An empirical study of certain tests for individual differences / by Mary Theodora Whitley.

AN EMPIRICAL STUDY OF CERTAIN TESTS FOR INDIVIDUAL DIFFERENCES

BY

MARY THEODORA WHITLEY, PH.D.

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. WOODWORTH

No. 19, AUGUST, 1911

COLUMBIA CONTRIBUTIONS TO PHILOSOPHY AND PSYCHOLOGY,
VOL. XX, NO. 1

NEW YORK
THE SCIENCE PRESS

ARCHIVES OF PSYCHOLOGY

Editorial communications should be addressed to PROFESSOR R. S. WOODWORTH, Columbia University, New York City.

Subscriptions and advertisements should be sent to ARCHIVES OF PSYCHOLOGY, Sub-Station 84, New York City. The subscription price is five dollars a volume, containing between six and seven hundred pages. Volume I comprises Nos. 1–10, and Volume II, Nos. 11–18. The numbers are as follows:

1. The Psychology of Mentally Deficient Children: NAOMI NORSWORTHY. $1.00.
2. (… FRANZ. 50c.
3.] … . 50 cents.
4.] … 50 cents.
5. …
6. … 50 cents.
7. …
8. … ction: SVEN …
9. … s.
10. …
11. …
12. …
13. …
14. …
15. …
16. C … o Emotional …
17. … Cloth, $1.25.)
18. A … s. (Cloth, 75 …
19. A … THEODORA …

Th… s of Philosophy, P… comprising the foll…

Meas…
Aven… s.
The …
The …
The …
Lingu…
The Diurnal Course of Efficiency: HOWARD D. MARSH. 90 cents.
The Time of Perception as a Measure of Differences in Sensations: VIVIAN ALLEN CHARLES HENMON. 60 cents.

ARCHIVES OF PHILOSOPHY

Editorial communications should be addressed to PROFESSOR FREDERICK J. E. WOODBRIDGE, Columbia University, New York City.

Orders should be sent to ARCHIVES OF PHILOSOPHY, Sub-Station 84, New York City. The numbers are as follows:

1. The Concept of Control: SAVILLA ALICE ELKUS. 40 cents.
2. The Will to Believe as a Basis for the Defense of Religious Faith: ETTIE STETTHEIMER. $1.
3. The Individual: A Metaphysical Inquiry. WILLIAM FORBES COOLEY. $1.00.

AN EMPIRICAL STUDY OF CERTAIN TESTS FOR INDIVIDUAL DIFFERENCES

BY

MARY THEODORA WHITLEY, PH.D.

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. WOODWORTH

No. 19, AUGUST, 1911

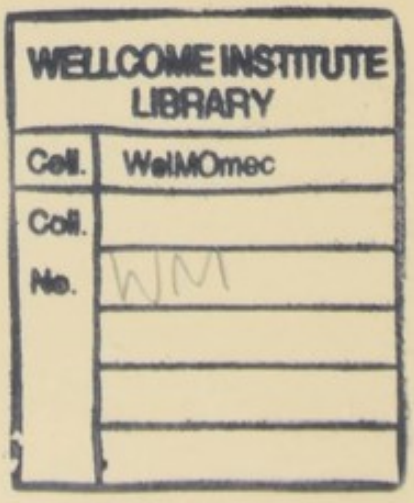
COLUMBIA CONTRIBUTIONS TO PHILOSOPHY AND PSYCHOLOGY,
VOL. XX, NO. 1

NEW YORK

THE SCIENCE PRESS

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

CONTENTS



AN EMPIRICAL STUDY OF CERTAIN TESTS FOR INDIVIDUAL DIFFERENCES

I

HISTORY OF THE INTEREST IN INDIVIDUAL DIFFERENCES

1. The Work of Various Investigators

The history of scientific inquiry into the nature and amount of individual differences dates back only about twenty-five years. Before that time experimental psychology had concerned itself chiefly with investigations into typical mental functions, especially those of perceiving the external world. For this purpose long and detailed tests were made upon a very few, or perhaps a single subject.

Galton in England was the first who devised and applied a series of tests, both physical and mental, to large numbers of subjects with a view to determining norms and studying the amount, causes and kinds of variation. Since the publication in 1883 of Galton's "Inquiries into the Human Faculty and its Development," the work done in this field in England has been chiefly confined in its application to school children; witness Bryant's experiments in 1886 in testing the character of school children[1] and the more recent work of Winch,[2] Spearman,[3] W. G. Smith,[4] Wimms,[5] and Burt.[6]

In Germany there is the general work of Münsterberg in 1891,[7] Kraepelin,[8] Aschaffenberg,[9] and Oehrn in 1896,[10] Cron in 1897,[11]

[1] *Journal of the Anthr. Inst. of Gr. Britain and Ireland.*

[2] *Brit. Jour. of Psych.*, **1**, 1904.

[3] *Am. Jour. of Psych.*, **15**, 1904.

[4] *Brit. Jour. of Psych.*, **1**, 1905.

[5] *Brit. Jour. of Psych.*, **2**, 1907.

[6] *Brit. Jour. of Psych.*, **3**, 1909.

[7] "Zur Individual Psychologie," *Centralblatt f. Nerv. in Psychiatrie*, **14**, 1891.

[8] "Der Psychologische Versuch in der Psychiatrie," *Psych. Arb.*, **1**, 1896.

[9] "Experimentelle Studien über Associationen," *Psych. Arb.*, **1**, 1896.

[10] "Experimentelle Studien zur Individuellen Psychologie," *Psych. Arb.*, **1**, 1896.

[11] "Ueber die Messung der Auffassungsfähigkeit," *Psych. Arb.*, **2**, 1897.

