thus assigned his relative position in any given trait makes this method a most convenient one for the instructor.¹⁴

At the University of Texas the same year Bell¹⁵ gave nine tests¹⁶ to about seven hundred and fifty freshmen. Bell definitely states that his aim was to devise a series of tests that would "be of assistance to college authorities in aiding freshmen to adjust themselves to their environment." The time required for testing was from forty to forty-five minutes. The tests were given not individually, but in groups averaging a little less than twenty each. The time-limit method was used. This, together with his arbitrary method of scoring the tests may account in some measure for the unsatisfactory nature of his results. He weighted each test so that a perfect

¹¹ Waugh, Dr. Karl T. *A new Mental Diagnosis of the College Student*. New York Times Magazine, January 2, 1916.

¹² Waugh's tests were: 1. Concentration of attention (cancellation of A's); 2. Range of information; 3. Speed of learning (substitution); 4. Quickness of association (opposites); 5. Ingenuity (puzzle-box); 6. Steadiness; 7. Memory for a passage (immediately after hearing it read and after an interval of two weeks).

¹³ Bingham, W. V. Some norms of Dartmouth Freshmen; *Journ. of Educ. Psychol.*, March, 1916, Vol. 7, pp. 129-142.

¹⁴ Bingham's tests were: 1. Endurance of grip; 2. Tapping; 3. Memory span for auditory digits; 4. Logical memory; 5. Cancellation; 6. Color Naming; 7. Logical relations; 8. Mixed relations; 9. Perception of form.

¹⁵ Bell, J. Carleton. *Mental Tests and College Freshmen*; *Journ. of Educ. Psychol.*, Sept., 1916, Vol. 7, pp. 381-399.

¹⁶ Bell's Tests include: 1. Cancellation of triangles; 2. Addition; 3. Association or learning pairs; 4. Recognizing forms; 5. Marking right statements; 6. Easy directions; 7. Hard Directions; 8. Alternatives; 9. Completion (using "The Strength of the Eagle" as material).

mark or the highest mark would approximate 100, and the other marks range downward from this to zero. For example, in the Triangles test there were fifty triangles to be crossed out. Each one correctly crossed out counted two points and five points were deducted for each error, positive or negative. For example, if a student crossed out 35 triangles, omitted 3, and crossed out one circle, his score was $70 \text{ minus } 20 = 50$. The other tests were scored in similar manner.

Bell also obtained the correlations of freshmen university grades with each other and of the university grades with the mental tests. His conclusions are:

1. The correlations between freshmen university grades vary from + .34 (mathematics—history) to + .59 (English—history, science—history).
2. The highest correlation between class marks and test scores is + .31 (English—Completion).
3. Among the tests themselves the highest correlations are found between the Association and Recognition tests, and between the Directions, Alternatives and Completion tests.
4. There is a considerable difference in the results of the tests with the best and the poorest students, but the scores are so variable as to be of little value for individual diagnosis.

The investigations of Calfee, Waugh, Bingham, and Bell illustrate the striking change that has taken place in the character of mental tests since the early Columbia tests were first instituted. In place of sensory and motor tests we now employ tests which will measure diverse mental functions. Motivated by this same desire to secure a group of tests for college students indicative of mental ability, and correlative with college grades, Rowland and Lowden¹⁷ began to try out groupings of psychological tests in 1912-13 and carried out their investigations over a period of three years. The tests were conducted individually on all the students in Reed College, twelve students of experimental psychology assisting in conducting the tests. The first grouping of tests was tried out on 54 students during 1912-13, after which the grouping was revised and given to 195 more subjects. No inter-test correlations are reported. The highest correlation between university grades and the groupings was between the grades and the letter-group g-r-s-t, cancellation, opposites, logical memory, judgment (syllogism), rote memory,

¹⁷ Rowland, E. and Lowden, G. Report of Psychological Tests at Reed College. *Journ. of Exper. Psychol.*, 1916, I, 211-217.

cancellation of words with *a* and *t*, (a correlation of + .37 with a P.E. of $\pm .06$).

Psychological tests have also been conducted for several years at Vassar College. Results of tests made upon Vassar freshmen during the years 1914,¹⁸ 1915, and 1916¹⁹ show data collected from four sources, namely: 1. Answers to a questionnaire calling for information regarding the student's imagery, interests, language facility, and habits; 2. Results of the tests;²⁰ 3. Freshmen academic grades; 4. Reports of promising students by their instructors. To determine roughly the correlation between academic marks and test scores, the difference between the average class standing of students having test scores in the first or highest quarter and the average class standing of students with test scores in the last quarter was found. If there was a marked difference the experimenters concluded that a positive correlation existed. According to this rough method they found a positive correlation between academic marks and the tests except Hard Directions. On the whole, the results of the Vassar tests appeared to indicate that ability in the tests correlates well with ability in freshman studies, while inability to do well in the tests is correlated with a similar inability to do well in freshman studies. Moreover, students designated as "promising" by their instructors tend to manifest a high grade of performance in the tests. (14.5% of 317 freshmen tested in 1917 who passed all the tests in the Terman Superior Adult Tests were rated by their instructors as being of only average ability.) The experimenters also found that the relation between success in freshman tests and academic success in three years' work is less than that between success in freshman tests and academic success in the freshman year. Inasmuch as there were thirty different testers, each one being assigned a small group of freshmen, little confidence may be placed in the accuracy of the data. The tests as conducted at Vassar are of value more for the opportunity they afford students of psychology to acquire training in experimental methods of procedure than for any contribution they make to our knowledge of freshman standards of performance in various tests.

¹⁸ White, Sophie D.; May, Sybil; and Washburn, M. F. A study of Freshmen. *Minor Studies from the Psychological Laboratory of Vassar College*, No. 31, *Amer. Jour. of Psychol.*, 1917, Vol. 28, pp. 151-154.

¹⁹ Montagne, M.; Reynolds, M. M.; and Washburn, M. F. A Further Study of Freshmen. *Amer. Jour. of Psychol.*, 1918, 29, 327-330.

²⁰ The tests described include: Verbal memory and memory for ideas; Reading Backwards; Hard Directions; Analogies; Sentence Building; Suggestibility; Free Association; Thurstone Reasoning.

An interesting contribution in connection with the application of psychological tests to college freshmen is that of Kitson²¹ at the University of Chicago. With the general purpose of devising a "system for measuring the mental capacity of college students in order to guide their college work," Kitson selected sixteen tests.²² About half the tests were given by the group method. The time required for testing was two and one half hours. From forty complete records Kitson computed norms of performance in the various tests. In addition, a graphic chart was arranged for each student to show his standing in each test and to furnish a net score combining his standing in all the tests. In the particular tests used, Kitson found a significant positive correlation only between: 1. Memory for meaningful material seen and heard (+ .54); 2. Between the first and second reproductions of this material (+ .49); 3. Between the Opposites and Constant Increment tests (+ .40). When correlations were computed of standings in each test with standings in the net score, they were found to be somewhat higher. The correlation between college marks and psychological tests was found to be + .44 (P.E. .09) but from forty records secured from a second group of freshmen tested the correlation was found to be only + .20 (P.E. .11). Kitson explains this low correlation on the ground that many other factors besides intelligence enter in to determine standing in school studies, such as the personal factor of the instructor, the student's will power, social surroundings, economic conditions, and physical condition. The correlation between the psychological tests and intelligence as estimated by the dean was + .57 (P.E. .05). Twenty-one of the 1915 freshmen were retested in seven of the tests in their Sophomore year and improvement was shown in every test except one. (Numbers heard.) Comparison between the net score for freshman and sophomore year shows a correlation of + .88 (P.E. .03).

Although his norms of performance in the tests and his inter-test correlations are not very reliable, based as they are upon only forty records, there is much to be said in favor of Kitson's general method of procedure. His emphasis upon the importance of studying the individual student in his relation to the college and his

²¹ Kitson, H. D. *The Scientific Study of the College Student*. Psychol. Monog., 1917, 23 (No. 98), p. 81.

²² The tests employed were: Number-checking; Memory for numbers heard; Memory for objects seen; Memory for logical material heard; Secondary memory for same; Immediate memory for logical material, seen; Secondary memory for same; Loss in logical material, heard; Loss in logical material, seen; Opposites; Constant increment; Hard directions, printed and oral; Word building; Sentence-building; and Business ingenuity.

realization of the fact that psychological measurements, however large the role they may play in determining a student's abilities and aptitudes, must not be considered the sole factor in such a determination, but rather should be so coordinated with measures of the student from various other aspects as to lead to our fuller understanding of the nature of the individual student and his potentialities, signify a decided advance in the method of treating the problem. The splendid cooperation of all the students and his success in dealing with delinquent cases speak much for Kitson's general method.

Other minor investigations have been made on freshmen with the same purpose. Sunne,²³ working at Newcomb College, found a low correlation between college grades and an information test tried on twenty-five freshmen, and with ninety-nine freshmen who were given a series of tests found correlations of tests with grades ranging from 0 to + .25. Haggerty²⁴ found a correlation of a quality of reading test and omnibus test with medical marks of + .62 and + .60, respectively, and of the two combined of + .65, in the case of sixty-nine candidates for medical school who had already completed two years of college.

At the University of Iowa King,²⁵ working with a little group of nineteen freshmen, found a tendency for the students with high academic marks to make higher scores in the completion, logical memory, and lanes test than the students with low academic marks. He gives no statistical evidence in support of this statement. Later, using a series of five tests with 56 freshman engineers, he obtained a correlation between students' ranks in all the tests combined and their academic grades of + .27. The tests employed by King were: 1. Curtis Arithmetic, Series B, (graded for speed and accuracy); 2. Hard Opposites; 3. Recognition of Forms; 4. The Kansas Silent Reading Test, (H.S. Series); and 5. "Hall Cube Test," a test of visual imagination.

A little later Irving King and James M'Crory²⁶ followed Kitson's method more definitely. In the fall of 1916 they tested 276 women and 268 men freshmen in seven different tests: the Curtis Standard

²³ Sunne, D. The Relation of Class Standing to College Tests, *Journ. of Educ. Psychol.*, 1917, 8, 193-211.

²⁴ Haggerty, M. E. Tests of Applicants for Admission to University of Minnesota Medical School. *Journ. of Educ. Psychol.*, 1918, 9, 278-286.

²⁵ King, I. The relationship of abilities in certain mental tests to ability as estimated by teachers, *School & Society*, 1917, 5, 204-209.

²⁶ King, I. and M'Crory, J. Freshman Tests at the University of Iowa, *Journ. of Educ. Psychol.*, 1918, 9, 32-46.

Arithmetic Test, Series B; mixed relations; two tests of "opposites;" a completion test used by Simpson; visualization; Whipple's information test, and a logical memory test. The group method of testing was used, the tests being given in groups of from ten to twenty-five. Their rather low inter-test correlations indicate, they state, that they are measuring a variety of mental functions. They find, moreover, fairly good correlations between the tests and academic grades (+ .14 to + .45 in the case of the girls, and + .21 to + .84 in the case of the boys). In their attempt to make practical application of the tests for the diagnosis of their students in general and cases of special ability and disability, as Kitson does, they have been fairly successful.

At Northwestern University Uhl²⁷ obtained inter-test correlations ranging from + .18 (Trabue Completion K and Information), to + .42 (Trabue Completion M and Information), for a group of one hundred freshmen tested in the fall of 1916. His series contained only four tests: Trabue Completion K and M, a hard opposites list of twenty words, and an information test which consisted of the seventy most familiar words in Whipple's list plus thirty new words. Test correlations with the first semester English and Mathematics grades were determined and found to range from + .48 (English and Mathematics), to + .16 (Completion K and Mathematics). When he had three mathematics instructors rate these one hundred students for ability, Uhl found a correlation of + .93 between their ratings and the Mathematics grades of the students. This high correlation was no doubt due to the tendency on the part of the teachers to make their judgments of the students practically equivalent to the students' course grades. The correlation between the instructor's judgments and the ranks of these same students in their last year of high school was + .59, and with all the tests combined was + .36. Uhl thinks his tests fail to measure accurately, the information test being the most unsatisfactory, and attributes his low correlations to the homogeneity of his group, the relative simplicity of the tests, and the unreliability of school marks.

Thurstone's²⁸ work represents a further development in the use of psychological tests. At the Carnegie Institute of Technology the attempt is made to use psychological tests as a criterion for admission. A series of six mental tests was given to 114 freshmen of the Margaret Morrison Carnegie School in October, 1917. The problem

²⁷ Uhl, W. L. *Mentality Tests for College Freshmen*, *Journ. of Educ. Psychol.*, 1919, 10, 13-28

²⁸ Thurstone, L. L. *Journ. of Educ. Psychol.*, March, 1919.

was to determine whether they could reduce the number of students who were dropped for poor scholarship or placed on probation for poor scholarship by the use of the mental tests, and to determine whether the mental test ratings correlated with faculty estimates concerning the general ability of the students. The tests which agreed well with the judgment of the faculty were retained. In working up his results Thurstone used the method of critical scores. After plotting scatter diagrams for each test, upper and lower critical scores were determined such that every student above the upper critical score is above the average in the opinion of the faculty, and every student below the lower critical score is below the average in the opinion of the faculty. The mental test rating was designated as the medium percentile rank in all six tests plus 5 points for each test in which the student is above the upper critical score, and minus 5 points for each test in which he is below the lower critical score. Students with a mental test rating of -10 or below were reported as doubtful.

Thurstone found a correlation between instructors' estimates of students' ability and the combined mental test rating of $+.60$. From his results he concluded that: 1. The mental test rating would have eliminated seven of the eleven total failures at the beginning of the year. 2. No average or good student would have been eliminated by the mental test rating. All students who scored below the lower critical mental test rating were, without exception, poor students.

Moreover, all the freshmen who were rated high by the faculty were above the average in the mental test rating. From all indications, this method is working out well at Carnegie.

The past three years have brought a further development in the use of psychological tests for measuring the intelligence of college freshmen. Since 1918 the Army Alpha test has been administered to freshmen in several colleges with varying degrees of success. Professor Stone²⁹ reports that its use at Dartmouth justifies the recent proposal to admit students scholastically in the upper quarter of their class in approved schools. Strictly speaking, the work at Dartmouth should not be included in this history, since it deals with the results obtained in testing all the college classes rather than freshmen only. We mention it here, however, because the college authorities are now devoting particular attention to admin-

²⁹ Stone, Charles Leonard. "Intelligence and Scholarship," *The Dartmouth Alumni Magazine* March, 1920.

istering the test to the freshman class. During the fall of 1918 the Army Alpha test was given to all the students in the Students Army Training Corps which included practically the entire student body. The average score in Alpha for the 677 S. A. T. C. men tested was 147.5. The average academic grade for the same men was 2.12, using the scale D = 1, C = 2, B = 3, A = 4. The correlation between the academic marks and Alpha scores was + .44. There is also a significant correspondence between a student's score in the Alpha test and the scholarship quintile his academic record places him in. Although less exact than Thurstone's method of assigning individuals their relative position in a group, this method serves to give a rough and quick estimate of a student's status.

Similar to this Dartmouth study is that of Walcott's³⁰ at Hamline University. Here, too, not freshmen alone, but all students were given the Alpha test in the fall of 1918. Walcott's results are based on data secured from 61 men and 145 women. As in the Dartmouth investigation, a far greater proportion of men and women students secure a score in Alpha in the high grade intelligence group than was found in any of the army camps. The median score is 129 for the Hamline men and 133 for the women, with the same sharp differentiation between the poor and the good groups as Stone found at Dartmouth. The correlation between the results of the Alpha test for the women and their first term academic grades was + .47, slightly higher than the Dartmouth result. Although Walcott does not consider the army test the best device for determining the fitness of students for college work, he sees in the significant difference in score between the upper and lower half of the students tested, the practical use to be made of this fact in the placing of students.

Similar investigations have also been conducted by Hill, Filler,³¹ and Hunter at the University of Illinois, Dickinson College, and Southern Methodist University, respectively. At the University of Illinois 3,500 students were tested in twenty-four groups in March, 1919,³² members of the faculty acting as experimenters. As at Dartmouth and Hamline, the scores of the students at each of these colleges show them to be a very select group compared to the army men. The median score of the freshmen in the school of liberal arts and sciences at the University of Illinois is 147. At

³⁰ Walcott, G. D. "Mental Testing at Hamline University." *School and Society*, 1919, 10, 57-60.

³¹ Filler, M. G. *A Psychological Test*. *School & Society*, 1919, 10, 208-209.