Cohn in 1898,[12] Stern in 1900,[13] and Wiersma in 1902.[14] In these cases experiments, if made at all, were usually in the form of a few carefully prepared tests given to a few subjects either with a view to studying their individual variations in detail or else for the sake of discussing the question of method of administration. There is also the other method of work, that of testing large groups of school children, as for instance the work of Ebbinghaus in 1897,[15] Netschajeff in 1900,[16] Lobsien in 1901,[17] and Meumann in 1905.[18]

In France under the influence of Binet and his publications in *L'Année Psychologique,* there has been an enormous amount of work done, especially with children—investigations into normal and abnormal conditions, both mental and physical, culminating in 1905 and 1908 in the Binet and Simon sets of graded tests of intelligence adapted to children of all ages from three years up. In 1904, Toulouse, in his "Technique de Psychologie expérimentale," gave, as the result of nearly ten years' work, a full and detailed exposition of the methods of giving certain tests, and of computing the results gained.

In America following the publication in *Mind,* 1890, of "Mental Tests and Measurements" by Cattell with comments by Galton there was a rapid development of the work represented by that of Bolton in 1892,[19] Gilbert in 1893–94,[20] Shaw in 1896,[21] Griffing in 1896,[22] Macdonald in 1897–98,[23] Kirkpatrick in 1900,[24] Bagley in 1901,[25] Seashore in 1901,[26] Smedley in 1901,[27] Swift in 1903,[28] and others

[12] "Experimentelle Untersuchungen . . .," *Zeitschr. für Psych.,* **15**, 1897.

[13] "Ueber Psych. der Individuellen Differenzen."

[14] "Die Ebbinghausche Combinationsmethode," *Zeitschr. f. Psych.,* **30**, 1902.

[15] "Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern," *Zeitschr. f. Psych.,* **13**, 1897.

[16] "Exp. Untersuchungen über d. Gedächtnissentwickelung bei Schulkindern," *Zeitschr. f. Psych.,* **24**, 1900.

[17] "Exp. Untersuchungen über d. Gedächtnissentwickelung bei Schulkindern," *Zeitschr. f. Psych.,* **27**, 1901.

[18] "Intelligenzprüfungen an Kindern der Volksschule," *Die Exp. Päd.,* **1**, 1905.

[19] "The Growth of Memory in School Children," *Am. Jour. of Psych.,* **3**, 1892.

[20] *Studies from the Yale Psychological Laboratory,* **1**, **2**, 1892, 1893.

[21] *Ped. Sem.,* **4**, 1896.

[22] *Psych. Rev.,* **3**, 1896.

[23] "Experimental Study of Children," in *Report United States Comm. of Ed.,* 1898.

[24] *Psych. Rev.,* **7**, 1900.

[25] *Am. Jour. Psych.,* **12**, 1901.

[26] *Ed. Rev.,* **22**, 1901.

[27] *Report Dept. of Child-study,* **3**, 1900–01 (Chicago Public Schools).

[28] *Ped. Sem.,* **10**, 1903.

on school children; that of Jastrow in 1893,[29] Thompson in 1903,[30] and Ternan in 1906,[31] on laboratory subjects (in the last instance children who came to the laboratory regularly), and further work of Cattell in 1893[32]–96,[33] and Jastrow in 1893,[34] on college students. A study of method and a somewhat extended inventory of seven subjects has also been made by Sharp.[35]

Columbia appears to be the only university still making tests upon the freshmen. An inquiry among the universities and larger colleges of the United States and Canada has resulted in fifteen replies in the negative.

This by no means exhausts the list, since a large proportion of recent investigations of whatever topic include a treatment or statement of individual differences in method of work or degree of achievement, and since, too, some treatises on the psychology of individual differences, Stern's for example, are largely reviews of other investigators' general work from this particular standpoint.

There are, aside from the questionaire method so largely used by Stanley Hall and others by which large quantities of crude, descriptive material are amassed from untrained observers, two customary methods of experimental procedure which have already been indicated. One is to use a few specialized tests upon a limited number of subjects, with a sufficient number of repetitions to establish the reliability of the reaction or to induce fatigue or practise. Oehrn, Kraepelin, Ternan, Wimms, and Binet make use of this method. The second method, scoffed at by Stern and criticized by Binet in his review of Wissler's work, is to use very simple tests, many of them physical, upon large numbers of subjects, usually without repetition. Cattell's tests for freshmen, Galton's tests and the many tests of all kinds on school children are of this nature. This latter method is the predominant one in this country to-day.

That this should be the case, is not surprising since the first laboratory work directly concerning itself with individual psychology was instituted by Cattell whose early work in individual differences has been noted. Already in the eighties his experiments on himself and others[36] on the time taken to recognize colors, letters of the alphabet, to see and name the same, and on three groups of as-

[29] *Ed. Rev.*, **5**, 1893.
[30] "The Mental Traits of Sex."
[31] *Ped. Sem.*, **13**, 1906.
[32] *Phil. Rev.*, **2**, 1893.
[33] *Psych. Rev.*, **3**, 1896.
[34] *Am. Jour. Psych.*, **4**, 1893.
[35] *Am. Jour. Psych.*, **10**, 1899.
[36] "Psychometrische Untersuchungen," *Phil. Stud.*, **2**, **3**, 1895–6.

sociation tests anticipate much that has since become part of the regular stock in trade of those who use the methods of simple mental tests of the higher psychic processes. His list of ten tests employed upon all freshmen and other volunteers in the University of Pennsylvania published in 1890,[37] was the first definite psychological inventory in this country.

In 1896 following Baldwin's suggestion at the annual meeting of the American Psychological Association a committee of five was formed consisting of himself, Jastrow, Sanford, Witmer and Cattell to consider the feasibility of cooperation among the various psychological laboratories in the collecting of mental and physical statistics. A suggestive but indefinite report was made by this committee through Witmer the next year.

In 1907 the Association again appointed a committee of five consisting of Angell, Judd, Pillsbury, Woodworth, and Seashore to determine a series of group and individual tests with reference to practical applications, and to determine standard experiments of a more technical character. Their first report appeared in December, 1910.

Not the least interesting feature of the development of the work, has been the fluctuating of opinion with regard to its value, and the criticism of the methods used in accordance with the aim in view, and the evident influence of parallel work in general psychology. For instance in Germany there is first the intensive work on some of the higher mental processes by Kraepelin and his school in the early nineties, contemporaneously with extensive work in America on simpler processes with emphasis on the accompanying physical measurements—the subjects being sometimes children—and with characteristic French investigations into abnormal and criminal types as well as into the thinking powers of school children.

The long article in Volume 2 of *L'Année Psychologique*, 1895, by Binet and Henri, is notable in that it formulates two distinct problems of individual psychology, definitely favors the use of tests complex in content and therefore less capable of precise treatment, and suggests a grouping of appropriate tests under ten functions. In this article the preceding work of Cattell, Münsterberg, Jastrow, Kraepelin, and Gilbert is illustrated and criticized. The lists of tests given by the first three men are termed too simple, incomplete and too partial—that is confined too entirely to tests of memory, sensations and physical abilities. Kraepelin's are criticized as being not only partial but impractical since the tests require five hours for completion, necessitating several visits to the laboratory. Gilbert's

[37] *Mind*, **15**, 1890.

are said to show the difference in degree but not in kind between the thinking powers of the child and the adult. Their own list of tests could be given in from one to one and a half hours. In describing them only vague directions for administration are given, and occasional illustrative results from some tests already used with school children. They conclude by saying that their tests probably need modification, and might not disclose the finer mental differences between individuals similarly trained and belonging to the same social group. The work is fruitful in suggestions, though with a sketchy indefiniteness rather than a diagrammatic precision.

Further progress, especially in the application of the tests to school children, was made in each country but along lines already indicated. Ebbinghaus[38] devised and applied a new sort of test since known as his "combination" or completion test, which aroused no little interest and discussion.

In 1899 Sharp[39] took up the question of method. The first half of her work is largely a review of the theses of Binet and Henri, while the remainder is a careful study of some of the tests suggested by them, as applied to seven college students. She considers the results unsatisfactory except that they show that a single trial of any of the tests, made in the suggested hour and a half among single trials of many other tests, would be practically valueless and most unreliable, especially in the case of the tests of a complicated nature.

The following year appeared Stern's work, "Über die Psychologie der Individuellen Differenzen." This contains a review of methods, but not of results to date, and criticisms which are largely destructive. Thus in pointing out the dangers of extensity and the probable resulting superficiality, he makes some enlivening remarks on the American fondness for the questionaire method, comparing it to the questions concerning favorite author, color, food, etc., compiled in the autograph books of the *Backfisch* of the day, which results in what he elsewhere calls "pseudostatistics." He would place no reliance on the results of any series of tests which could be completed in an hour and a half, and considers the individual differences found in sensation and perception to be due to lack of experience with the material, since practise reduces those differences. He also says that tests on memory should seek to discover ways of memorizing and length of retention rather than content, and that as a measure of association, the spoken first idea is too erratic to be trustworthy, and measures too much else besides association. He offers few definite suggestions as to methods of procedure.

[38] *Zeitschrift für Psychologie,* **13**, 1897.

[39] *Op. cit.*

In 1901 Wissler, in working over the results of the Columbia freshmen tests from the point of view of correlation, finds so little that he concludes that they tell nothing as to the general intelligence of individual college students or adults. If a functional relationship exists it must be more complex than is usually supposed and it needs further testing. He remarks that correlating successive trials would help show the precision of a test.

Two years later appeared Binet's[40] account of careful and repeated tests, extending over several months, on his two little daughters. Methods and results are given in detail and the conclusions drawn from them as to the characteristics of the two subjects. Many of the twenty different tests were those already utilized in work among school children, notably the written descriptions of objects and pictures. His object was qualitative and descriptive rather than normative, and in consequence the actual tests are supplemented by long and careful questioning as regards imagery and analysis of associations.

The same year, in the introduction to the first volume of the "Beiträge zur Psychologie der Aussage," Stern again criticizes current methods of investigation. He points out that by them either time or numbers is sacrificed, whereas data from many people should be amassed by trained observers, and similarly treated. Instead of one experimenter using a few volunteer students as subjects, another large or selected groups of school children, another his own patients, another criminal cases, and still another results of a few experiments on himself and treated by original methods—the general results being confusion rather than cohesion—there should be an Institute for Applied Psychology, to act as a centralizing and unifying agency, a sort of clearing house, with the services of a trained statistician always available! The tests used should represent actual life conditions as nearly as possible and not be at all of the type of immediate memory for colors, tones, etc., which tell as much about the memory as a microscopic study of the finger would tell of its function. How well he has succeeded in justifying his position may be gathered from the successive volumes of the *Beiträge* and the *Zeitschrift für angewandte Psychologie*.

The next year a distinct advance towards synthesis and standardization of tests was made in the carefully prepared work of Toulouse, Vaschide, and Piéron.[41] Without quoting results to be expected or norms to be employed, explicit directions are given for the administration of nearly fifty tests, more than half of which are on memory.