³² Hill, D. S. *Results of Intelligence Tests at the University of Illinois*; *School & Society*, 1919,

Southern Methodist University³³ the effort was made to secure select groups of students in order to compare their scores with the average score for the school. Each student was asked to name men and women students whom they thought would make high scores in the Alpha test. For 16 men and 8 women named by from five to forty students as being able to make the highest scores, the average score for the men was 154, and for the women 156, justifying the judgment of the students. With a similar group of students named by the faculty as being able to make the highest scores even better results were obtained, the average for the men being 161 and for the women 167. In selecting a group of men and women whom they judged would make low scores the faculty were equally successful. Both faculty and students thus showed themselves fairly good in their ability to select students on the basis of intelligence, though this method of selection is inferior to selection on the basis of actual scores. The correlation between the Alpha scores of the women students and their college grades for the fall term was + .52. No correlations are given in the Illinois and Dickinson reports, which are only preliminary.

The following is a comparative table showing scores obtained at the University of Illinois, Dickinson College, and Southern Methodist University:

	University of Illinois	Dickinson College	Southern Methodist University
Total number tested	3,254	213	321
Number of freshmen	489	72	128
Lowest freshman score	52	75	60
Highest freshman score	188	195	188
Median freshman score	147	141	127

Hunter explains the lower median score at Southern Methodist University as due to a difference in the method of conducting the test.

More fully developed than these three preliminary investigations is the work being done at Brown University.³⁴ Colvin reports the results obtained from 103 freshmen with the Alpha test and two series of psychological tests, known as Brown University Series I and II, which were separated by an interval of several days. Each series consisted of four tests: mutilated sentences, vocabulary, analogies or mixed relations, and a reasoning test. The distribution

³³ Hunter, H. T. *Intelligence Tests at Southern Methodist University*; School & Society, 1919, 10, 437-440.

³⁴ Colvin, S. S. *Psychological Tests at Brown University*; School & Society, 1919, 10, 27-30.

of scores for both Series I and Series II separately and for the combined scores of Series I and II conformed closely to a normal probability curve. The correlation between Brown University Series I and II is + .75, and between the average of these two series and the Alpha test is + .79. The correlation between the Brown University tests and the average academic marks of the first and second terms is + .59, and between the army test and the average of the marks of the first and second terms is + .45. Practical application was made of the tests to foretell a student's probable academic success and to aid in diagnosing cases of failure in school work. Colvin found that two-thirds of 80 students reported as doing unsatisfactory work in the first term had made low scores in their psychological tests, while only one-sixth of the men had a satisfactory grade. Most of the cases of students doing poor college work who had obtained high scores in the tests were due not to lack of ability, but to other reasons. So satisfactory have the tests been in determining the students' mental status and helping them that they are still being employed.

In a recent article in the *Educational Review*³⁵ Professor Colvin compares in greater detail the scores and correlations obtained in the Brown University tests and the Alpha test, and reports results secured in giving the Brown tests and the Thorndike tests to 300 freshmen. The Brown tests require about fifty-five minutes of actual working time as contrasted with about three hours required by the Thorndike tests. The median score for the Brown tests is 62.4 with a standard deviation of 10.59, compared to the median score for the Thorndike tests of 76.5 with a standard deviation of 14.89, the difference being due to the fact that the Brown tests have a maximum score of 100, while the Thorndike tests have a maximum score of about 150. The correlation between the scores obtained by students in the two tests is + .816 with a P.E. of .0138, but the Thorndike tests show a higher correlation with academic marks (+ .53) than the Brown tests (+ .46). While the Thorndike tests show a slight superiority in prognostic value, nevertheless results show that men receiving scores in the lowest fifteen percentile of either the Brown or the Thorndike tests have a relatively small chance of graduating from college. Colvin warns against the danger of refusing men admission to college solely because of a low psychological record. He advocates the conservative position of

³⁵ Colvin, S. S. The Validity of Psychological Tests for College Entrance. *Educational Review*, June, 1920.

regarding the psychological record as one among many factors to be considered in diagnosing cases of individual students.

At Ohio State University the Alpha test was successfully given to 5,950 students October 10, 1919, in groups of one hundred to two hundred and fifty. The distribution of scores for the entire group conformed to the normal probability curve, the students being grouped into five classes as follows:

Class	Score		Approximate Percentage in Each Class
I.	178-212	Very superior intelligence	5
II.	155-177	Superior "	20
III.	115-154	Average "	50
IV.	85-114	Fair "	20
V.	0-84	Poor "	5

The percentage of students falling into each of these five classes was then determined for the various university units separately, such as the Graduate School, Commerce and Journalism, Law, Medicine, Engineering, Arts—Education, Agriculture, Pharmacy, etc. The median, highest, and lowest scores, and the number examined for each class (college year), in each college and in the whole university, are reported. The highest median score, 157, was obtained by the Graduate School; Arts received second place with a median score of 147; Commerce and Journalism third, with a median score of 146; and so on down to a median of 112, (Veterinary Medicine group). The report gives an interesting comparison of the various college groups.

The Thorndike tests, previously mentioned, are rapidly becoming more widely employed for freshmen testing than the Army Alpha. Jones,³⁶ writing in the *Educational Review*, clearly describes the general nature of these tests. Although conceding their practical value, he urges that they should be employed "not to the exclusion of other measures for determining fitness, but along with them." Evidence of a student's fitness to undertake college work should, in Professor Jones' opinion, include his preparation for college work, his character and promise, his health, and his intelligence denoted by his score in the Thorndike test. In a brief report before the New York Branch of the American Psychological Association this year Mr. Wood stated that the purpose of the Thorndike tests was fourfold: 1. To select those fit for a college course; 2. To aid college committees; 3. To assist the progress of schools; 4. To

³⁶ Jones, A. L. *Psychological Tests for College Admission*; *Educational Review*, 1919, 58, 271-278.

assist the Dean in the administration of the college. Results from a large number of freshmen showed a correlation between the total Thorndike score and the average college grade of + .52, and the median college grade of + .54. Although no published reports of results secured with the Thorndike tests have appeared, investigators who are employing the tests find them highly satisfactory.

SECTION II

STATEMENT OF THE PROBLEM WITH A LIST OF THE TESTS EMPLOYED

The present investigation, begun at Barnard College in the fall of 1915, about two years before the Army Alpha and the Thorndike Tests were originated, was carried on during the years 1915-16, 1916-17, the fall of 1917, and the spring of 1919. The general purpose underlying the investigation was similar to that underlying the investigations of other experimenters during this period—a purpose which continues to motivate present studies. The aim was first, to establish norms and standards of performance in mental tests for Barnard freshmen, and second, to give students a clear conception of their abilities and aptitudes along various lines. More specifically, this investigation concerns the trial of a group of tests with the object first, of determining their reliability as measures; second, their correlation with freshman university grades; and third, with physical records taken in the gymnasium.

In selecting the particular group of tests to be used several factors contributed. Paramount in importance was the desire to select a series of tests of such nature as to call into play various mental functions. In addition, it was desired to secure tests which previous investigators had found to have a positive correlation with such factors as age, ability along some vocational line, or general intelligence. Equally important in determining the final selection was the time-limitation factor. Owing to unwillingness on the part of students to act as subjects for a longer period, and to the factor of fatigue which would probably influence the results of tests completed after that time, it was found necessary to have a series of tests such as could be completed in one hour. Consideration of all these factors finally lead to this selection of tests:

- | | | |
|-----------------|--------------------|--------------------|
| 1. Coordination | 8. Verb-object | 14. Word Memory |
| 2. Tapping | 9. Mixed Relations | 15. Logical Memory |
| 3. Cancellation | 10. Word Building | 16. Substitution |
| 4. Checking | 11. Word Naming | 17. Completion |
| 5. Color Naming | 12. Knox Cube | 18. Information |
| 6. Directions | 13. Digit Span | 19. Vocabulary |
| 7. Opposites | | |

SECTION III

METHOD AND TECHNIQUE OF THE INVESTIGATION

Shortly after the beginning of the academic year, in the fall of 1915, the series of tests selected according to the manner described in the preceding section was submitted to a preliminary trial in order to determine the best method of conducting the tests, and to afford the writer practice in their administration. After determining the general method of procedure, a notice was posted in the Freshman Study of Barnard, stating that a series of psychological examinations had been instituted for Barnard freshmen, and giving a description of the nature and purpose of the tests. It was stated that the time required for the examination was one hour, and an accompanying schedule indicated the hours at which the test might be taken. The place where the examinations were to be held was also indicated, and all freshmen interested were requested to sign their names on the schedule opposite the hour at which they could take the test. This method of permitting the student to take the test at the hour most convenient for her, rather than at a time prescribed by the experimenter, seems advisable in that it establishes a certain uniformity in conditions, the student usually being in her best physical condition at the time of testing. In addition, letters were sent to individual students in the class, reminding them of the examination, and an account, written by Professor Hollingworth, of the widespread use of similar tests by reliable business firms and their value in selecting candidates for positions along various lines, appeared in the college weekly. A similar notice of the tests was posted in Freshman Study in the fall of 1916, and in the fall of 1917. Letters were also sent to individual students at these times.

The subjects, as indicated, were Barnard students in their freshman year. The fact that they had had no training in experimental psychology, and were unfamiliar with the tests employed, made them a suitable group for testing. Out of a class of about one hundred and forty freshmen during 1915-16, one hundred were tested. This constitutes our first group of subjects whom we will designate as Group I. During the year 1916-17 (class of 1920),

eighty-five freshmen were tested, and in the fall of 1917 fifteen more (class of 1921) were given the tests. These last two groups together constitute our second group of one hundred freshmen whom we will designate Group II. In addition, in order to determine the reliability of the tests, the series was divided into two equivalent parts in a manner to be described later. In the spring of 1919, during the period extending from March 14 to May 15, forty-five freshmen from the class of 1922 were tested twice on the same day, each test requiring forty-five minutes of the student's time.

All the tests were given individually. This enabled the experimenter to supervise personally the performance of each subject and to stop her at any indication that she did not fully understand the directions given. It was likewise an important factor in contributing to the standardization of the conditions of the experiment. The subject was by this means freed from any feelings of irritation or discouragement that might have arisen if she had taken the test with a group of students whom she knew to be more rapid workers than herself. In such a case the knowledge that others were accomplishing their work in a shorter period of time would operate to arouse in some subjects such feelings of the futility of competing with their companions that their resulting performance would have been much slower than would have been the case where the tests were taken under more favorable conditions. Each freshman, then, was examined individually, and every effort was exercised to make the conditions of the experiment as uniform as possible. The room employed for the testing was one regularly used by the Department of Psychology for advanced experimental work, and from the point of view of light and ventilation it is well adapted for research. Except during the tapping and coordination tests, the subject sat at a small laboratory table, opposite the experimenter. As the room was so situated as to be almost unaffected by sounds from neighboring rooms, and was itself kept in a quiet condition, there was nothing to distract the subject's attention from her work.

As previously indicated, attempt to secure uniformity in administering the tests was also made. Besides giving the tests individually, the order in which the tests are listed was followed. In a few cases circumstances rendered it necessary to deviate slightly from this order, but in general it was followed rigidly. The result of the preliminary trial had been to indicate the most satisfactory manner in which the tests should be administered. The aim was to make the directions as clear, simple, and direct as possible. As a detailed

account of the instructions given for each test will be considered in the next section, it is only necessary to mention here that the method of procedure agreed upon was carefully followed with one or two exceptions where misinterpretation of the directions resulted in the experimenter's repeating the instructions in a slightly different form.

SECTION IV

DISCUSSION OF THE TESTS, INCLUDING MATERIALS USED, METHODS OF PROCEDURE, AND RESULTS

Test No. 1. Coordination

This test, popularly termed the "three-hole test" calls for both speed and accuracy of movement and gives an indication of the subject's motor ability and coordination.

Apparatus: An oak plate tilted at an angle of 45 degrees to the base board, containing three brass-line holes arranged in the form of an equilateral triangle, about 8 cm. apart. Contact of the metal rod with the bottom of the hole makes an electrical connection recorded by the automatic counter. Stop watch.

Instructions: "I want you to hold this (stylus) in your right hand and to touch the bottom of each one of these targets *as quickly as possible*, going around in a circle without skipping any of the holes. You see every time you do so, the contact is registered on the electric counter. I want to see how many contacts you can make in one minute. You start then when I say, 'Go' and stop when I say, 'Stop.'"

Method of scoring: The score represents the number of contacts made in one minute.

Results: The average, standard deviation, and range for groups I and II (200 freshmen in all), is indicated in Table I below:

TABLE I

Test No. 1 Coordination	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	82.7	10.77	63.8	109.0
Group II	84.1	11.92	60.8	110.4

Test No. 2. Tapping

This test has been widely used as a test of motor speed and endurance and has been considered by some experimenters to afford the best index of motor capacity.

Apparatus: Tapping board with metal plate and electric counter.

Tapping stylus with flexible connecting wire attached. Two dry cells. Stop watch.

Instructions: "I want you to hold this (stylus) in your right hand and tap on here (indicating the brass plate) *as quickly as possible*. I want to see how many times you can tap in a minute. Start when I say 'Go' and stop when I say 'Stop.'" These instructions were accompanied by an illustration of tapping by the experimenter. For this test the subject sat directly in front of the tapping board, resting her arm on the table, and assumed the position most convenient for her.

Method of scoring: The score represents the number of taps made in one minute.

Results: Table II shows the results obtained in this test:

TABLE II

Test No. 2 Tapping	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	376.26	51.69	263.2	499.0
Group II	368.54	39.32	283.0	451.4

Test No. 3. Cancellation

This test is well adapted for measuring concentration and alertness of attention, maximum effort being required to accomplish the task quickly and accurately. In addition to involving such factors as "speed of perception" and "discrimination" it is partly dependent upon the subject's muscular reaction to stimuli presented. Owing to the fact, previously mentioned, that it was necessary to complete all the tests in one hour, it was found advisable to limit some of the tests. Inasmuch as we desired to include the Checking Test which involves functions similar to those involved in Cancellation and as it was believed that these two tests together would exert an unfavorable influence upon the results of following tests due to the eye-strain they would cause, it was deemed advisable to use only one half of the Cancellation blank and one half of the Checking blank. The halves of these blanks have been found by Woodworth and Wells to be equal in difficulty and they suggest that one half of the blank in the case of both these tests is a sufficient test. Thus we were able to avoid undue eye-strain and were further able to spend the extra time, saved from halving these two tests, in lengthening three of the Association tests.

Materials: Woodworth-Wells number blank, Form A.³⁷ Stop watch. A pencil was used for checking.

Instructions: After placing the blank on the table before the subject, face downwards, the following instructions were given: "When I say 'Go' I want you to turn over this sheet of paper, and cross out all the 3's, *as quickly as possible*, going across the paper like this (illustrating). There are five 3's on every cross line so you want to be sure to cross out all those on the first line before passing to the second line. Start when I say 'Go.'"

Method of scoring: The time taken to complete the cancellation was the score. Errors were very rare and were therefore entirely disregarded.

Results: Table III indicates the performance in this test.

TABLE III

Test No. 3 Cancellation	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	76.51 sec.	17.51	128.28	52.12
Group II	76.77 sec.	13.82	105.60	50.76

Test No. 4. Checking

This test measures functions similar to those employed in the Cancellation test, although here the functions involved are more complex. To quote Woodworth and Wells, "The detection of a pair of digits in a group is a specialized performance, not reducible to the acts of detecting the single digits. The difficulty of this test is mainly perceptual and the overlapping which is effective in finding pairs of digits must occur in the perceptive process."³⁸ Inasmuch as Professor Woodworth found the first half of his number blank, Form B, to be equal in difficulty to the second half, for the reason mentioned under "Cancellation" only one half of this blank was employed.

Materials: Woodworth-Wells' number blank, Form B. Stop watch. Pencil.

Method of procedure: As in the Cancellation Test, the blank was placed before the subject, face downwards, and the following instructions were given: "When I say 'Go' I want you to turn this

³⁷ Woodworth, R. S., and Wells, F. L. Association Tests. Psychological Monograph, No. 57, 1911, p. 24.

³⁸ Woodworth, R. S., and Wells, F. L., Op. cit.

paper over and check any way at all, *as quickly as possible*, all the numbers that contain both a 9 and a 6. Start when I say 'Go.'"

Method of scoring: The total number of checks to be made was 35. Therefore the score was obtained by dividing the time taken by the subject by the number of correct checks made and then multiplying by 35. No account was taken of wrong checks made as it was believed that the time spent in making them sufficiently penalized the subject.

Results: Table IV shows the performance attained in this test.

TABLE IV

Test No. 4 Checking	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	102.93 sec.	19.64	152.28	72.6
Group II	105.98	20.45	161.0	76.86

Test No. 5. Color Naming

"This is a test of discrimination-reaction, involving prompt decision and correct reaction to a situation."

Materials: Woodworth-Wells' Color Naming blank.³⁹ Stop watch.

Method of procedure: Preliminary to the actual test the blank was placed before the subject with only the sample line of five colors showing. The subject was then asked to give the names of each color. Then the following directions were given: "I want you to name all these colors for me, *as quickly as possible*, going across the paper, from left to right, as in reading. Start when I say 'Go.'"

Method of scoring: The score was the time taken by the subject to complete the entire series of 100 reactions.

Results: The results are shown in Table V.

TABLE V

Test No. 5 Color Naming	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	56.01 sec.	8.75	78.84 sec.	41.16 sec.
Group II	58.55 sec.	9.36	81.32 sec.	39.0 sec.

³⁹ Op. cit.

Test No. 6. Directions

This test measures the subject's speed in apprehension and her general intelligence.

Materials: Woodworth-Wells' Hard Directions blank. Stop watch.

Instructions: "When I say 'Go' I want you to turn this blank over and follow directions—do just what the directions say, *as quickly as possible.*"

Method of scoring: The score is the time in seconds required to complete the test. Errors were counted separately.

Results: Table VI indicates the performance in this test.

TABLE VI

Test No. 6 Directions	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	126.15	52.00	296.6	64.08
	sec.		sec.	sec.
Group II	119.76	41.65	243.2	61.6

Test No. 7. Opposites

For a test which would indicate a general tendency or "adjustment to react according to instructions" and also measure the quickness and accuracy of association of ideas, the two equal lists of opposites proposed by Woodworth and Wells were combined into one list. Our reason for combining the lists was in order to get a real measure of the individual's ability to name opposites. If we had taken only the short list we would have obtained an adequate measure of the subject's alertness of attention and ability to adapt herself to a situation, but we desired to go further than this and find out whether the individual really had any special ability for naming opposites. This test also indicates facility in handling words and is generally considered to have a high correlation with general intelligence.

Materials: Woodworth-Wells' Lists of Opposites printed on cardboard. Stop watch.

Method of procedure: These instructions were given: "I want you to name the opposite for each one of these words (showing card with lists, at a distance) *as quickly as possible*, not repeating the words themselves but just naming the opposite. For instance, if the word were 'tall,' you would say 'short.' Be sure you give the

exact opposite of each word before proceeding to the next. Do you understand?"

The subject was stopped if a wrong opposite was given and not permitted to proceed with the other words until the right opposite was given.

Method of scoring: As no errors were permitted to be made in the test, the score represents the time taken for completing the task.

Results: Table VII indicates the results obtained in this test.

TABLE VII

Test No. 7 Opposites	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	51.08	10.33	79.00 sec.	34.84 sec.
Group II	50.88	8.55	71.52	35.92

Test No. 8. Verb-object

This is also one of the association tests and measures ability to handle verbal relations. As in the Opposites Test we combined the two equivalent lists of verbs proposed by Woodworth and Wells into one test. Desire to obtain a real measure of the subject's innate ability to name objects was the reason for lengthening this test.

Materials: Two equal lists of verbs combined into one list and printed on cardboard. Stop watch.

Method of procedure: These instructions were given: "In this case I want you to name an object for each one of these verbs, *as quickly as possible, not repeating the verbs themselves* but simply naming the objects. For instance, if the verb were 'bake,' you would say 'bread' or 'cookie.' Do you understand?"

Method of scoring: As no errors were permitted to be made, the score presents the time required to complete the test.

Results: The results are indicated in Table VIII.

TABLE VIII

Test No. 8 Verb-object	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	65.55 sec.	12.32	99.56	45.48
Group II	67.35 sec.	12.91	99.08	47.24

Test No. 9. Mixed Relations or Analogies

This test measures facility in handling associations, and ability to perceive relationships among logical material. As in the two preceding Association Tests the two equal lists proposed by Woodworth and Wells ("Eye: see = Ear——; Oyster: shell = Banana:——" and "Good: bad = Long:——; Man: woman = Boy:——") were combined into one long list for a reason similar to that which led us to lengthen the Verb-object and Opposites tests.

Materials: Combination of Woodworth-Wells' two equal lists for Mixed Relations test, printed on cardboard. Stop watch.

Method of procedure: The subject was shown sample analogies and the following instructions given: "In this case there are three words given and you are to supply a fourth word that has the same relation to the third word as the second word has to the first. For example, in this case, 'Box: square = Orange:——,' square gives the shape of the box. Then the shape of an orange is round, so you would supply 'round' as the fourth term. (Two other illustrations were then given.) The relations involved won't always be the same; it may be the case of shape, or opposites, etc. But you look at the first pair of terms in every case and then make the second pair express the same relationship as the first pair. Do you understand?"

Method of scoring: As no mistakes were allowed, the score is the time required to complete the test.

Results: The results are shown in Table IX.

TABLE IX

Test No. 9 Mixed Relations	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	139.64 sec.	42.97	266.6 sec.	82.88 sec.
Group II	131.66 sec.	32.97	227.2 sec.	79.56 sec.

Test No. 10. Word Building

For a test that would indicate ingenuity and skill in the manipulation of letters and give a measure of the subject's command of vocabulary, the word building test was used. The number of words written in a given time depends in part on whether the subject

proceeds with a definite plan, combining, for example, "a" with all the other letters, then "e" with all the other letters, etc., or goes about the task in a vague or random fashion.

Materials: Sheet of paper at the top of which were written the letters *a e i l p r*.

Method of procedure: The procedure as given by Whipple⁴⁰ was followed with the exception that the time-limit was three minutes instead of five.

Method of scoring: The score represents the number of words written. A word was considered correct if it is included in Whipple's list of admitted words.

Results: Table X shows the results secured in this test.

TABLE X

Test No. 10 Word Building	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	16.33 words	4.93	6.0	27.2
Group II	16.23	4.52	6.4	24.6

Test No. 11. Word Naming

This uncontrolled association test appears to be a good test for determining individual differences, the subjects tending to write words belonging to various categories. Such differences as the tendency to write series of rhymed words, to write a series of words that are grouped about one central idea, then to write another series of words grouped about a second central idea, suggested perhaps by the last word in the first series, etc., are revealed in this test. It also depends in part on the subject's speed of writing.

Materials: Stop watch. Sheet of paper and pencil.

Instructions as follows were given: "I am going to give you three minutes in which to write all the words you can. It makes no difference what sort of words they are—they can be anything you want to write."

Method of scoring: The score equals the number of words written.

Results: Table XI shows the results for this test.

Test 12. Knox Cube

This test gives an indication of the subject's power of observation, memory, and ability to concentrate her attention. It involves

⁴⁰ Whipple, G. M. *Manual of Mental and Physical Tests*. Part II, p. 275.

the ability to handle concrete objects and to imitate another's performance with accuracy.

TABLE XI

Test No. 11 Word Naming	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	67.14 words	12.78	40.8 words	94.2 words
Group II	67.87	11.86	45.0	93.0

Materials: Five one-inch cubes.

Method of procedure: Pintner's standardization of the Knox test was followed. Care was exercised to execute all movements slowly and deliberately and at a uniform rate.

Method of scoring: The score represents the number of lines correctly imitated.

Results: Results are indicated in Table XII.

TABLE XII

Test No. 12 Knox Cube	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	9.20 lines	1.56	5.8	11.4
Group II	8.82 lines	1.64	4.8	12.0

Test No. 13. Digit Span

To measure ability to reproduce with accuracy disconnected and non-logical material, the digit span test was employed. It tests the subject's power to concentrate her attention upon the series of digits as they are read aloud to her by the experimenter and to so retain said series in her mind that she may reproduce it with absolute accuracy immediately after the experimenter has ceased speaking. It affords an opportunity also to observe individual differences.

Materials: Digit Span blank. Stop watch.

Method of procedure: These instructions were given: "I am going to read some numbers to you and as soon as I have finished saying them, I want you to repeat them in exactly the same order." The smallest number of digits given was five. Three trials were

given for each number. The attempt was made to repeat the numbers without rhythm.

Method of scoring: The score represents the highest number of digits correctly repeated two trials out of three.

Results: Table XIII indicates the results of this test.

TABLE XIII

Test No. 13 Digit Span	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	7.39 digits	1.31	5 digits	10.2 digits
Group II	7.67 "	1.29	5.2 "	10.2 "

Test No. 14. Word Memory

Test No. 15. Logical Memory

Both of these tests call into play functions similar to those demanded in the digit span test. However, here the material to be reproduced has meaning, consisting in Test 14 of a series of concrete words and in Test 15 of a list of familiar proverbs.

Materials: Cards containing a list of 25 words and a list of 25 proverbs, respectively. Also two blanks containing 50 words and 50 proverbs, respectively. The cards and blanks were those employed by Edith Mulhall Achilles.⁴¹

Method of procedure: Instructions were given as follows: "I am going to let you look at a list of words (or proverbs as the case might be) for *one minute*, after which I am going to ask you to write as many of the words (or proverbs) as you remember." The subject was allowed one minute in which to write down the words she remembered and two minutes to write the proverbs. After recording the words remembered the subject was given a second list in which there were 25 words previously seen and 25 new words, and was asked to mark "y" all the words she recognized as having seen before and "n" those she thought she had not seen. Similar procedure was followed for the test with proverbs.

Method of scoring: For Recall the number of words or proverbs written constitutes the score. No account was taken of the order in which they were recalled, or any false recollections recorded.

In scoring *Recognition* this formula was employed to derive the score:

⁴¹ Achilles, Edith Mulhall. Archives of Psychology, No. 44, 1920.

50 (which is the total number of words or proverbs) minus 2 x number of errors = score.

Results: Tables XIV and XV indicate the results of these tests.

TABLE XIV

Test No. 14 Word Memory— Recollection	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	11.59 words	2.70	6.6 words	17.4 words
Group II	10.91 "	2.79	6.2 "	18.0 "
Word Memory—Recognition				
Group I	35.84 "	7.44	20.0 "	47.2 "
Group II	35.07 "	8.33	14.8 "	48.4 "

TABLE XV

Test No. 15 Logical Memory— Recollection	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	6.19	1.74	3.0	9.6
Group II	6.50 proverbs	1.76	3.2 proverbs	9.8 proverbs
Logical Memory—Recognition				
Group I	36.75	8.95	17.2	47.6
Group II	37.47	7.69	18.4	48.4

Test No. 16. Substitution

For a test which would measure speed of learning new associations the Substitution test was employed. In this test a key is constantly referred to and as the test proceeds it is gradually learned, the subject depending less and less upon it. Comparison between the time taken to complete the first and second halves of the blank gives a measure of the amount of time saved from learning the key.

Materials: Substitution test blank. The blank with 5 geometrical forms was used. Stop watch.

Method of procedure: The key was explained to the subject and then the blank was placed face downwards before her and she was instructed to turn over the Substitution blank at the signal "go" and to begin with the first form and take each one as it came, going across the paper from left to right, and to write the proper number in each form according to the key at the top.

Method of scoring: Three scores were taken, representing the time for the first half of the blank, the second half and the whole blank, respectively. Errors, being rare, were counted separately.

Results: The data for this test are found in Table XVI.

TABLE XVI

Test No. 16	Average	S. D.	Range	
			Poorest	Best
Substitution—1 Half	seconds		seconds	seconds
Group I	64.33	9.69	87.68	46.8
Group II	66.68	12.14	97.60	46.0
Substitution—2 Half				
Group I	59.10	11.62	86.2	37.0
Group II	61.51	13.15	91.8	38.4
Substitution—Whole				
Group I	123.09	19.61	167.72	86.48
Group II	128.19	23.89	187.0	87.40

Test No. 17. Completion

For measuring correctness and facility in the use of words, readiness in perceiving and comprehending situations and affording some indication of creative ability, the Completion test was employed. To quote Trabue, "On the whole it will be found that ability to complete these sentences successfully is very closely related to what is usually called 'Language ability.'"⁴²

Materials: Trabue Language Scal A. Stop watch.

Method of procedure: The standard procedure suggested by Trabue was followed, a time-limit of four minutes being employed.

Method of scoring: In general, the method was to follow Dr. Trabue's scoring; "A score of 2 being given each sentence if perfectly completed, a score of 1 if almost but not quite perfectly completed, and a score of 0 if not attempted at all or if imperfectly done." Total of 48 points is the maximum score attainable in Scale A.

Results: Table XVII represents the performance of the freshmen in this test.

TABLE XVII

Test No. 17 Completion	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	36.08	4.33	26.8	44.8
Group II	35.78	4.36	25.2	44.4

⁴² Trabue. Completion-Test Language Scales.

Test No. 18. Information

To measure range of information and obtain some conception of the number and kind of objects known and the degree to which they are known, the information test was used. It tests the individual's knowledge rather than her ability.

Material: The information test blank as specified in Whipple's Manual, containing 100 words and directions for marking them.

Method of procedure: The subject followed the directions at the top of the blank, marking each word with a certain letter which indicated the degree to which it was known to her. There was no time-limit in this test, the subject being allowed all the time she desired to finish the blank.

Method of scoring: The score represents the number of words marked "D," "E," "F," and "N," respectively. As no check was used in this test, the score probably shows over-estimation. The total score was obtained by assigning these values: D = 3; E = 2; F = 1; and N = 0, and taking their sum.

Results: The table following indicates the results of this test.

TABLE XVIII

Test No. 18	Average	S. D.	Range	
			Poorest	Best
Information D	21.47 words	9.71	3.6 words	41.6 words
Information E	13.70 "	6.16	3 "	28 "
Information F	14.81 "	6.43	1.8 "	26.2 "
Information N	50.01 "	10.35	69.6 "	29 "
Total Score:				
Group I	106.63	25.51	59.8	158.2
Total Score:				
Group II	104.71	26.79	55.4	161.8

Test No. 19. Vocabulary

This test merely indicates the number of words in the individual's vocabulary.

Materials: Vocabulary test blank as specified in Whipple's Manual.⁴³

Method of procedure: The subject was asked to follow the directions given at the top of the test blank and to mark the words carefully according to the directions.

⁴³ Op. cit. Vol. 2, p. 310.

Method of scoring: The score represents the number of words marked plus (+). This number indicates the vocabulary-index; the index taken as a per cent. is multiplied into 28,000.

Results: Table XIX shows the results for this test.

TABLE XIX

Test No. 19 Vocabulary	Average	S. D.	Range	
			Poorest (Av. of lowest 5)	Best (Av. of best 5)
Group I	74.81 words	6.86	59.6	86.6
Group II	73.90 "	7.60	59.4	87.4

SECTION V

NORMS OF PERFORMANCE AND THEIR PRACTICAL APPLICATION

To summarize the results of the preceding section, Table XX shows the norms of performance for the two hundred Barnard freshmen (Groups I and II), in all the various tests. The average, probable error, and range from the poorest to the best score are shown for each test. To avoid misrepresentation of facts by undue weight being given extreme cases, the average of the five poorest scores is in each case taken as the poorest score, and the average of the five best scores as the best score.

The following is a comparative table comparing our results with those of other investigators who have employed some of these tests with freshmen. Only those cases are considered where the tests are identical, and the method of scoring the same.