[40] "L'étude expérimentale de l'intelligence," 1903.

[41] "Technique de Psychologie Expérimentale," 1904.

Ways of scoring are also illustrated at some length. The tests suggested have been selected from a wide and lengthy laboratory and clinical experience, and are, some of them, unduplicated in America, so far as I know. A condensed list will be given later. The methods of scoring too, do not seem so well known as Kraepelin's, for instance, perhaps because England and America are more apt to borrow from German than from French sources.*

There have been since then two types of test series in use, one of a simple nature useful in determining differences of large classes of people, the other of a more elaborate sort, applicable to a study of individual differences within a group, or to stages of development, or in some studies to the elucidation of the tests themselves. Thus epileptics, feeble minded, backward and truant children are studied as different from the normal type; twins, bright and dull children, younger and older children are compared, and individual differences in fatiguability by mental work, etc., investigated by the use of tests.

2. Representative Lists of Tests

By way of comparison some of the more representative lists are here given. They are not all complete, since the purely anthropometric tests have been omitted. It will be noted that a given test such as cancellation or tapping may be differently classified by different investigators.

Cattell's list, for students at Pennsylvania includes—

Rate of movement— of hand and arm through 50 cm.
Least noticeable difference in weight—lifted pairs (similar to Galton's test).

* After the experiments to be reported in this study had been made, there appeared Burt's article in the *British Journal of Psychology,* 1909, on "Experimental Tests of General Intelligence" and Whipple's "Manual of Mental and Physical Tests." The former contains four new and interesting tests, and an elaborate treatment by the method of correlation. The latter is exactly what its title would indicate. Besides minute and explicit directions for administration and statistical interpretation of the fifty-four tests described, the published norms and extensive bibliographies are particularly helpful. The present study is a more specific attempt to determine relative values in the case of certain of the tests from which on the basis of general experience and a critical survey, Professor Whipple has chosen his standard series.

Finally there are now being published reports of the Committee on Tests of the American Psychological Association, which began its work in 1907. So far three studies have been reported: "Methods for the Determination of the Intensity of Sound," by W. B. Pillsbury; "The Measurement of Pitch Discrimination," by C. E. Seashore; "The Determination of Mental Imagery," by J. R. Angell; all in Monograph Supplement No. **53** of the *Psychological Review,* December, 1910.

Reaction time for sound.	
Time for naming colors—	ten colors.
Space judgment—	bisection of a 50-mm. line.
Time judgment—	equate an interval to a 10-sec. standard.
Memory and attention—	number of letters correctly repeated after one auditory presentation.

JASTROW'S list for students at Wisconsin includes—

Rate of movement—	touching two reaction keys 38 inches apart in natural time.
	touching two keys 3 inches apart in quickest time.
Sense judgment—	estimate an ounce.
	equate two weights.
	estimate 1 inch on the skin.
	estimate position in guided movements.
	equate bilaterally symmetrical free movements.

JASTROW'S list for volunteer subjects at the World's Fair.

Sensibility, of touch—	distances in length.
	kinds of surface.
	weights.
of touch and sight—	bilateral symmetry.
	lengths.
	direction.
	location.
	aiming at a target.
of sight only—	lengths of lines.
	bisection, trisection, etc., of lines.
	number of letters, words, squares, colors, etc., seen in an exposure of 1/20 sec.
Memory—	visual immediate.
	recognition method for colors and forms.
Reaction time.	

This description of the list follows Binet's analysis.

GILBERT'S list for testing school children.

Muscle sense—	threshold for lifted weights.
Suggestibility—	size weight illusion.
Voluntary motor ability, Fatigue	rate of tapping.
Reaction time.	
Discrimination reaction.	
Memory of time.	

OEHRN'S list for 10 subjects.

Perception—	counting letters.
	proof reading.
	cancellation test.

Memory—	time to learn 12 nonsense syllables.
	time to learn 12 numbers.
Association—	adding one place numbers.
Motor—	speed of writing from dictation.
	speed of reading.

BINET AND HENRI'S suggested list.

Memory—	of a geometrical design.
	of 60-word sentences.
	of musical phrases.
	of colors (recognition method).
	number of repetitions needed to learn 12 numbers.
Images—	letter square.
Imagination—	questions as to tastes, etc.
	ink blots.
	suggestion from abstract words.
	coordination of a theme.
	completion of a drawing.
	construction of many sentences with given nouns or verbs.
	a ten-minute theme on a given subject.
	development of a musical theme.
Attention—	regularity of reaction times.
	reproduction several times of a line seen once.
	speed at which two metronomes at different rates can be counted.
	simultaneous reading and writing of different content.
Comprehension—	understanding of simple puzzle mechanisms.
	differentiation of synonyms.
	criticism of absurdities, fallacies.
Suggestibility—	an increase-in-length-of-line trap.
	discrimination of odors (odorless flasks).
	name and unannounced sensation from imposing-looking apparatus (none given).
	apprehension at second, slow trial of algometer.
	involuntary movements.
Æsthetic choice—	constancy in selection of rectangles, colors, etc.
	series of musical phrases.
Moral feelings—	kind of reaction to one photograph of brutal horrors included in a series of neutral scenes.
	behavior at a sudden loud noise.
Force—	dynamograph.
Motor skill—	(vaguely indicated) some form of maze test.
	throwing 10 balls at a target.

It will be noticed that the emphasis is on the qualitative rather

than the quantitative side, even in a series to be given at one sitting only. Following these suggestions, but with repeated sittings there is

SHARP'S list, used with seven subjects.

Memory—	immediate for 12 letters, visual.
	immediate for 12 numbers, visual.
	immediate for words, auditory, disconnected.
	immediate for sentences, short and long, auditory.
	for sounds, by question method.
Images—	letter square test.
	questions.
Imagination—	ink blots.
	puzzle watch and box.
	development of themes.
	questions on suggestions from abstract terms, etc.
Attention—	cancellation (in four variations).
	reading time of concrete and abstract material.
	simultaneous reading aloud and writing.
Observation—	description of picture exposed for 2 minutes.
	memory of colors exposed for 5 seconds.
	comparison of synonyms.
Tastes—	range of information about pictures.
	number of pieces of sculpture, artists, musical composers named in 5 minutes.
	naming one production of each of 10 composers.
	naming an author from hearing a selection read.

STERN'S suggested list.

Type of perception—	things highly colored named in 5 minutes (written).
	things of vivid sound named in 5 minutes (written).
	color recognition, after 10 minutes' interval.
	pitch discrimination with several minutes' interval.
	kind of mistakes in letter square test.
	reproduction of melodies and rhythms after several days' interval.
	estimate of location of a rotating hand on a dial after a given interval.
Memory—	time to learn lists.
	time to re-learn next day, noting accuracy.
	reproduction of an anecdote immediately, next day, a week, a month later.
Apperception type—	reproduction of a story.
	description of a picture, object, etc.

Attention—	distractibility during work from alteration in light.
	distractibility during work from interrupting sounds.
Combination (construction)—	formation of as many words as possible out of a given selection of letters.
Judgment—	suggestibility by weights, odors, changes in pitch.
Natural tempo—	constancy in rate on different days of beating a three-fold rhythm.

BINET'S list, used with his daughters.

Association and imagery—	writing a list of 20 words.
	first idea on auditory presentation of a word with many questions for introspection.
	writing sentences (time before beginning noted).
	completing sentences.
	developing a theme.
	writing down events recalled.
	description of objects.
	description of occurrences (pictures).
Attention—	cancellation test, varied.
	immediate memory of numbers heard.
	number of glances needed to copy figures and lines of prose.
	copying a drawing exposed .07 of a second, number of exposures needed.
	regularity and judgment of reaction time.
Memory—	amount of poetry learned in 10 minutes recalled immediately and 6 months later.
	immediate memory for unrelated words, auditory.
	immediate memory and description of objects seen.
	immediate memory for drawings of objects seen for 20 sec.
	immediate memory of hieroglyphs seen for 15 seconds.
Space and time perception—	reproduction in movement of a given length of line.
	equating an interval to varied standards.

TOULOUSE, VASCHIDE and PIÉRON list.

Memory—	visual, of colors, lines, angles, curves, location of dots in a circle, rates of movement.
	auditory, of tones, chords, arpeggio intervals.
	muscular, of lines, curves, positions.
	verbal, of numbers, letters, words, phrases (auditory).

	objects, pictures of.
	positions, jointed model of a human figure.
	sketches.
	musical, phrases, rhythms.
	logical, of a prose passage, auditory and visual.
	localization, grouped and serial order of 16 printed nouns.

(All the above to be studied by both reproduction and recognition methods.)

	time to learn long lists of numbers and letters, length of retention of lists.
	recognition of words in lists too long to have been learned.
	lists of words with prefix or suffix in common.
Attention—	cancellation test of letters, hieroglyphs.
	reaction time with discrimination, and irregular intervals.
Suggestibility—	algometer.
Perception type (objectivation)—	rate of tapping.
	reaction time to sight, sound, touch.
Association and imagery—	first idea, orally, from a starting word or object drawn.
	words with or without specified letters.
	associate or dissociate of a verb.
	free association, orally, for 30 seconds from a word or object drawn.
Imagination—	spelling words backwards, visual or auditory.
	giving syllables backwards, auditory.
	theme about a picture or drawing.
Abstract synthesis—	species-genus first idea.
Judgment and observation—	detection of absurdities, fallacies, etc. (oral presentation) and in drawings.
Reasoning—	completion of syllogisms.
	criticism of given syllogisms.

CATTELL'S *Columbia* freshmen tests. (* = discontinued.)

Sense discrimination and perception of space and time—	reproduction and bisection of a line.
	pitch discrimination.
	æsthesiometer.
	reproduction of regular rhythm.
	perception of weight (distance).
Memory—	numerals heard, immediate.
	numerals seen, immediate.
	logical, of a prose passage read aloud to them.
	retrospective, of line drawn and bisected, after 50 minutes' interval.
Imagery—	questions.
Motor—	ergometer.

	rate and accuracy of dotting.
	reaction time to sound.
	tremor in drawing a line.*
Perception—	reaction with discrimination.*
	cancellation test.
	naming 100 colors.
Association—	first idea, written.
	(opposites, written).
Æsthetic choice—	color liked and disliked of models shown.
Attention, Apperception, Suggestibility	*

Whipple in his Manual[42] does not propose his list as one to be used in its entirety as an inventory of an individual, but would probably claim, and with much justice, that an adequate inventory would require his 54 tests or more and an expenditure of something like an equal number of hours. His list is not quoted, though it is the most important single contribution of the last decade to the topic, because it is readily accessible. It should be carefully studied by any one whose interests lead him to read the present report.

3. Aim of the Present Study

Without discussing the difference in aim revealed in the character of these series nor the results obtainable by the different methods, this study is concerned with only the usefulness of simple tests now employed or of similar tests designed to supplement or replace them because of greater significance or greater adaptability in content or method. With the exception of one or two association tests all are of the simplest type, and the question raised is, "If this kind of test is the sort frequently used, is it the best of its kind for the purpose?" To answer this adequately would necessitate collecting every simple test of intelligence known and experimenting with it from the points of view of make-up of the test, method of administration, results, change with practise, with maturity, with fatigue, etc.—too long and complicated a task for this study. By limiting the field, however, is caused the main defect of this work. If more of the time which has been spent over the statistics resulting from the data gained had been given in the first place to administering more tests of one function more carefully to more subjects there might be some definite value. Nevertheless, for such as it is, this study is now presented.