Test	Barnard Norm	Bingham	Kitson	Other Investigators	
Cancellation .	76.6 sec.	48.3 sec.	69.2 sec.		
Color Naming .	57.2 sec.	56.2 sec.			
Hard Directions	122.9 sec.		110.9 sec.	Washburn,	153 sec.
Opposites . .	50.9 sec.		52.6 sec.		
Word Building .	16.2 words		21.4 words	Sunne,	18
Digit Span . .	7.53 digits	7 digits	8.4 digits	Cattell,	7.6
Information .	20.4 words			Waugh,	24
				King &	
				M'Crory,	25
				Smith,	10.9

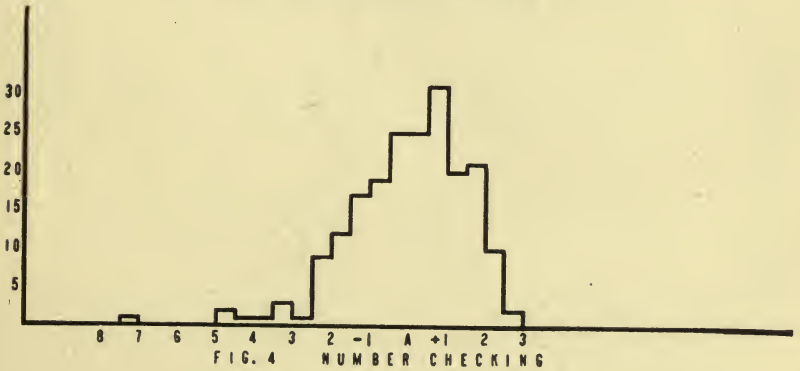
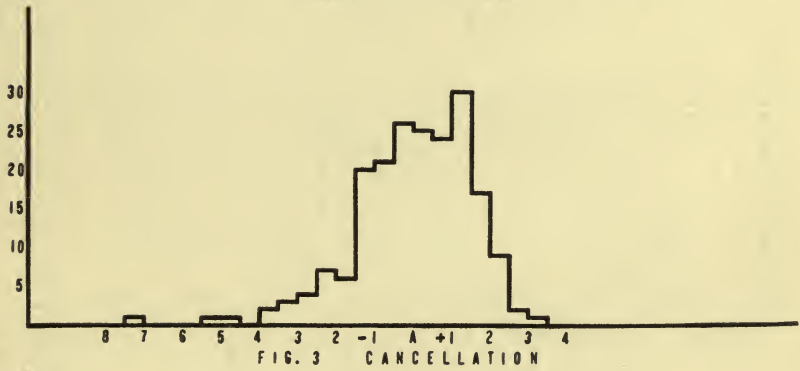
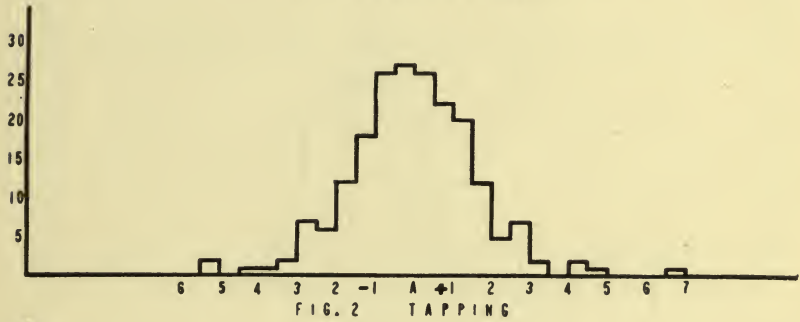
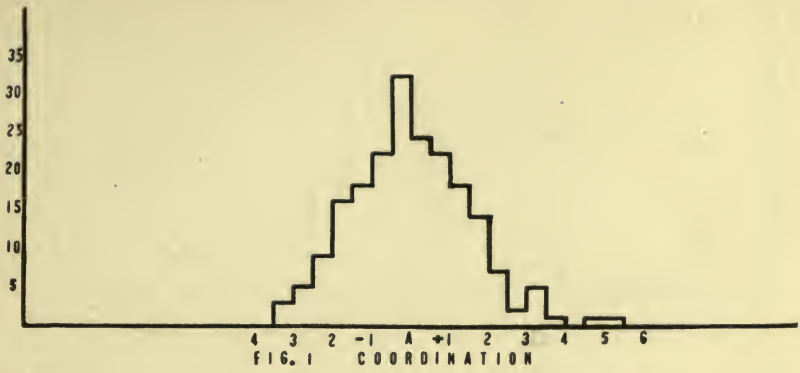
Figures 1 to 23 inclusive, show graphically the dispersion of measures about the average in the case of the Barnard freshmen. To secure uniformity and facilitate comparison, the charts are constructed with the average in each case as the mid-point and the scores expressed in terms of P.E. units from the average as a center. The P.E. was taken as the unit because it is a convenient and familiar measure. The vertical scale is also kept constant except in three tests where it is changed for reasons to be specified later. Inspection of these figures reveals many interesting features.

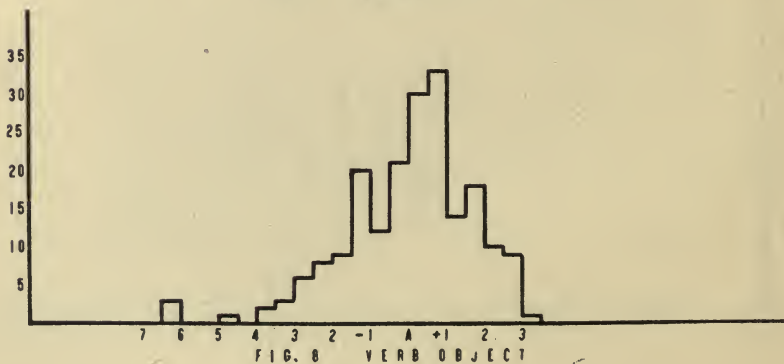
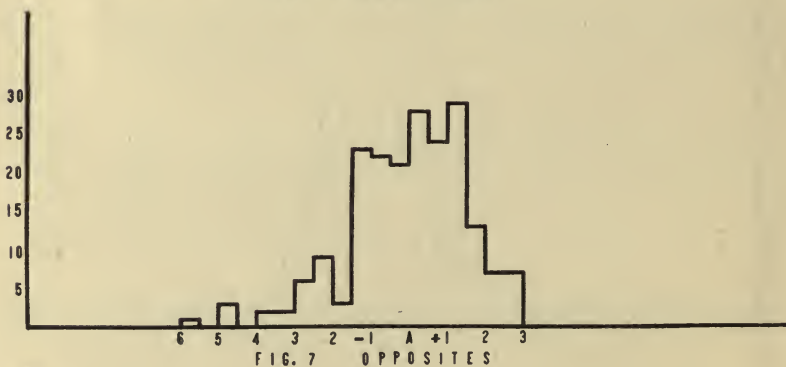
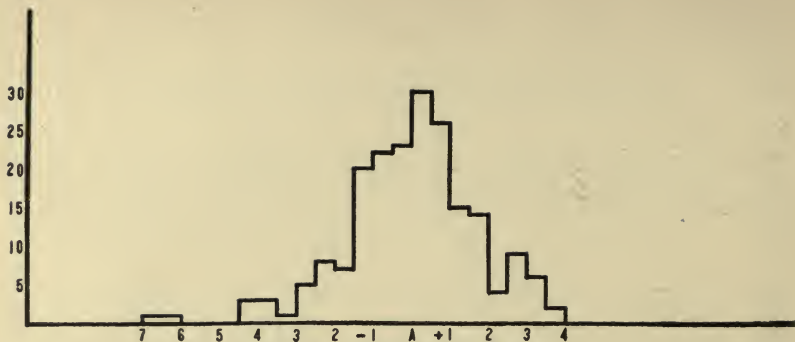
We may divide the tests roughly into five groups.⁴⁴ The first group contains the two motor tests—Coordination and Tapping.

⁴⁴ Justification of this division of the tests will be given in Chapter VI.

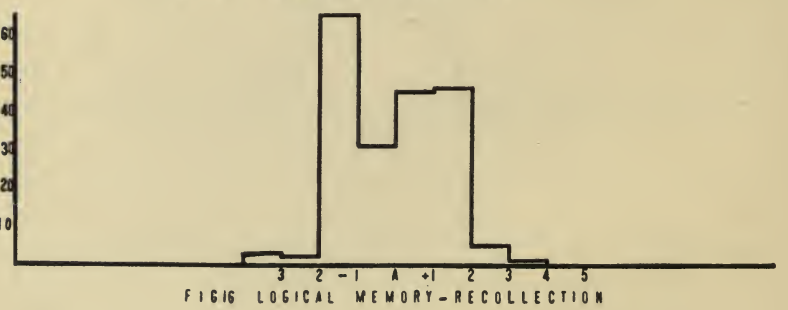
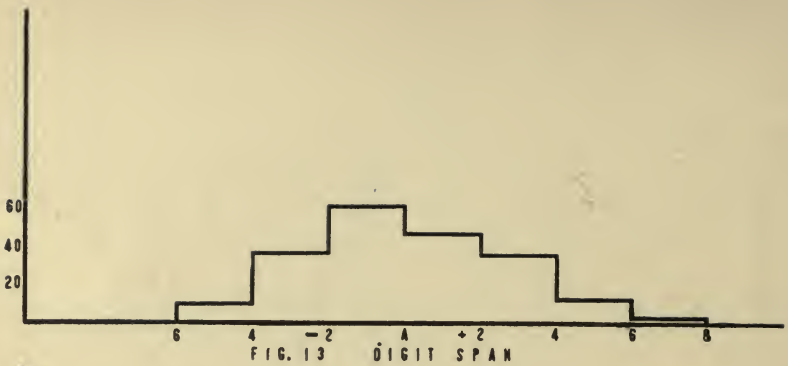
TABLE XX
NORMS AND STANDARDS OF PERFORMANCE (200 Cases)

Name of Test	Average Group I	Average Group II	Average Groups I and II	Probable Error	Range	
					Poorest; (Average of poorest 5)	Best (Average of best 5)
1. Coordination	82.7	84.1	83.42 strokes ¹	7.5	60.00	114.20
2. Tapping	376.26	368.54	372.4 taps	27.6	251.40	501.20
3. Cancellation	76.51	76.77	76.64 sec.	10.35	128.28	48.56
4. Checking	102.93	105.98	104.46	13.10	169.16	72.08
5. Color Naming	56.01	58.35	57.28	5.28	84.72	38.76
6. Directions	126.15	119.76	122.95	24.95	312.00	59.16
7. Opposites	51.08	50.88	50.98	6.01	79.80	33.92
8. Verb-object	65.55	67.35	66.45	7.54	107.80	44.96
9. Mixed Relations	139.64	131.66	135.65	21.34	279.00	77.96
10. Word Building	16.33	16.23	16.28 words	2.71	4.80 words	27.60
11. Word Naming	67.14	67.87	67.50	7.50	39.40	97.20
12. Knox Cube	9.20	8.82	9.01	1.01	4.80	12.00
13. Digit Span	7.39	7.67	7.53	.53	5. digits	10.40
14. Word Memory—Recollection	11.59	10.91	11.25 words	1.74	5.8	18.60
15. Word Memory—Recognition	35.84	35.07	35.45	5.45	13.60	48.80
16. Logical Memory—Recollection	6.19	6.50	6.34	1.34	2.20	10.20
17. Logical Memory—Recognition	36.75	37.47	37.11	5.11	13.20	48.40
18. Substitution 1st Half	64.33	66.68	65.51	7.51	101.28 sec.	43.80
19. Substitution 2nd Half	59.10	61.51	60.32	8.12	98.40 sec.	35.20
20. Substitution Whole	123.09	128.19	125.64	14.40	196.60 sec.	83.20
21. Completion	36.08	35.78	35.93	2.93	23.80	46.00
22. Information	106.63	104.71	105.66	18.33	54.20	168.20
23. Vocabulary	74.81	73.90	74.35	5.35	57.40	89.00







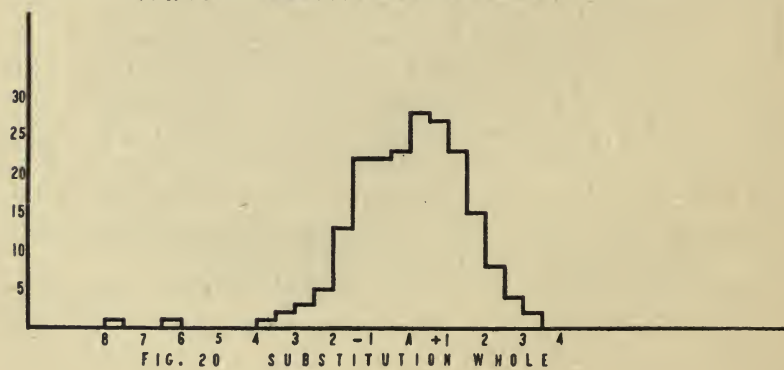
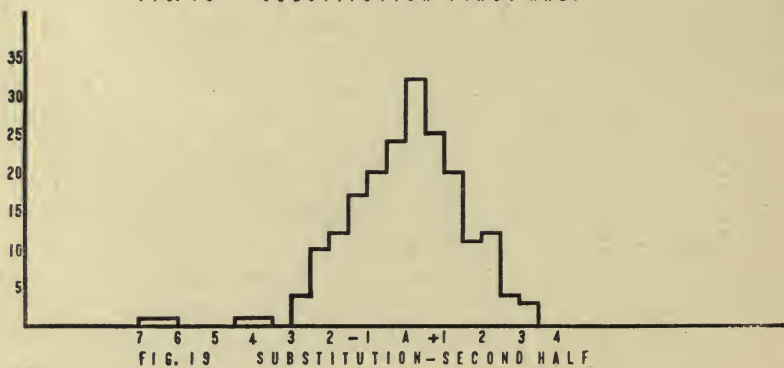
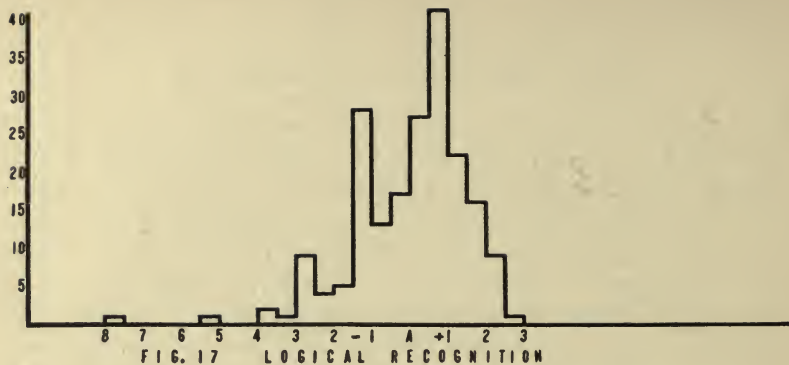


Here we have fairly uniform distributions. The actual range for Coordination is from $-3\frac{1}{2}$ P.E. to $+5\frac{1}{2}$ P.E. (skewed at the positive end), and for Tapping from $-5\frac{1}{2}$ P.E. to $+7$ P.E. But to take the actual range as the basis of our comparison is misleading. A clearer conception of the facts is obtained by noting the closeness with which the measures distribute themselves about the central tendency. In these two motor tests we find a fairly uniform distribution, suggesting that the tests are adequate for selecting good and poor subjects even in a group as homogeneous as college freshmen.

In the second group we may place those tests which involve powers of perception and comprehension, namely, cancellation, checking, color naming, word naming, and substitution. Here again we find a distribution approximating the normal curve of distribution. At first glance it would appear that in four of these tests the curves are skewed toward the negative or poor end. In both Fig. 3 and Fig. 4, (Cancellation and Number Checking), we find a case at $-7\frac{1}{2}$ P.E.; in Fig. 5 (Color Naming) we find one at -7 P.E.; and in Figures 18, 19, and 20 (Substitution), we find cases at -9 P.E.; -7 P.E., and -8 P.E.; while at the good end no case exceeds $+4$ P.E. We must take care, however, not to let these extreme cases mislead us as to the general character of the distribution. If we count up the cases on either side of the average we find 108 cases above the average in Cancellation, 109 in Number Checking, 106 in Color Naming, 107 in Substitution, and 98 in Word Naming. Thus we really have a more or less uniform distribution with a tendency of the number of scores above the average to exceed the number below it. Disregarding the few extreme cases, we find the majority of the scores contained within the normal limits of the P.E. distribution, (-4 P.E. to $+4$ P.E.).

In the third group we may place the tests involving associative relations, namely, Directions, Opposites, Verb-object, Mixed Relations, Word Building, and Completion. Here, likewise, as in the two preceding groups, we find fairly uniform distributions with a greater number of cases above than below the average, (except in Word Building, where the distribution is about equal). The majority of cases are likewise contained within the normal range of 8 P.E., but there are a few extreme cases at the poor end in Completion, Opposites, Verb-object, Mixed Relations, and an extreme case at both the good and bad end in the Word Building test.

The fourth group contains those tests which call into play powers



of learning, viz: observation and retention, namely: Word Memory, and Logical Memory.

A word of explanation is needed here regarding the construction of the chart for Logical Memory (Recollection). The categories into which the scores fall are so few that the finest grouping possible is in 1 P.E. units instead of $\frac{1}{2}$ P.E. units as in the other tests. As we said before, to secure uniformity we let the P.E. represent the same interval along the base line in all tests. Now, in order to keep the area of a given number of cases constant for all tests, it is necessary where we have scores in terms of 1 P.E. units to reduce the vertical scale proportionately. Therefore, we regard the measures as distributed evenly over the P.E. intervals and reduce the vertical scale one-half. In this test and in Word Recollection we find a greater number of cases below the average than above. The curve is skewed toward the poor end in Word Recollection, and toward the good end in Word Recognition and Logical Recognition.

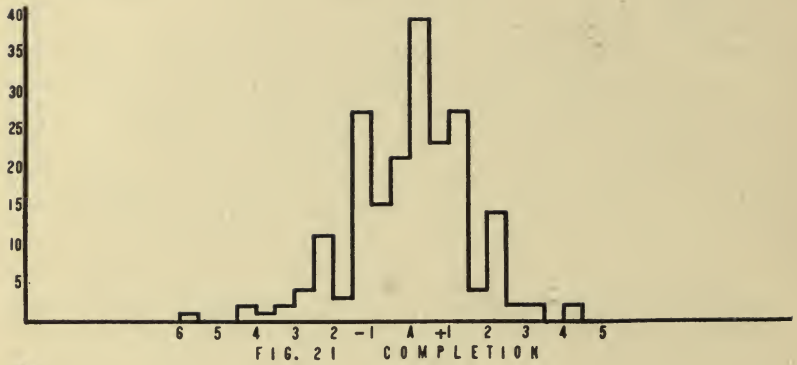
In our fifth group we have tests which depend on the subject's knowledge rather than her innate ability, namely, Information and Vocabulary. Here we find fairly uniform distributions with no extreme cases. This suggests the tendency of education to make a homogeneous group of individuals approach a general level of performance in a test of mere learning.

We have, finally, a miscellaneous group which comprises the Digit Span and Knox Cube tests—tests which showed both a low intercorrelation and low correlations with the other tests of the series. In the Knox Cube test the small number of categories makes it necessary to use 1 P.E. units and in the Digit Span test it is necessary to use 2 P.E. units.

To sum up then, these surfaces of distribution are fairly symmetrical, if we disregard the few extreme cases. In addition, the fact that the averages and surfaces of distribution for the first group of one hundred freshmen (Group I) are approximately the same as for the second group of one hundred (Group II), corroborates this conclusion and supports the view that the norms here presented are reliable.

ACADEMIC GRADES

Besides their score in the psychological tests we have additional information about the first group of one hundred freshmen (Group I) in the form of university grades and records taken in the gymnasium. The college subjects may be grouped into five classes:



1. Language (including English, Latin, Greek, German, French, Italian, and Spanish); 2. Mathematics; 3. Science (physics, chemistry, botany or geology); 4. Philosophy (including psychology); and 5. History. Due to the freedom allowed the students in making out their programs, the same subjects are not taken by all, and the number of cases in each class therefore varies. The letter system of marking is employed at Barnard, the letters A (excellent), B (good), C (fair), D (Poor), and F (failure), being used. For the statistical treatment of the data the letter grades were transformed into numbers according to the scale: A = 90, B = 80, C = 70, D = 60, and F = 50. Norms for these freshmen in their college work are shown in Table XXI.

TABLE XXI

Academic Record	Number of		P. E.	Range (Actual)	
	Cases	Average		Lowest	Highest
1. Language	97	75.31	4.69	50	90
2. Mathematics	88	76.99	6.99	50	90
3. Science	41	72.26	7.74	50	90
4. Philosophy	27	78.15	3.15	60	90
5. History	26	72.88	2.88	60	90

The averages tend to be approximately equal for all subjects with a nearly equal range of distribution.

PHYSICAL MEASUREMENTS

Table XXII gives averages, P.E.'s, and range from lowest to best score of the physical measurements taken in the gymnasium.

TABLE XXII

Test	Number of Cases	Average	P. E.	Range (Actual)	
				Poorest	Best
Height	97	159.92 cm.	4.08	137	172.9
Weight	97	120.59 lbs.	12.59	90	182
Lung Capacity	94	171.05 cu. cm.	13.50	118	230
Strength of Grip, r. h.	97	30.02 kg.	4.02	13	43
Strength of Grip, l. h.	97	27.27 kg.	4.27	16	38
Upper Back	97	20.60 kg.	3.4	12	42
Chest	97	19.60 kg.	2.6	11	36

One of the main purposes of this investigation, as we remarked in a preceding section, was to give the individual student a knowledge of her strengths and weaknesses. Accordingly, at the com-

Coordination 96
 Tapping 368
 Cancellation 60
 Checking 93.4
 Color Naming 55
 Directions 100
 Opposites 37
 Verb Object 43
 Mixed Relations 81
 Word Building 20
 Word Naming 75
 Completion 48
 Knox Cube 8
 Digit Span 10
 Word Recoll. 15
 Word Recog. 46
 Logical Recoll. 10
 Logical Recog. 44
 Substitution 104
 Information 118
 Vocabulary 75
 Language 90
 Mathematics 90
 Philosophy 90
 History 90
 Height 165.2
 Weight 120.5
 Lung Capacity 188
 Grip rt. hand 28
 Grip left 25
 Upper back 23
 Chest 22

96

25

50

38
37

53

61

67

67

71

73

73

75

79

81

80

80

83

87

86

85

90

96

93
91

94
98
95

98

99
100

100

CHART 1. A.M.

pletion of the entire series of examinations each year, an individual report was sent to each student who took the tests. This consisted of two blanks giving a description and interpretation of the various tests, with whatever significance each test was known to possess from a vocational standpoint. In addition to these explanatory blanks, there was a third blank which indicated the standing of the individual student in each of the tests, together with the average standing, (with the P.E.), in each test for the entire group of one hundred freshmen, so that the individual could compare her own record with that of the average in every case.

The ideal plan would have been for the experimenter, after sending each student her report, to have had a personal interview with her. In this she could have cleared up any difficulties the student might have had in interpreting her results and understanding their significance. She could also have rendered distinct aid by suggesting means whereby the student could make the best use of her abilities, or strengthen her weak points. Where the girl was doing academic work of a grade below the level her test record showed her capable of, the experimenter could have sought to determine the cause of the girl's academic failure—whether due to too many distractions, outside work, or what not—and given advice accordingly. Lack of time made it impossible to do this, however. We therefore have no record of these girls in their last three years of college to show whether they benefited from their test results. It is worth while at this point, nevertheless, to indicate how one may proceed to make practical use of these tests.

Charts 1 to 6, inclusive, represent the psychographic records of six students from Group I. They are constructed as follows: Reading along the heavy horizontal base line, we have the names of the nineteen psychological tests, (Substitution First Half and Substitution Second Half are omitted since ability in this test is adequately measured by Substitution Whole), the academic subjects varying from two to four, according to the programs of study, and seven physical measurements. Opposite the name of each test, subject, and physical measurement is the individual's score, and below this, the amount of her plus or minus deviation from the average scores expressed in P.E. units. To make the individual's relative standing more concrete, her score in P.E. units is also expressed in terms of what her position would be in a group of one hundred freshmen, selected at random.

The vertical line (reading up from the base line) is divided into

98

88

84

79

68

68

83

54

53

75

75

51

19

22

8

25

0

34

4

5

1

3

1

19

33

25

19

- Coordination 89
- Tapping 407
- Cancellation 61
- Checking 81
- Color naming 57
- Directions 157
- Opposites 87
- Verb Object 61
- Mixed Relat. 226
- Word Building 13
- Word Naming 89
- Completion 30
- Knox Cube 8
- Digit Span 5
- Word Recoll.
- Word Recog. 32
- Logical Recoll.
- Logical Recog. 38
- Substitution 104
- Information 58
- Vocabulary 61
- Language 55
- Science 50
- History 60
- Height 160.5
- Weight 103.5
- Lung Capacity 162
- Grip rt. hand 34
- Grip left 23
- Upper back 24
- Chest 16

0

0

0

Coordination 84
 Tapping 410
 Cancellation 65.4
 Checking 73.6
 Color Naming 49
 Directions 135
 Opposites 44.6
 Verb Object 63.6
 Mixed Relat. 108
 Word Building 16
 Word Naming 83
 Completion 37
 Knox Cube 10
 Digit Span 7
 Word Recoll. 8
 Word Recog. 16
 Logical. Recoll. 7
 Logical Recog. 36
 Substitution 104.8
 Information 76
 Vocabulary 62
 Language 71.67
 Mathematics 70
 Weight 172
 Weight 131
 Lung Capacity 208
 Grip rt. hand 43
 Grip left 38
 Upper back 31
 Chest 36
 History 70

52
38
25
11
1
15
6
31
25
43
48
59
59
62
74
70
76
80
82
85
81
76
74
70
94
92
98
96
98
95
98
100
25
25
25
25

CHART 3. G.S.

CHART 4. S.E.

Coordination	81				<u>42</u>
Tapping	312	<u>7</u>			
Cancellation	152.2				
Checking	129.5	<u>10</u>			
Color Naming	70	<u>6</u>			
Directions	99				<u>54</u>
Opposites	50				
Verb Object	60				<u>74</u>
Mixed Relat.	108				<u>80</u>
Word Building	7	<u>1</u>			
Word Naming	56		<u>16</u>		
Completion	37				<u>59</u>
Knox Cube	9				<u>50</u>
Digit Span	8				
Word Recoll.	9		<u>20</u>		<u>72</u>
Word Recog.	26		<u>13</u>		
Logical Recoll.	7				<u>62</u>
Logical Recog.	38				<u>54</u>
Substitution	149.4		<u>14</u>		
Information	101				<u>51</u>
Vocabulary	70				<u>30</u>
Language	74.65				<u>47</u>
Mathematics	60	<u>5</u>			
Science	60		<u>15</u>		
Height	153.5		<u>15</u>		
Weight	111				<u>31</u>
Lung Capacity	160				<u>30</u>
Grip rt. hand	26		<u>25</u>		
Grip left	25				<u>37</u>
Upper Back	23				
Chest	19				<u>44</u>

equal divisions, indicating position in a group of one hundred freshmen selected at random, using the norms of Table XX as the basis. No. 1 is considered the poorest individual in each case, No. 100 the best. The heavy horizontal black line in the center represents the average individual or the 50th individual in the group. To illustrate the use of these charts let us consider Chart I, A.M.'s record. In coordination this individual scores 96. Referring to Table XX, we see that the average freshman score for this test is 83.42 with a P.E. of 7.5. A.M.'s deviation from the average score is, therefore, $+ 12.58$ ($96 - 83.42$) $\div 7.5$ (the P.E.) or $+ 1.67$ P.E. units above the average. We know from the normal curve of distribution that between the average and $+ 1$ P.E. are found 25% of the cases, or 25 cases in a group of one hundred individuals. Between 1 P.E. and $+ 2$ P.E. there are approximately 17% more cases, or 17 in a group of one hundred individuals, so that if a girl made a score of $+ 2$ P.E. she would rank 50 (average) $+ 25 + 17$, or 92 in the group. A.M., however, does not quite reach this score. Her score reaches only .67 of the interval between $+ 1$ P.E. and $+ 2$ P.E., or .67 of the 17 cases contained within these limits. Now $.67 \times 17 = 11.39$, *i. e.*, A.M.'s score is that of the 11th individual in this group. This is only her approximate position, of course, since the scores are not distributed evenly over the interval. To secure her exact position we would transform her P.E. score into rank according to proper table.⁴⁵ She therefore stands $50 + 25 + 11$, or 86 in a group of one hundred freshmen in coordination. In Tapping her score is 368 taps. The average freshman score in this test is 372.4 taps with a P.E. of 27.6. A.M.'s deviation from the average, accordingly, is $- 4.4$ ($372.4 - 368$); her deviation in terms of P.E. is $- 4.4 \div 27.6$ (the P.E.), or she is $- .15$ P.E. units below the average. Her score therefore reaches .15 of the 25 cases in the interval between the average and $- 1$ P.E. Now, $.15 \times 25 = 3.75$. Her score therefore gives her a rank 3.75 or approximately 4 places below the average or 50th individual, *i. e.*, she stands 46 in a group of one hundred freshmen. A similar method was employed in finding out the psychographic records of the other five students. Considering the net scores in the psychological tests, A.M. ranked 97 in Group I, only three individuals surpassing her. When we group the tests under the five divisions suggested above, we see that although she would stand well above the average in a random group of one hundred freshmen, she makes her highest rank (88 average rank for

⁴⁵ Thorndike, E. L. *Mental and Social Measurements.*

93
CHART 5. L.J.H.

91
91

90

95

87

72

59

70

60

71

67

42

30

42

39

50

44

51

41

37

5

7
5

3

12

22

20

25

17

Coordination 81
 Tapping 397
 Cancellation 85.2
 Checking 136.8
 Color Naming 55.4
 Directions 177.8
 Opposites 66
 Verb Object 69
 Mixed Relat. 118.4
 Word Building 13
 Word Naming 64
 Completion 42
 Knox Cube 9
 Digit Span 6
 Word Recoll. 9
 Word Recog. 34
 Logical Recoll. 4
 Logical Recog. 32
 Substitution 135
 Information 113
 Vocabulary 79
 Language 88.33
 Mathematics 90
 Philosophy 90
 Height 153.9
 Weight 121
 Lung Capacity 166
 Grip rt. hand 40
 Grip left 25
 Upper back 23
 Chest 24

this group) in the group of tests which involve the association processes, *i. e.*, in those tests involving more complex and higher abilities. Moreover, she made the highest standing in academic marks of any freshman in Group I, being the only one to secure grade A in all the subjects she pursued during the year. It is of interest to note also that the subject's score in physical measurements is above the average. The tests therefore give an adequate measure of this student's ability.

Chart 2. L.H.C. This freshman presents the other extreme of ability. With the lowest academic standing of Group I, (having no mark higher than D grade), she also ranks only 26 in net test score. She is especially deficient in the association tests. In a random group of one hundred freshmen she would rank only 1 in Opposites and Mixed Relations, showing poor powers of associating ideas and perceiving relationships among logical material, and 8 in Completion, which measures readiness in perceiving and comprehending situations. She is also poor in the memory tests. In the second group of tests which involves ability to perceive what is wanted and to carry out simple instructions, she ranks above the average, suggesting that she would do well at simple types of clerical or stenographic work, though she lacks ability to perform work requiring a higher level of intelligence. In Information and Vocabulary her low rank of 4.5 is what we would expect. Having no aptitude for study, it is only natural that she should be uninterested in it. Her physical report was also below average. All indications confirmed her psychological report that she was unfitted to pursue college work. Her failure to meet the academic standard set for freshmen necessitated her withdrawal from college at the end of the year—a course justified by her psychological record.

Chart 3. G.S. Although in academic work this individual ranked only 21 in the group of one hundred, her net score in the psychological tests gave her a rank of 74. Her record in Group 3, *i. e.*, in the tests requiring the highest mental abilities, indicated that she was doing work of a grade far below her ability. Her net score in the tests of Group 4 suggested, and her record in the Information and Vocabulary tests, which depend chiefly on knowledge acquired, corroborated the hypothesis that she was neglecting her college work. In her case interest in athletics furnished the explanation for her college record. Not only was her physical record the highest in the class, but G.S. was a prominent figure in all college athletic events, especially in the swimming meets and in basket-ball games.

Coordination 78
 Tapping 394
 Cancellation 59.2
 Checking 96.6
 Color Naming 43.8
 Directions 94.4
 Opposites 37.6
 Verb Object 54.8
 Mixed Relat. 116
 Word Building 14
 Word Naming 58
 Completion 43
 Knox Cube 9
 Digit Sparr. 9
 Word Recoll. 9
 Word. Recog. 20
 Logical Recoll. 4
 Logical Recog. 26
 Substitution 89.2
 Information 85
 Vocabulary 74
 Language 71.25
 Mathematics 73.33
 Height 167.2
 Weight 122
 Lung Capacity 220
 Grip rt. hand 32
 Grip left 25
 Upper back 25
 Chest 20

32

29

21

20

3

12

7

23

29

37

37

69

65

70

87

77

84

95

93

94

97

95

50

49

53

62

54

88

99

80

CHART 6. M.M.

Chart 4. I.E. This case parallels L.C.'s. I.E.'s net academic rank was only 3 and her rank in the tests was also below freshman standard. Like L.C., also, I.E.'s withdrawal from college at the end of her first year was fully justified.

Chart 5. L.J.H. Here we have a case of a girl with a physical record above the average, and a rank of 95 in academic standing, but whose net score in the psychological tests is only 17. Having no other information about this girl besides the test data and her school marks, we cannot definitely explain this case. In only six of the tests does she rank above average, but two of these—Mixed Relations and Completion—involve the most complex mental functions, powers of understanding, and reasoning. It may be that, lacking powers of immediate recall, this girl was willing to devote long hours to grasping the subject matter of her studies so that by extra effort she was able to make high grades. Her score in Information and Vocabulary also suggests her attention to her studies.

Chart 6. M.M. This case presents the other extreme. Here we have a freshman who is in fine physical condition and has a net score of 77 in the psychological tests, but whose net academic standing is only 26. Inasmuch as she stands well above the average in all the tests involving the higher mental processes, her academic failure is probably due to lack of interest in her studies, or to too many outside activities.