My best thanks are due to a friend who assisted in standardizing and correcting 360 pages of one of the cancellation tests, to the three friends who cheerfully served as subjects for so many hours in the

[42] *Op. cit.*

hot summer days of 1907, to N. who also helped in many of the later calculations, and lastly to Professor Thorndike for his ever ready counsel and patient assistance in the revision of both data and treatment.

In general this study is divided into two sections, one in which about 45 different tests repeated on from three to seven subjects are discussed from the point of view of correlation of the tests, change with short practise and reliability of a single trial, the other in which five very different tests practised with nine subjects are discussed from the point of view of change in each, and similarity of changes.

II

EXPERIMENTAL WORK WITH SEVERAL GROUPS OF TESTS

Concerning certain of the tests supposed to inventory an individual's mental functions and measure his differences from the type which are frequently given, as, for instance, the Columbia freshman tests,[43] we are still undecided as to their exact value. We need to know, (1) whether they test fundamental qualities slowly changing by general mental growth and the effects of training in general, or whether they measure degree of attainment in some specialized ability. If large areas of the mind are reached, then much might be predicted from them; if only narrow habits are tested, then little could be predicted from them. One line of evidence is their susceptibility to practise; for a test in which there is much change in a short period of practise is evidently measuring something other than a general function—it might be specialized ability, or the fact of becoming adjusted to test conditions, or the adoption of some device with regard to certain material.

We need to know, (2) in case general qualities can be measured by these tests, whether the test chosen is the best of its kind, the most typical. One line of evidence here is the correlation of different tests all supposed to measure the same thing.

We need to know, (3) how accurately the few trials made, often only one, will measure the function directly tested, how far, for instance, the result may be affected by the understanding of the subject of what he is to do and how he is to do it. The reliability of first trials can be worked out to give light here.

We need to know too, (4) how far results are influenced by differences in the method of administration. Can differences in attitude be made in the subject by varied direction of the attention? Practically the question is—"How could the tests now in use be improved in significance and accuracy?"

The methods at present in use with the students from Columbia and Barnard colleges must of necessity be more or less rough and ready, since only from fifty to sixty minutes are occupied in giving

[43] For a full list and descriptions of these, see Wissler, "The Correlation of Mental and Physical Tests," *Psych. Rev. Mon. Suppl.*, Vol. **3**, No. 6, 1901.

some twenty to twenty-four tests; and in successive years they are given by different experimenters. Some of the subjects, particularly the girls, are too nervous to do themselves justice at the beginning of the hour, a fact which, as seniors, they frequently recall with amusement or deprecation. Comparison in such cases between performance as freshman and as senior will tend to overweight the gain shown in the results of the seniors' tests, and the consequent inferences as to the beneficial effect of college training.

The problems with which this first section deals are—

A. How far is each test susceptible to practise, especially to short practise?

B. What is the value of each test as a measure of the individual's ability in some general function or group of functions, such as memory, association or sensory discrimination?

C. How can we get the best possible measure from a single trial?

In general the procedure was as follows:

1. Three subjects, a highly selected group, made twenty trials of each of certain selected tests during six weeks in the summer of 1907. Of these three, N. had had comparatively little linguistic training, but, on the other hand, had exceptional preparation in psychology, particularly in giving tests similar to these. She was unusually quick in thinking and talking, also in writing and hand movements. W. and F. both had a more inclusive linguistic training, F. particularly so. Both had done graduate work in psychology, not including, however, much work of this nature. W. was somewhat variable in speed, F. was rather slower on the whole, with two notable exceptions, and was the least likely of the three to be put out or upset nervously. Conditions were made as uniform as possible during the tests, and record kept of the weather and temperature conditions from day to day. The association, perception, and memory tests were practised by the three subjects in a group. The discrimination and motor tests were practised by each separately, as individual attention and timing were necessary. The group work took about three quarters of an hour daily, the individual work from 20 to 30 minutes for each subject. The last two sets of trials were made under rather forced circumstances, as it became necessary to complete the twenty sets a little earlier than had been expected. The general trend of the practise curve was not affected however.

2. From experience with this group, called the "long-term-practise group" for convenience, certain of these tests, along with others supposed to be of a similar nature or to test the same mental process, were repeated in the spring of 1908 with a larger group of

subjects varying from six to eight members. These were junior or senior women students in Teachers College, four rather young, three rather more mature, and one man, some of whose records in the association tests had to be omitted owing to some difficulty with the English language. As much as possible was done with these subjects working in a group, for which purpose they met once a week for two hours for six weeks. They made from two to ten trials with different tests. Later, each came alone for work with some of the tests requiring special apparatus or individual attention. These subjects are referred to as the "short-term-practise group."

3. Certain random groups of college students were used either as opportunity offered or definitely in order to procure a larger number of control cases. One such group of nineteen summer session students spent an hour in 1908 in taking various association and perception tests; another group of similar size in the winter term spent half an hour on some of the tests. These have been called the "instructed group." Single tests are frequently given to large groups for demonstration purposes, and where available, these records have been utilized to get a standard average and deviation for maturer students working in a group. These are referred to as "control cases."

In discussing the work each test is taken separately and report made, first of general experience with the test, including the freshman results for men and women, then of the instructed group, men and women separately where so distinguished, next of the short-term-practise group, last of the long-term-practise group. Thus there is quoted first the result as found by the present test and method; next the results from more mature students, sometimes by a slightly different method; then the change taking place in naïve mature subjects with only a few repetitions; last, what change may take place even in habituated, mature subjects with more extended practise.

A test in which there is not much change will, other things being equal, be the more reliable to use for a single trial with naïve subjects. The "other things" must of course include ease with which directions are understood, simplicity of required reaction, and freedom from all pitfalls or traps for the well-intentioned but unwary subject.

For each group of tests the questions of change by practise, intercorrelation and precision are then taken up and recommendation made of one or another of the tests tested.

1. Tests on Association

A. Descriptive

The first group of tests to be reported on will be those on association.

The Columbia freshmen are given one test only, the *first idea*, the Barnard freshmen that and an *opposite* test.

First Idea.

This consists of the blank given below.

House
Tree
Child
Time
Art
London
Napoleon
Think
Red
Enough

(*N. B. This and many other blanks appear here in reduced size.*)

The test is explained to the students as one of rapidity in thinking rather than of quality. They are told to write as quickly as possible after each word the first idea—preferably one word—that occurs to them. Practise is given orally with a sample word, then the students are handed the blank. The time taken to finish the blank is taken on a stop-watch, and the blank is filed.

One's common observation in giving this test to the freshmen is that it is particularly hard to follow the directions, and to write down actually the first idea that occurs on reading the word. Subjects will sit blankly, stopped by a word, obviously choosing the fittest of several ideas, however well it may have been explained to them that it is primarily a test of the rate rather than of the quality of thinking. The averages calculated from 250 Columbia and 100 Barnard freshmen show that the men take 55.4 seconds to write down 10 ideas, the girls 71.8 seconds. The P.E. for Columbia students is 22.9, quite the largest P.E. found for any of the freshman tests. To make these figures easily comparable with those to be given for subjects after short and long practise, they may be put thus: in 15 seconds men, as tested in the regular manner, wrote 2.7 first ideas, girls wrote 2.1 first ideas, or the average time to call up and write one idea is 5.54 seconds for men and 7.18 seconds for women. In this test then, the girls seem specially hampered; for the results of other tests of the rate of association, such as adding, and giving the opposites of words show no such superior speed for males.

The method used in the present investigation was to explain very carefully just what was wanted, giving oral practise with two sample words. Subjects were told to begin at the signal "go" and get as much as they could done till the signal "stop" was given. They were warned that they would not have much time, though the actual number of seconds was not told them in advance. (The three subjects who took the long term of practise soon came to know the time allowed for the different tests.) For the *first idea* test, the time-limit was 15 seconds. The score was kept in number of words written. Three letters counted as a word if the subject could explain that he had surely thought of something.

A single trial with 37 unpractised subjects, 19 men and 18 women, with the time-limit of 15 seconds gave an average of 5.6 words written, with an average deviation of 2.19 or an average of 2.68 seconds to call up and write a word. The men and women had exactly the same average, but the A.D. for the men was 2.58, for the women 1.78. Unless then, the apparent sex difference in the freshman results is due to difference in the relative immaturity of the subjects, it may be produced by the method of giving the test. (For convenience, the method by which a subject is told to work as quickly as possible and the time taken to finish the test is noted will be called the "amount-limit" and the method by which the subject starts and stops at a given signal, and a certain time-limit unknown to the subject beforehand is allowed, will be referred to as the "time-limit" method. The latter has obvious conveniences in testing groups of subjects.) In each test where both methods were used, comparison will be made of the results by each method, and a special section devoted later to a summing up of these results.

By the amount-limit method 2.7 first ideas were written in 15 seconds by the men, by the time-limit method 5.6; by the women the averages are 2.1 and 5.6 respectively. These differences suggest first, that the amount-limit method leaves the test ambiguous, the time being a measure partly of slowness in associations and partly of associations called up and rejected; second, that a time-limit acts as a spur, making subjects work more quickly than if simply directed to write as quickly as possible, and making them less fastidious in selection of associations when speed is so much emphasized. It is known that "controlled association time" is often shorter than free association time, the theory being that the setting of the attention and judgment beforehand holds certain paths open for use more readily than others; it may be then that attention is aided in a somewhat analogous fashion by the incentive to do as much as pos-

sible in a given time. The anticipation of the signal "stop" seems to give a more definite aim than merely one's best effort after speed.

TABLE I

	Words written in 15 seconds		Seconds required per word	
	Men	Women	Men	Women
Amount limit	2.7	2.1	5.54	7.18
Time limit				
Instructed	5.6	5.6	2.68	2.68
reversed		4.6		3.26
Short 1st		7.0		2.14
Short Average		7.85		1.91
Short 4th		8.2		1.83
Long 1st		7.0		2.14
Long Average		7.83		1.92
Long 20th		8.6		1.75

It was, however, suggested, that the list of words as printed lent itself to higher scores by the time-limit method than by the amount-limit, as the more concrete words come near the beginning, and the most difficult are the three last. To test this point, the list was typewritten in reverse order and then used as a time-limit test with two other groups of students, 29 rather young women, and 34 in a mixed group of men and women somewhat older. The average number written in 15 seconds was 4.6 words. Asked to repeat the test commencing with the bottom word, the average in 15 seconds was 4.8 words. Thus the greater speed does not seem to be entirely due to the kind of words encountered at the outset.