SECTION VI

INTER-TEST CORRELATIONS AND THEIR SIGNIFICANCE

The psychographic charts showed that a freshman rarely did equally well in all the psychological tests. Whereas she tended to make approximately the same standing in all her academic subjects, she showed no such uniformity in the psychological tests. There were, of course, a few extreme cases where a good student scored above average in the majority of the tests, (for example, A.M.), or a poor student scored below average, (for example, L.H.C.). This raises the interesting question: Just what is the nature of the relationship existing between these tests? Are some more closely related than others? Is there any evidence to support our division of the tests into the groups suggested in the preceding section? For determining the relationship between the tests the particular method of correlation used in this investigation was one suggested by Professor Woodworth for combining the results of several tests.⁴⁶ By the use of his method it is possible to assign each individual her position in the distribution of the group; she stands, in other words, "above or below the group average and so and so much above or below as compared with the average variation of the group." The method of procedure is as follows: The average of any test is regarded as zero, and the individual's standing is expressed as a deviation above or below the average. Then the measure of variability (in this case the S.D.) is taken as the unit of deviation from this zero, and all deviations are expressed as fractions or multiples of the unit. Each individual deviation, then divided by the S.D. of the series, gives a resulting quotient called the "reduced measure." Having obtained the reduced measures, by appropriate substitution in the Pearson formula for correlation, we may easily obtain the correlation of two given tests A. and B., for, given the reduced measures of two arrays, the coefficient of correlation between them is the average of the products of the various reduced measures. The advantage of using this method is that the net position of an individual in a group of tests, for example, in the twenty-three tests

⁴⁶ Woodworth, R. S. Combining the Results of Several Tests; A Study in Statistical Method. From *Psychological Review*, March, 1912.

here used, may be easily obtained by dividing the sum of her reduced measures in those tests by the number of tests, (twenty-three in this instance).

Table XXIII gives the inter-test correlations computed according to this method. The test records used in obtaining these correlations are those of the one hundred freshmen of Group I. Inspection of this table reveals many interesting features. The correlations range from + .77 (between Cancellation and Digit Span) to .00 (between Tapping and Word Recollection, and between Mixed Relations and Word Recollection). The highest correlations are + .77 (between Cancellation and Digit Span); + .58 (Word Recollection and Word Recognition); + .57 (Opposites and Mixed Relations); + .56 (Logical Recollection and Logical Recognition); + .51 (Cancellation and Checking); + .48 (Coordination and Tapping); + .48 (Mixed Relations and Completion); + .44 (Opposites and Verb-object); and + .40 (Cancellation and Word Naming). That the Cancellation test furnishes the highest single correlation is interesting because it contradicts the old compensation theory and McCall's finding of a negative correlation ($- .28$) between this and the Trabue Completion test. All our correlations with Cancellation are positive, ranging from + .03 to + .77. Especially noteworthy are the correlations of + .40 with Word Naming, + .30 with Word Building, and + .31 with Substitution—all tests calling into play the higher thought processes. The fact that the correlations are all positive is suggestive of a definite relationship between cancellation and these various tests.

Checking and Word Naming show the highest average correlation (+ .25) with the other tests (omitting Information, Vocabulary, Word Recollection, and Word Recognition). Then, in order, Opposites, Verb-object, and Cancellation; Color Naming, Directions, Mixed Relations, Word Building, and Completion; then, Logical Recollection and Substitution Whole; Knox; Tapping, and Digit Span; Coordination; Logical Recognition. The Information and Vocabulary tests were omitted because they showed no correlation with the other tests. The Vocabulary test has an average correlation with the other tests of .00, indicating chance relationship. The correlations of Information with the other tests were not worked out because inspection of the scores showed that approximately the same result would be obtained as for the Vocabulary test.

On the whole, the inter-test correlations, although mostly posi-

tive, are low. This would indicate that we are testing here different mental abilities. The fact that we can group certain tests together on the basis of relationship shown by the correlation coefficients further supports this view. It is possible to find several groups of tests which correlate closely among themselves, but loosely with the other tests. The following table gives the various groupings with their correlations:

TABLE XXIV

GROUPING OF TESTS ON THE BASIS OF THEIR CORRELATION COEFFICIENTS

Group I. Coordination and Tapping, Correlation $+ .48$ with each other.			
Group II. Cancellation, Checking, Color Naming, Word Naming, Substitution.			
	Average Correlation of tests within group		$+ .32$
"	"	Cancellation with all others	$+ .35$
"	"	Checking " " "	$+ .36$
"	"	Color Naming " " "	$+ .27$
"	"	Word Naming " " "	$+ .34$
"	"	Substitution " " "	$+ .30$
Group III. Directions, Opposites, Verb-object, Mixed Relations, Word Building, and Completion.			
	Average Correlation of tests within group		$+ .32$
"	"	Directions with all others	$+ .25$
"	"	Opposites " " "	$+ .40$
"	"	Verb-object " " "	$+ .31$
"	"	Mixed Relations " " "	$+ .40$
"	"	Word Building " " "	$+ .25$
"	"	Completion " " "	$+ .30$
Group IV. Word Recollection, Word Recognition, Logical Recollection, Logical Recognition.			
	Average Correlation of tests within group		$+ .38$
"	"	Word Recollection with all others	$+ .39$
"	"	Word Recognition " " "	$+ .37$
"	"	Logical Recollection " " "	$+ .40$
"	"	Logical Recognition " " "	$+ .35$
Group V. Information and Vocabulary.			
Miscellaneous: Digit Span, Knox Cube.			

Thus Tapping and Coordination correlate $+ .48$ with each other, but both tests show a much lower correlation with the other tests. (The correlations outside of the group range from $+ .33$ to $+ .01$). This agrees with Thorndike's theory that tests of the motor sensory level correlate rather closely with each other, but only loosely with tests of other levels. In Group II, Checking has an average correlation of $+ .36$ with the others of the group, and also a much lower

correlation with tests outside Group II, (ranging from + .30 to - .04). Similarly, in Group III, Opposites and Mixed Relations both have an average correlation of + .40 with the other tests in this group, but a lower correlation with any test outside the group, again conforming to Thorndike's contention that tests on the associative level correlate closely with each other, but rather loosely with tests on other levels. (The average correlation of Opposites with the tests outside Group III is + .15; the average correlation of Mixed Relations with tests outside Group III is + .10). In Group IV, also, Logical Recollection has an average correlation of + .40 with the other tests in the group, but a lower correlation with any test outside this group. (The correlations outside the group run from + .30 to + .01). Information and Vocabulary differ from the other tests of the series in that they are indicative of one's learning rather than one's innate ability. There is only a chance correlation between them and the other tests. A more detailed discussion of this relationship we will postpone till the following section. As for Knox Cube and Digit Span, perhaps the best plan is to consign them to the miscellaneous class. Knox Cube shows on the whole the closest correlations with the tests in Group II, but the average group correlation is not high enough to warrant us definitely placing it in this group rather than in Group IV. In like manner, aside from its surprisingly high correlation with Cancellation (+ .77), Digit Span shows no close relationship with any other test. If we omit these four tests, (namely, Information, Vocabulary, Knox Cube, and Digit Span), we do get very definite groupings of the other tests, as shown in Table XXII above, indicating that we are measuring different abilities. The rather high intercorrelations between the tests of each group, together with their low correlations with tests outside their own groups would support this view. There is no evidence from these results to support Spearman's theory that correlations are produced between all sorts of performance, the amount of the correlation being simply proportional to the extent that the performances concerned involve the use of a general common factor or "general ability." Our data give evidence neither of a common factor nor of a hierarchial arrangement of the correlations. Attempts to arrange the correlations to form a hierarchy met with even greater failure than Simpson reports.

The simplest and clearest way to explain the existing relationships between our tests seems, therefore, to arrange them in the groups indicated in Table XXIV—a grouping supported by the

TABLE XXV
INTER-TEST CORRELATIONS CORRECTED FOR ATTENUATION

	Coordination	Tapping	Cancellation	Checking	Color Naming	Directions	Opposites	Verb-object	Mixed Relations	Word Building	Word Naming	Knox Cube	Digit Span	Word Memory—Recollection	Word Memory—Recognition	Logical Memory—Recollection	Logical Memory—Recognition	Substitution	Completion	
1. Coordination																				
2. Tapping	+0.67																			
3. Cancellation	+0.48	+0.67																		
4. Checking	+0.39	+0.27	+0.48																	
5. Color Naming	+0.13	+0.40	+0.23	+0.33																
6. Directions	+0.01	+0.27	+0.13	+0.35	+0.32															
7. Opposites	+0.11	+0.10	+0.17	+0.30	+0.36	+0.36														
8. Verb-object	+0.19	+0.25	+0.19	+0.38	+0.48	+0.52	+0.59													
9. Mixed Relations	+0.17	+0.01	+0.05	+0.22	+0.28	+0.52	+0.83	+0.59												
10. Word Building	+0.16	+0.13	+0.47	+0.33	+0.27	+0.12	+0.46	+0.33	+0.35											
11. Word Naming	+0.39	+0.42	+0.62	+0.44	+0.38	+0.26	+0.23	+0.45	+0.34	+0.50										
12. Knox Cube	+0.18	+0.19	+0.26	+0.35	+0.24	+0.28	+0.23	+0.19	+0.39	+0.13	+0.21									
13. Digit Span	+0.14	+0.32	+0.00	+0.12	+0.29	+0.18	+0.22	+0.09	+0.26	+0.12	+0.18	+0.20								
14. Word Memory—Recollection	+0.35	+0.00	+0.18	+0.33	+0.33	+0.16	+0.14	+0.14	+0.00	+0.45	+0.50	+0.37	+0.23							
15. Word Memory—Recognition	+0.43	+0.23	+0.07	+0.07	+0.13	+0.22	+0.04	+0.17	+0.09	+0.17	+0.25	+0.15	+0.25	+0.00						
16. Logical Memory—Recollection	+0.48	+0.02	+0.13	+0.26	+0.12	+0.32	+0.31	+0.19	+0.47	+0.51	+0.36	+0.42	+0.11	+0.00	+0.71					
17. Logical Memory—Recognition	+0.24	+0.13	+0.12	+0.24	+0.07	+0.46	+0.20	+0.13	+0.26	+0.03	+0.15	+0.37	+0.04	+0.61	+0.53	+0.95				
18. Substitution Whole	+0.37	+0.29	+0.48	+0.34	+0.41	+0.16	+0.28	+0.39	+0.03	+0.34	+0.41	+0.26	+0.11	+0.62	+0.60	+0.33	+0.20			
19. Completion	+0.01	+0.09	+0.13	+0.22	+0.27	+0.22	+0.44	+0.18	+0.71	+0.52	+0.39	+0.33	+0.28	+0.32	+0.04	+0.35	+0.21	+0.20		
Average correlation between each test and all the others	+0.25	+0.21	+0.32	+0.28	+0.27	+0.28	+0.30	+0.30	+0.28	+0.30	+0.36	+0.22	+0.23	+0.32	+0.26	+0.39	+0.26	+0.32	+0.27	

actual correlation coefficients. The tests within each group seem to be closely related to each other because they possess elements in common—elements serving to bind them closely to each other, but loosely to tests without their own groups. Thus, Group I involves motor capacity and skill; Group II powers of perception and comprehension; Group III associational relations; Group IV pure memory. Though there is some slight overlapping in the qualities called into play in the various groups, nevertheless it is not sufficient to spoil our classification.

Table XXV gives the inter-test correlations corrected for attenuation. The correlations are all higher but show in general the same relationship. They range from + 1.00 (Cancellation and Digit Span; Word Recollection and Word Recognition; Word Recollection and Logical Recollection) to + .00 (Tapping and Word Recollection; Mixed Relations and Word Recollection). When the correlations are corrected for attenuation, Logical Recollection shows the highest average correlation (+ .39) with the other tests (Omitting Information and Vocabulary). Then, in order, Word Naming; Substitution, Word Recollection and Cancellation; Opposites, Verb Object and Word Building; Checking, Directions and Mixed Relations; Completion and Color Naming; Word Recognition and Logical Recognition; Coordination, Digit Span, Knox and Tapping.

The corrected coefficients of correlations also support the groups of tests given in Table XXIV. It is possible to arrange the attenuated correlations in the same groups as those given by the raw correlations. The corrected coefficients of correlation are higher than the raw correlations but the relationship between the tests is similar.

To determine the reliability of the test scores, an investigation was conducted three years after the testing of the first group of one hundred freshmen (Group I). Two trials of the tests were given to a group of 45 freshmen during the period extending from March 14 to May 15, 1919, inclusive. The two trials occurred in every case on the same day and required approximately 45 minutes of the student's time. Table XXVI gives a list of the tests employed in two trials.

The method of procedure in conducting these tests with the 45 freshmen was the same as that employed with the 200 freshmen in Groups I and II. Moreover, all the tests were conducted individually just as was done in testing the freshmen in Groups I and II, and the room employed for the testing was the same as in the former

TABLE XXVI

1. Coordination	Trials 1 and 2 identical, same as with Groups I and II.
2. Tapping	Trials 1 and 2 identical, same as with Groups I and II.
3. Cancellation	First half of Woodworth-Wells' blank used in Trial 1, and second half in Trial 2.
4. Checking	First half of Woodworth-Well's blank used in Trial 1, and second half in Trial 2.
5. Color Naming	Trials 1 and 2 identical.
6. Directions	Woodworth-Wells' blank used in Trial 1; Wells' alternative form used in Trial 2.
7. Opposites	The first half of each of these Woodworth-Wells' blanks was used in Trial 1, and the second half in Trial 2.
8. Verb-object	
9. Mixed Relations	
10. Word Building	Letters <i>a e i l p r</i> used in Trial 1. (Same as in groups I and II). Letters <i>a e o b m t</i> used in Trial 2.
11. Word Naming	Trials 1 and 2 identical.
12. Knox Cube	Trials 1 and 2 identical.
13. Digit Span	Trial 1 as in Groups I and II; equivalent form used in Trial 2.
14. Word Recollection	Trial 1 the same as in Groups I and II; equivalent Mulhall form used in Trial 2.
15. Word Recognition	
16. Logical Recollection	
17. Logical Recognition	Given only once. (The closeness with which the correlations of the first half of the test with the other tests agreed with the correlations of the second half of the test with the other tests, measures the reliability of this test.) The correlation between the score in the first half of the blank and the score in the second half of the blank was taken as the measure of reliability.
18. Substitution	
19. Completion	Given only once. The correlation between the score in the odd numbered sentences and the score in the even numbered sentences was taken as the measure of reliability.

investigations. Just as we found the average and P.E.'s for the various tests to be approximately the same for both groups I and II, so the norms for this group of 45 freshmen are approximately the same as those obtained for Groups I and II. Thus, since one group of Barnard freshmen appears very similar to any other group of Barnard freshmen selected at random, we may fairly assume that the coefficients of reliability secured with any one group will also be indicative of the relationship that would exist between two trials with any other group selected at random. If, then, we find the reliability of the tests high for this group of 45, it is fair to judge that it would have been equally high with the group of 100 fresh-

men, (Group I), whose test scores were used in computing the correlations given in Table XXIII.

TABLE XXVII
TEST CORRELATIONS BETWEEN TRIAL I AND
TRIAL 2—GROUP OF 45 FRESHMEN

1. Coordination	+ .66
2. Tapping	+ .77
3. Cancellation	+ .60
4. Checking	+ .88
5. Color Naming	+ .88
6. Directions.	+ .76
7. Opposites	+ .79
8. Verb-object	+ .70
9. Mixed Relations	+ .60
10. Word Building	+ .70
11. Word Naming	+ .71
12. Knox Cube	+ .69
13. Digit Span	+ .83
14. Word Memory—Recollection	+ .18
15. Word Memory—Recognition	+ .33
16. Logical Memory—Recollection	+ .48
17. Logical Memory—Recognition	+ .73
18. Substitution	+ .70
19. Completion	+ .77

Table XXVII shows the correlation between the first and second trial for each of the 19 psychological tests. With three exceptions—Word Recollection (+ .18), Word Recognition (+ .33), and Logical Recollection (+ .48)—the correlations are high enough to indicate a high degree of reliability. These reliability correlations range from + .88 in the case of checking and Color Naming to + .60 in the case of Cancellation and Mixed Relations. If we disregard Word Recollection, Word Recognition, and Logical Recollection on the ground that their low reliability coefficients suggest that their correlations with the other tests do not give us an exact measure of the existing relationship, we have remaining a series of 16 reliable tests. The inter-test correlations based upon the scores in these 16 tests are accurate indicators of the true relationship existing between these tests. Our conclusions drawn from these inter-test correlations are, moreover, strengthened by our knowledge that they are based upon reliable test scores which give an accurate measure of the freshman's ability in these tests.

SECTION VII

CORRELATIONS BETWEEN THE TESTS AND
ACADEMIC MARKS

TESTS VERSUS MARKS AS MEASURES OF MENTAL
ABILITY

The charts discussed in Section V showed that the freshman scores in the psychological tests were distributed according to the normal probability curve. Tables XXVIII to XXXII inclusive, show the distribution for the five groups of academic marks, based on grades of freshmen in Group I.

TABLE XXVIII

LANGUAGE

Grade	Frequency
F (50-60)	2
D (60-70)	14
C (70-80)	49
B (80-90)	30
A (90-100)	2

TABLE XXIX

MATHEMATICS

Grade	Frequency
F (50-60)	1
D (60-70)	14
C (70-80)	33
B (80-90)	30
A (90-100)	10

TABLE XXX

SCIENCE

Grade	Frequency
F (50-60)	4
D (60-70)	6
C (70-80)	16
B (80-90)	12
A (90-100)	3

TABLE XXXI—PHILOSOPHY

Grade	Frequency
F (50-60)	0
D (60-70)	1
C (70-80)	10
B (80-90)	12
A (90-100)	4

TABLE XXXII—HISTORY

Grade	Frequency
F (50-60)	0
D (60-70)	4
C (70-80)	16
B (80-90)	4
A (90-100)	2

Not only is there a coarse grouping (only five units) as compared with the fine grouping of scores in the various psychological tests (15 to 20 units), but the distributions fail to follow the normal error curve as is the case in the test scores. With the academic marks there is a decided skewing of the distribution curves toward the good or positive end. It seems as though instructors made a deliberate effort to avoid failing their students. As for the passing grades, inspection of the marks suggests that there is insufficient care in rating students according to their relative abilities in various courses.

Observation of the uniform surfaces of frequency obtained when these one hundred freshmen were given the twenty-three psychological tests, compared with the decidedly skewed distributions for the same students in academic marks, prepares us for correlation tables XXXIII and XXXIV.

Table XXXIII shows the correlation between the scores of all the psychological tests (excluding Information), and the marks in each of the five academic groups for the freshmen in Group I. Language shows a fair positive correlation with Mixed Relations (+ .20), Word Building (+ .31), Completion (+ .30), and Vocabulary (+ .41), *i. e.*, with the tests in which the language factor performs a significant role. Mathematics shows a fair positive correlation with Cancellation (+ .28), Checking (+ .22), tests involving simple mathematical processes, and Knox Cube (+ .24). Science shows positive correlations with Opposites (+ .33), Verb-object (+ .23), Mixed Relations (+ .30), tests involving the higher thought processes needed in understanding the science courses given at Barnard, Knox Cube (+ .34), a test involving powers of perception and observation which are necessary in scientific laboratory work, and Logical Recollection (+ .21), which is also an important factor in scientific work.

The correlations of Philosophy with Cancellation + .37, Word Naming (+ .29), Knox Cube (+ .28) and Digit Span (+ .22) are unexpected.

TABLE XXXIII
CORRELATIONS BETWEEN TESTS AND ACADEMIC RECORDS

	Language	Mathematics	Science	Philosophy	History
Coordination	-.12	+.05	-.03	+.03	+.15
Tapping	-.16	+.01	-.10	+.15	+.00
Cancellation	+.14	+.28	+.04	+.37	+.10 -
Checking	-.01	+.22	+.06	+.10	+.02 -
Color Naming	+.11	+.07	+.12	-.07	-.05
Directions	+.03	-.10	-.03	-.22	+.13 -
Opposites	+.17	-.01	+.33	+.01	+.30 -
Verb-Object	+.04	+.03	+.23	+.17	-.05 -
Mixed Relations	+.20	+.01	+.30	+.12	+.19 -
Word Building	+.31	+.15	+.00	-.17	+.24 -
Word Naming	+.10	+.06	+.02	+.29	+.09 -
Knox Cube	+.18	+.24	+.34	+.28	+.08
Digit Span	+.19	+.19	+.05	+.22	+.33 -
Word Memory—Recollection	-.01	-.23	-.07	-.27	-.03
Word Memory—Recognition	+.06	+.02	+.12	+.10	+.13
Logical Memory—Recollection	+.13	+.13	+.21	-.03	+.40
Logical Memory—Recognition	-.03	+.06	+.03	-.08	+.02
Substitution 1st Half	-.08	+.11	+.09	-.19	+.18
Substitution 2nd Half	-.05	+.08	+.06	-.14	+.26
Substitution Whole	-.10	+.11	+.00	-.19	+.14 -
Completion	+.30	+.02	+.05	+.17	+.14
Vocabulary	+.41	-.05	+.12	+.09	+.23 -

TABLE XXXIV

CORRELATION BETWEEN TESTS AND INTELLIGENCE QUOTIENT

	Intelligence Quotient
Coordination	+.18
Tapping	+.17
Cancellation	+.22
Checking	+.20
Color Naming	+.23
Directions	+.20
Opposites	+.24
Verb-object	+.23
Mixed Relations	+.20
Word Building	+.22
Word Naming	+.26
Knox Cube	+.22
Digit Span	+.16
Word Memory—Recollection	+.14
Word Memory—Recognition	+.17
Logical Memory—Recollection	+.23
Logical Memory—Recognition	+.18
Substitution—First Half	+.27
Substitution—2nd Half	+.25
Substitution—Whole	+.27
Completion	+.21
Vocabulary	+.03

History shows positive correlations with Opposites (+ .30), Word Building (+ .24), Digit Span (+ .33), Logical Recollection (+ .40), and Substitution (+ .26), *i. e.*, with the tests involving ability to memorize logical material and ability to perceive relationships between facts—two essentials for successful performance in the required first-year history course at Barnard.

In general, then, the five academic groups show positive correlation with tests which we would expect to correlate with them. Table XXXIV gives the correlations between the tests and the composite score of all the academic groups. The correlations are all positive, ranging from + .14 to + .27 (excluding Vocabulary), suggesting a positive relationship. They are, however, too low to be used for diagnostic purposes. Aside from a few correlations in Table XXXIII previously mentioned, the correlations between the various tests and each of the five academic groups are even less susceptible to use for practical purposes.

In view of these low correlations and the wide variation in corre-

lations obtained between tests and marks by other experimenters, the question arises: Do the academic marks or the psychological tests give the more reliable estimates of the student's mental ability? The present writer believes that the psychological tests give the more adequate measures.

What meager experimental data there is relevant to this question of the reliability of school marks, corroborates this view. The skewed distributions in the case of the Barnard academic grades were indicated before—a fact which has been noted by investigators in the case of other institutions.⁴⁷

Professor Max Meyer,⁴⁸ making a statistical study of all the marks of forty instructors given during a period of five years at the University of Missouri, found a striking lack of uniformity in the standards of grading used. So striking was the non-uniformity that the college authorities were moved to establish a definite system of marking in 1908, with the aim of overcoming the tendency of the instructors to distribute grades according to personal opinion. Following Meyer, a study of the distribution of marks at the University of Wisconsin was made by Dearborn,⁴⁹ and of the marks at Harvard University and the University of California by Foster.⁵⁰ These, and studies made at the University of Chicago, Amherst College, and Columbia University, agreed in showing the same wide variation in the standards of grading employed by instructors.

Aikins⁵¹ found a slight difference in the relative positions assigned to 17 students in a philosophy class by the students themselves on the basis of several ten-minute tests, and the positions he assigned them on the basis of four hour tests. Smith gives several plates, illustrating clearly the great discrepancies and marked lack of uniformity in marking systems at the University of Iowa.⁵²

Zerbe, in a detailed study of the distribution of grades assigned for academic work and those assigned for shop work at the School of Applied Industries, Carnegie Institute of Technology, found that the grades as distributed for the shop work were based on a much lower standard than the grades assigned for the theoretical

⁴⁷ Kelly in a monograph entitled "Teachers' Marks" has given a history of the standards of marking in elementary schools, high schools, and colleges.

⁴⁸ Meyer, Max. *The Grading of Students*, Science, 28; 243-252.

⁴⁹ Dearborn, W. F. *School and University Grades*.

⁵⁰ Foster, William T. *Scientific vs. Personal Distribution of College Credits*; *Popular Science Monthly*, 78; 378-408.

⁵¹ Aikins, H. A. *The Reliability of "Marks," Science, N. S.*, 1910, 32; 18-19.

⁵² Smith, A. G. *A Rational College Marking System*, *Journ. of Educ. Psychol.*, 1911, 2; 383-393.

subjects.⁵³ He also observed a marked lack of conformity to a standard in the case of grades given by individual instructors. When Jones⁵⁴ gave an opposites test and a memory test to each of two elementary psychology classes, taught by different instructors, he obtained these interesting results:

	Instructor "A" (28 students)	Instructor "B" (33 students)
Class standing and opposites09	.49
Class standing and memory44	.07

These correlations were explained when further investigation revealed that instructor A taught by the outline method, emphasizing the memory factor, whereas instructor B discouraged verbatim statements taken from the text book. Both instructors were teaching the same subject, but assigning grades according to entirely different standards.

After an exhaustive study of the question at Harvard and other institutions, President Foster of Reed College concluded that "Not only are there extreme variations among different courses, but there are variations in the same course from year to year that cannot be accounted for, apparently, by any of our scientific studies in the distribution of abilities among human beings. From Maine to California the administration of college credits, although alike in no other particular, agrees in this: "That its basis is personal rather than scientific."⁵⁵ Recognition of this personal equation factor has led Smith, Weiss,⁵⁶ Zerbe, Foster, Starch, and other investigators to emphasize the need of a uniform system of grading. They agree, moreover, in maintaining that the distribution of college grades, when properly assigned, should conform to the normal probability curve. In 1914, a committee on standardizing grades at George Washington University made a similar proposal. Definite attempts to enforce such systems of marking are now being used at the University of Missouri, Reed College, and other institutions.

Even in a more restricted and more objective situation when instructors are asked to assign grades according to performance in a definite task—as for example, in a written examination paper, there is great variability due to the widely different subjective

⁵³ Zerbe, J. L. Distribution of Grades. *Journ. of Educ. Psychol.*, 1917, 9; 575-588.

⁵⁴ Jones, E. S. A Suggestion for Teacher Measurement. *School and Society*, 1917, 6; 321-322.

⁵⁵ Zerbe, J. L. Distribution of Grades.

⁵⁶ Weiss, A. P. School Grades—To what Type of Distribution shall they Conform? *Science*, 1912, 36; 403-407.

standards employed by the teachers in judging.⁵⁷ Jacoby found a variation of 1.5 points out of 10 in the grades of six professors of astronomy in marking eleven astronomy papers.⁵⁸ Starch and Elliott had facsimile reproductions made of two first-year English papers and a geometry paper, printed on the same kind of paper the students had written them on.⁵⁹ These they then had rated by 142 high school teachers of these two subjects. The English papers were also rated by a class in the Teaching of English in the University of Wisconsin and by a Summer School class of teachers in the University of Chicago. They found that the grades assigned to the two English papers by 142 English teachers ranged in the case of one paper from 64 to 98, with a probable error of 4.0, and in the case of the other from 50 to 98, with a probable error of 4.8. The grades of the mathematics paper assigned by 118 mathematics teachers ranged from 28 to 92, with a probable error of 7.5 points.⁶⁰

In a later investigation Starch had ten college freshman English papers graded independently by ten instructors of the various sections of freshman English.⁶¹ He found as wide a range of marks as he obtained with the English and Mathematics papers of his former investigation. Moreover, when ten papers were regraded by the same instructor after a certain interval of time, Starch found an average difference between the first and second grading of 4.4 points. He also found a mean variation of the grades assigned by teachers in different schools of 5.4 points, by teachers in the same department and institution of 5.3 points, and of grades assigned at different times by the same teachers to their own papers of 2.2 points. On the basis of all his data, he concluded that the best marking scale is 100, 95, 90, 85, 80, etc., and that the distribution of grades should follow the probability curve.

All the studies thus far made in this field indicate this same variation in standards of grading. There are, moreover, additional factors which render school marks absolutely unreliable measures of a student's mental ability, and cause low correlations between psychological tests and marks.

⁵⁷ For illustrations of the variability of Civil Service examiners in rating the same papers, the variation between the marks of teachers in New York State on the one hand, and the Regents on the other, see Kelly's monograph.

⁵⁸ Jacoby, H. *The Marking System in the Astronomical Course at Columbia College, 1909-1910*, Science, 31; 819.

⁵⁹ Starch and Elliott, *Reliability of Grading High School Work in English*, School Review, September, 1912.

⁶⁰ Starch and Elliott, *School Review*, 21, 254-259.

⁶¹ Starch, D. *The Reliability and Distribution of Grades*, Science, 1913, 38; 630-636.

James, from work done at Whitewater Normal School, gives these three reasons for the low correlations obtained by him:⁶²

"1. The reluctance of nearly all teachers, and their inability because of the limitations of our poor rating methods, to rate the good students as high as they should be rated, or the poor ones as low as they should be rated."

"2. The rather closer application to their studies made by the less able, due to greater anxiety and more time at their disposal."

"3. The easy-going satisfaction displayed by many able minds content with what is for them mediocre accomplishment, and the greater drain on their time imposed by fellow-students for outside activities of all kinds."

From data obtained from a questionnaire sent to 127 delinquent college freshmen and to their high school principals, Miner concluded that such traits as "lack of purpose, laziness, and lack of resistance to social and other distractions" often explain a student's failure in school work.⁶³ Their marks in such cases are unreliable measures of their ability. Scott manifested agreement with Miner when he stated that: "Where students stood high in the tests, but low or medium in estimates, their failure to succeed in class work was usually due to laziness, timidity, or disgust for the idea of struggling for marks."⁶⁴

Abundant statistical evidence, therefore, supports our contention that the striking lack of uniformity in standards of grading among instructors, making for skewed distributions of marks, the differences in grades assigned the same paper by teachers at different times, the personal equation in marking, the tendency of many able students to neglect studies for outside distractions and of poorer students to apply themselves more assiduously, the role played by such factors as lack of purpose or incentive, interest in outside or in college activities, economic pressure causing students to devote much time to earning money, etc., make college marks totally inadequate measures of students' ability. All these factors are influential, moreover, in making Barnard marks as unreliable as marks given in other colleges. No attempt is made by Barnard instructors to distribute their grades according to the normal

⁶² James, B. B. *Mutual Correlations of Intelligence, Scholarship, and Vocabulary*. *School & Society*, 1919, 9; 427. In *School & Society*, 1918, 7; 238-239, James gives similar factors as influencing the correlations between marks and tests.

⁶³ Miner, J. B. *The College Laggard*. *Journ. of Educ. Psychol.*, 1910, 1; 263-271.

⁶⁴ Scott, C. A. *General Intelligence or "School Brightness."* *Journ. of Educ. Psychol.*, 1913, 4; 509-524.

probability curve. Absolute freedom is permitted the teachers. As a result, the personal bias of the teachers plays a large part in the marks received by students. This, combined with the contributory causes above mentioned, renders Barnard marks untrustworthy.

The psychological tests, on the other hand, have much to recommend them as giving reliable estimates of freshmen's mental ability. All the tests employed are standard tests. They were, moreover, administered by one experimenter according to a carefully standardized method of procedure. All conditions were kept constant—the place of testing, the attitude of the experimenter, the method of conducting the tests, and the method of scoring. Every student undertook the examination with a determination to do her level best. Whereas, in school subjects, lack of interest or incentive often caused a girl to do a lower grade of work than she was mentally capable of doing, here there was a definite incentive impelling her to exert maximum effort. Each freshman expected to receive vocational guidance based on her test scores. She accordingly took the psychological test at an hour convenient for her—when she was feeling in good condition. Genuine interest in the tests, (noted in the case of all students), coupled with a keen desire to make a favorable record, renders their test scores reliable estimates of their ability. The fact that the scores conform to normal distribution curves further indicates the reliability of these measures.

We do not claim, however, that we can predict a student's future success in college from her psychological test record. The psychological examination gives an adequate measure of what each freshman can do. From it we can make an authentic psychograph of her mental abilities indicating in which processes she is strong, and in which she is weak. Whether she will make high academic grades or attain success in later life depends not only upon her mental capacity, but upon such other factors as interest, incentive, will-power, economic stress, environmental conditions, etc. The tests, not her academic marks, measure her mental capacity; to predict her future performance in school or her success in a particular vocation, we must also consider these other factors.

SECTION VIII

CORRELATIONS BETWEEN PSYCHOLOGICAL TESTS AND PHYSICAL MEASUREMENTS. THEIR SIGNIFICANCE

There is one further problem to be considered—the relation existing between the psychological tests and the physical measurements. The correlations shown in Table XXXIII, based on the records of the one hundred freshmen in Group I, furnish an important contribution to our existing meagre data on this subject.

Most investigators who have hitherto reported correlations between physical traits and mental ability have used school marks or teachers' estimates as indicators of mental ability. Their subjects, moreover, have been school children. Porter, Smedley, De Buck, MacDonald, Gilbert, Baldwin, Pyle, King, Arnold, Wilson, and Schuyten are some of the chief workers in this field. Widely varying results have been reported, some experimenters finding positive correlations between physical traits and school progress, others negative, and still others indifferent or zero correlations. Discussing the significance of these varying correlations, Whipple says: "The trend of evidence is to the effect that all such correlations, where found, are largely explicable as phenomena of growth, *i. e.*, as correlations with relative maturity. This makes intelligible the fact that, in general, the positiveness of all such correlations lessens with age, and that many of them, indeed, become difficult or impossible of demonstration in adults."⁶⁵

Of the investigations in which adults have been used as subjects, the work of Dr. Karl Pearson is perhaps the most extensive. He made measurements of 1,000 Cambridge University students, obtaining these correlations:

Mental ability and dolichocephaly	+ .03	± .03
Mental ability and short heads	- .08	± .03
Mental ability and broad heads	+ .04	± .03

His method of rating his subjects for mental ability was extremely rough, consisting merely in grouping the men into two big classes—pass men and honor men. Similar correlations obtained by Pearson between head measurements and mental ability as measured by

⁶⁵ Whipple, G. M. *Manual of Mental and Physical Tests. Part I, p. 71.*

teachers in the case of 1856 school boys twelve years of age, lead Galton to conclude "that there is no marked correlation between ability and shape or size of the head."⁶⁶

In another investigation with Cambridge students, Pearson found zero correlations between mental ability, determined roughly as indicated above, and strength of pull, strength of squeeze, long sight, weight, and ratio of weight to stature.⁶⁷ Continued testing of Cambridge students and school children lead Pearson to conclude in 1906 that "The results (of our investigations) confirm the previous conclusion that: While there exists a slight but sensible relation between size of head and intelligence, there is no possibility of using this relation to make even rough individual predictions."⁶⁸

These investigations, although interesting, have no direct bearing upon our problem, however, which is concerned with the relationship existing between the performance of college freshmen in psychological tests and their physical measurements taken in the gymnasium.

We have good reason to feel that these physical measurements are fully as reliable and accurate estimates as are the psychological test scores. The physical examinations were all conducted in the Thompson Gymnasium of Teachers College. They were given individually, the head of the Department of Physical Education of Barnard College making all the measurements. These were then immediately recorded on the student's physical record card by an assistant. Thus any inaccuracy in taking the measurements would be a constant one, and would not disturb the relative ranking of the freshmen.

Experimental conditions were as uniform as in the case of the psychological tests. Each girl came to the gymnasium at an hour convenient for her and went through all parts of the examination according to a standardized method of procedure. No clothing was worn during the examination, save for two light cloth flaps which were fastened loosely about the shoulders by means of a draw string and two similar flaps fastened about the waist which could easily be raised in taking measurements. These were provided by the physical director for the occasion.

⁶⁶ Pearson, K. On the Correlation of Intellectual Ability with the Size and Shape of the Head. *Proc. Roy. Soc.* 1902, LXIX, 333-342.

⁶⁷ Lee, A., Lewenz, M. A., and Pearson, K. On the Correlation of the Mental and the Physical Characters in Man. II *Proc. Roy. Soc.*, 1902, LXXI, 106-114.

⁶⁸ Pearson, K. On the Relationship of Intelligence to Size and Shape of Head, and to other Physical and Mental Characters. *Biometrika*, 1906, 5; 105-146.

The physical records taken were: height measured in centimeters with a stadiometer; weight, measured in pounds with the Fairbanks scale; lung capacity, measured in cubic centimeters; and four other strength tests—grip right and left hand, upper back and chest, measured in kilograms with a dynamometer. The norms for these measurements obtained for these one hundred freshmen were given in Section V.

The curves of distribution for these seven measurements (which lack of space prevents us from printing), conform approximately to the normal probability curve. The subjects, moreover, with a very few exceptions, were all eighteen years of age or over, so that the factor of relative maturity does not affect the correlations. The freshmen are a rather homogeneous group with respect to age. These facts, coupled with the accuracy of both the physical and psychological measures give us good reason to believe in the reliability of the correlations in Table XXXIII.

It is interesting to note that six of the seven physical measurements—all except lung capacity—manifest zero or chance correlations with all the psychological tests. The average correlation of each of these six measures with all the psychological tests is as follows: Height with all the tests, $+ .05$; weight $+ .06$; strength of grip, right hand, $+ .04$; strength of grip, left hand, $+ .02$; strength of upper back, $+ .02$, and strength of chest, $+ .05$. As these correlations are all less than the probable error ($\pm .068$) they indicate clearly that there is no connection between these physical measurements and a freshman's mental ability as indicated by her psychological test records. In the case of lung capacity, all the correlations (except with vocabulary) are positive. They are markedly low, though, the average correlation between lung capacity and all the psychological tests being only $+ .10$. This is little more than the probable error, indicating the existence of only a chance relationship.

The uniformity of the single correlations in exhibiting this tendency toward chance relationship is significant. In only eight cases out of the total number of 154 correlations, or, in fact, we might say in only six cases, since the correlations between Substitution First-half and lung capacity ($+ .20$) and Substitution Second-half and lung capacity, ($+ .26$) duplicate information yielded by the correlation between Substitution Whole and lung capacity ($+ .24$)—are there correlations of $+ .20$ or over. The highest correlation is only $+ .26$ (Substitution Second-half and lung capac-

ity), which is too low to admit of diagnostic purposes. With these few exceptions, all the correlations between physical measurements and the tests—146 correlations in all—show approximately zero relationship. The large number of these correlations justifies us

TABLE XXXV
CORRELATIONS BETWEEN TESTS AND PHYSICAL MEASUREMENTS

	Height	Weight	Lung Capacity	Strength Grip r.h.	Strength Grip l.h.	Upper Back	Chest
1. Coordination	+09	+01	+10	+08	+11	+02	+15
2. Tapping	-.05	-.07	+10	+17	+16	+11	+19
3. Cancellation	+15	-.08	+17	+07	+02	+06	+08
4. Checking	-.01	-.14	+10	+04	-.00	-.06	+09
5. Color Naming	+12	-.05	+14	+18	+12	-.01	+03
6. Directions	-.05	-.14	+07	-.00	-.09	-.08	+04
7. Opposites	+05	+11	+10	-.01	+02	+08	+21
8. Verb-object	+02	-.13	+03	+01	-.10	-.05	+07
9. Mixed Relations	+04	+08	+13	+00	-.04	-.00	+11
10. Word Building	-.02	+01	+06	-.11	-.21	-.07	-.01
11. Word Naming	+12	+03	+04	+15	+09	+17	+15
12. Knox Cube	+10	-.03	+21	+12	-.05	+08	+01
13. Digit Span	-.07	-.04	+11	-.12	+00	+04	-.03
14. Word Memory—Recollection	-.04	+13	+04	+02	+15	-.07	+02
15. Word Memory—Recognition	+06	+07	+05	-.14	-.04	-.06	-.10
16. Logical Memory—Recollection	+18	+22	+22	-.05	+01	+10	+09
17. Logical Memory—Recognition	+09	+11	+14	-.05	+04	-.01	+05
18. Substitution—First Half . .	+19	+05	+20	+02	+07	-.00	+02
19. Substitution—2nd Half . .	+17	-.02	+26	+05	+07	+09	+02
20. Substitution—Whole	+19	+00	+24	+04	+07	+06	+01
21. Completion	-.12	-.05	+04	-.02	-.05	+02	-.03
22. Vocabulary	-.02	+07	-.17	+07	-.05	-.04	-.15
Average	+05	+06	+10	+04	+01	+01	+04

in concluding that the relationship between the physical measures and the tests is one of chance only.

It is interesting to know that the only other experimenter who has reported the results of a similar study with college freshmen supports this view. Although Wissler in his study of the results of the old Columbia freshman tests reports only two correlations

between the physical tests and the psychological tests—namely, a correlation between length of head and logical memory of $+ .21$, and between breadth of head and logical memory of $- .05$ —the observation of the records of freshmen in other physical tests compared with their records in the psychological tests lead Wissler to conclude: "That the physical tests show a general tendency to correlate among themselves, but only to a very slight degree with the mental tests."⁶⁹

Although the physical measurements exhibit only a chance connection with a freshman's psychological test score, they should be taken into consideration by an instructor or advisor whose duty it is to give guidance to a student in planning her college course. In Section V we pointed out the case of a freshman (Chart 3, G.S.), whose net score in the psychological examination was well above the average freshman record, but whose standing in academic work was in the lowest quintile of the class. The fact that she made the best record in the class in the physical measurements, together with the information we later acquired concerning her athletic activities, explained her academic failure. The more varied measures of a student we have, the better qualified we will be to make an adequate psychograph of a student's relative abilities and disabilities, in various lines.

⁶⁹ Wissler, Clark. *Psychological Review Monograph Supplement*, June, 1901.

GENERAL SUMMARY OF THE RESULTS WITH SUGGESTIONS FOR THE PRACTICAL USE OF THE TESTS

A series of nineteen psychological tests was given to two groups of one hundred Barnard freshmen each with the aim first of establishing norms and standards of performance and giving students a clear conception of their abilities and aptitudes along various lines and second of determining the reliability of the tests and their correlations with freshmen university grades and physical measurements.

All the tests were given individually according to a standardized method of procedure and under standard conditions.

The averages and surfaces of distribution for the first group of one hundred freshmen (Group I) are approximately the same as for the second group of one hundred (Group II) and for a third group of forty-five freshmen—showing that Barnard freshmen are a homogeneous group, differing little from year to year.

The inter-test correlations range from + .77 (between Cancellation and Digit Span) to .00 (between Tapping and Word Recollection and between Mixed Relations and Word Recollection). The positive correlations between Cancellation and the other tests (+ .03 to + .77) contradict the old compensation theory. The fact that the correlations are all positive is suggestive of a definite relationship between Cancellation and these various tests.

Checking and Word Naming show the highest average correlation (+ .25) with the other tests (omitting Information, Vocabulary, Word Recollection and Word Recognition); then, in order, Opposites; Verb-object and Cancellation; Color Naming; Directions, Mixed Relations, Word Building, and Completion; Logical Recollection and Substitution Whole; Knox; Tapping and Digit Span; Coordination; Logical Recognition.

On the whole, the inter-test correlations, although mostly positive, are low, indicating that we are testing different mental abilities.

On the basis of the relationship shown by the correlation coefficients we may divide the tests into three groups: (1) motor tests (Coordination and Tapping); (2) tests involving powers of perception and comprehension (Cancellation; Checking, Color Nam-

ing, Word Naming and Substitution); (3) tests involving associative relations (Directions, Opposites, Verb-object, Mixed Relations, Word Building and Completion); (4) tests which call into play powers of learning, viz., observation and retention—(Word Memory and Logical Memory); (5) tests depending on the subject's knowledge more than on her innate ability (Information and Vocabulary); (6) miscellaneous group (Digit Span and Knox Cube). There is only a chance correlation between Information and Vocabulary and the other tests. With the exception of this group and Digit Span and Knox Cube, the remaining groups of tests correlate closely among themselves but loosely with the other tests.

There is no evidence from these results of a general common factor nor of a hierarchial arrangement of the correlations.

The tests within each group seem to be closely related to each other because they possess elements in common—elements serving to bind them closely to each other but loosely to tests without their own groups.

The coefficients of correlation corrected for attenuation are considerably higher than the raw correlations but show in general the same relationships.

The coefficients of reliability are low for Word Recollection (+ .18), Word Recognition (+ .33) and Logical Recollection (+ .48). For the other tests they range from + .88 (Checking and Color Naming) to + .60 (Cancellation and Mixed Relations). We have, thus, a series of sixteen reliable tests. Inter-test correlations based upon the scores in these sixteen tests are accurate indicators of the true relationship existing between these tests.

The psychological tests show low correlations both with each of five academic groups (1) Language, (2) Mathematics, (3) Science, (4) Philosophy and (5) History, and with the composite score of all the academic marks (+ .14 to + .27).

Lack of uniformity in standards of grading among instructors, causing skewed distribution curves of marks, the personal equation in marking, the role played by such factors as lack of incentive, interest in outside or college activities, economic pressure, etc., make college marks inadequate measures of the students' ability.

There is evidence that the psychological tests give a true estimate of each freshman's mental capacity. To predict her performance in school or in a future vocation both her capacity and such other factors as interest, incentive, will-power, environmental conditions, etc., must be considered.

The correlations between the physical measurements and the psychological tests show approximately zero or chance relationship.

Psychographic charts may be constructed, showing each student her relative rank in the tests, academic grades and physical measurements. Such psychographs may be put to practical use as for example, in cases where a student is doing academic work of a grade below the level her test record showed her capable of.

The results of this investigation make it possible to offer a few tentative suggestions to college administrators who desire to institute a system of student guidance. The first step in such a plan might well be to put each member of the freshman class through a thorough physical examination to determine her physical fitness for undertaking college work. This examination should be made by the director of the Physical Education department or a competent assistant in the department. Students with correctible physical defects should be given proper treatment—eyeglasses, special physical exercises or what not, according to their needs. Those suffering from a slightly run down condition might be advised to take a light program until they regained their normal condition; those too far below par might be advised not to enter college.

The second step might be to obtain an estimate of her mental capacity on the basis of her score in a psychological examination. A psychologist (who might also act as vocational advisor) with an assistant might well be in charge of this work. If possible, each freshman should be tested individually, the same experimenter conducting all the tests according to a standard method of procedure. As for the particular tests to be used, they should be varied in character, adapted to measure various mental abilities. A series that may be divided into several groups, each group testing a rather definite mental ability and such that tests within each group correlate highly among themselves but loosely with all tests outside their own group, as in the present investigation—perhaps represents the ideal type of tests. The particular series of tests employed in this study is not, however, recommended as the best series of tests that might be used. It is very probable that a series could be found that will test more significant mental abilities and such that the tests within each group will correlate more closely with each other and more loosely with other tests. Only by empirically trying out different series can the ideal series be found.

Where lack of time or the size of the freshman class makes it impossible to test each freshman individually, a comprehensive

group test that has been found successful—as for example, the Army Alpha or the Thorndike Group test—may be employed. In view of the successful results secured with these group tests and the speed with which they may be administered, it may well be that such a comprehensive group test as the Thorndike test would be the best to employ. In the case of students who barely passed or who failed in this group test, such a series of tests as that used in the present investigation might be used to supplement the results of the group test. It would seem that a group test which might be supplemented, where necessary, by an individual examination would be the ideal arrangement.

As we stated before, a psychologist and an assistant should preferably be in charge of the psychological testing. Perhaps a group of fifteen to twenty persons with some experience in scoring psychological tests might be employed to score the tests immediately after the psychologist has given them. In this way the examinations might be easily scored within three or four days and the reports made out for each student very soon after. The results of the psychological examination and the physical examination together with the student's academic entrance record, might then be submitted to the psychologist or vocational advisor. On the basis of these records, psychographic charts might be made out for each student indicating her strengths and weaknesses. The vocational advisor might then have an immediate interview with such students who showed any marked disabilities. In this personal conference the advisor might try to obtain from the student pertinent information concerning her interests, economic status, environmental conditions, etc. All these supplementary items of information would then enable him to form a comprehensive idea of the student's mental and moral calibre. With this as a basis vocational advice could be given the student regarding her choice of subjects, study habits, participation in extra curricula activities, etc. Perhaps such students might be asked to report at stated intervals for further conference. Much the same procedure might be followed with the other students except that here fewer conferences would be necessary.

The advisor should be free to devote all his time to supervising the academic career of the students and to rendering needed advice. Obviously such a man should be a psychologist with both ability to interpret the various measures secured of each student's ability and tact in persuading students to follow his suggestions. From the

attempts that have thus far been made in certain institutions to guide students' academic careers, it seems probable that with an able vocational advisor aided by a competent assistant such a system would be a distinct help in stimulating students to exert maximum effort in doing their college work.

VITA

The writer was born in Norwich, Chenango County, New York, on October 1, 1895. She graduated from Girls' High School, Brooklyn. In 1916 she received the degree of Bachelor of Arts from Barnard College, and in 1917 the degree of Master of Arts from Columbia University. Her work in psychology has been done under Professors R. S. Woodworth, H. L. Hollingworth, L. S. Hollingworth, E. L. Thorndike and A. T. Poffenberger. She was assistant in psychology in Columbia University 1918-20; assistant in Summer School, 1918, 1919, and 1920; assistant and instructor, Extension Teaching, 1918-19; and instructor, Extension Teaching, 1920.

THIS BOOK IS DUE ON THE LAST DATE
STAMPED BELOW

AN INITIAL FINE OF 25 CENTS
WILL BE ASSESSED FOR FAILURE TO RETURN
THIS BOOK ON THE DATE DUE. THE PENALTY
WILL INCREASE TO 50 CENTS ON THE FOURTH
DAY AND TO \$1.00 ON THE SEVENTH DAY
OVERDUE.

SEP 18 1934

12 Apr '55 DS

16 Apr '56 RE

APR 2 1955 LB

80 Nov '57 RK

REC'D LD

NOV 17 1957

23 May '62 SM

REC'D LD

MAY 9 1962

LB1181

C3

517506

UNIVERSITY OF CALIFORNIA LIBRARY