In the short term of practise, 4 trials on different days by 6 subjects by the time-limit method, the average was 7.85 first ideas written in 15 seconds, or 1.91 seconds per word. In the long term of practise, 20 trials by 3 subjects, the average was 7.83, or 1.92 seconds per word. The number written at the first trial by each group was 7.0. Taking all the trials of these two groups into account, 85 in all, there were 14 occasions, or 16 per cent. of the total number, when the test was completed in 15 seconds. The two lowest records, made only once each, were 3 and 5 first ideas, both considerably higher than the freshman results by the amount-limit method.

The difference appears even more striking when the fairly constant factor of speed of writing is discounted. Three subjects were given six trials each in writing ten words of some familiar sentence* under each other in a vertical column. The average time for the 18

* Two clauses from the Lord's Prayer: (1) Our Father, etc.; (2) Lead us, etc.; and (3) "Little Jack Horner sat in corner eating his Christmas pie." The number of letters were 40, 43, and 48.

trials was 13.38 seconds or 1.34 seconds a word. Thirty subjects, naïve except for an hour's work in other tests, were asked to write a single word similarly with a time limit first of 10 seconds, then of 15 seconds. Half of them wrote the word "watch" in the 10-second test, the word "father" in the 15-second test; the other half wrote "father" in the first test, "watch" in the second. The results were for the 10-second test 5.1 words, for the 15-second test 7.75 words, or an average time of 1.95 a word or .355 second a letter. Thus the average extra time needed for association over mere writing is, in the case of the amount-limit method, about five seconds a word; in the case of the time-limit method less than 1 second a word.

In absolutely free association—*i. e.*, when a starting word only was given and the subjects wrote down whatever series of things they thought of, an average of 11.5 words was written in 15 seconds, or at the rate of 1.31 seconds a word. (Incidentally it is interesting to note that serial connections are more rapidly written than even the same word in repetition, thus:

Familiar sentences, 3 subjects, 18 trials, 1.34 seconds per word, .307 per letter.
Free association, 6 subjects, 30 trials, 1.31 seconds per word, .240 per letter.
"Father" or "Watch," 30 subjects, 60 trials, 1.95 seconds per word, .355 per letter.

though this difference is partly due to the fact that the 18 trials came from 3 practised subjects on different days, the 30 trials from 6 subjects after the short term of practise, the 60 trials from 30 subjects after 1 hour's work with various tests.)

It seems certain then that the *first idea* test, as usually given, does not measure the rate of association. Nor apparently can any test involving the writing of words do so. For not only is the average rate of mere writing no less per letter than the average rate of writing words under some associative requirement, but in certain cases where the description of the association involves writing a phrase or long word such as "eyes, nose and mouth," "kerosene oil" or "pussy-willow," the writing time entirely obscures the association-time.

Considering it from the point of view of practise, in the short irregular practise with the average score of 7.85 the fourth trial showed a gain of 1.17 or 17 per cent. over the first. With the three subjects who repeated the test twenty times there was a practise gain of 1.6 or 23 per cent.

In the five trials with the absolutely free association test there was quite the reverse of practise effect. The starting words used at

the five trials were, respectively, *house, read, black, table, ball.* The average amount done in 15 seconds was 11.5 words, or one word in 1.31 seconds, the deviation of the first trial from this 11.5 being + 1.8, of the fifth — .8.

The correlation of the *first idea* test with other association tests will be taken up later.

Opposites Test.

In giving this test the usual experience is that some words are uniformly hard, and that when once at a loss for the opposite to any word that has presented difficulty, an enormous amount of time may be spent. Some subjects will go on writing the easier ones, returning afterwards to those that have proved puzzling. If these have been retained subconsciously there is probably a saving of time. Usually no hint is offered about "skipping" in this way to the freshmen, though where this test has been used in group work with children and others, with a time limit, usually no skipping is allowed. It then becomes impossible to know how much of the time is spent over perhaps one word in the list, so that the final record is very much affected by the inherent difficulty of the test-words. The standard set prepared by Woodworth and Wells* is not in common use yet, and the Columbia set presents several difficulties. It is as follows:

Write as quickly as you can beside each word in the column a word which means the opposite thing from it.

barbarous
simple
rude
obscure
gentle
to expand
elation
adroit
loquacious
to degrade
to hinder
precise
permanent
repulsion
to respect
genuine
separate
deceitful
grand

* To be reported as a publication of the American Psychological Association.

Other sets used in comparison were:

OPPOSITES TESTS

I

day	vertical	right	good
asleep	to spend	love	outside
absent	to reveal	rude	quick
brother	level	just	tall
best	ignorant	lie	big
above	past	tidy	loud
big	part	cruel	white
backwards	motion	run away	light
buy	to hold	best	happy
come	generous	quick	false
cheap	proud	remember	like
broad	diligent	dressed	rich
dead	stupid	to be hit	sick
land	serious	lose	glad
country	frequently	mend	thin
tall	weary	disobey	empty
son	wicked	clean	war
here	to create	noisy	many
less	to enrage	rough	above
mine	stormy	cross	friend

II

serious	high	great	vertical
grand	up	hot	ignorant
clumsy	wet	dirty	rude
to win	new	heavy	simple
to respect	soft	late	deceitful
frequently	wider	first	stingy
to lack	wrong	left	permanent
apart	yes	morning	over
stormy	young	much	to degrade
motion	laugh	near	weary
forcible	winter	north	to spend
to float	weak	open	to reveal
straight	forget	round	genuine
to hold	wild	sharp	level
after	beginning	east	broken
unless	straight	known	wild
rough	raise	something	part
to bless	rough	stay	past
to take	love	push	permit
exciting	noisy	nowhere	precise

In scoring these, a mark of 2 was given for the best choice, 1 for a second best choice, and 0 for a bad choice. The key used in scoring will be found, alphabetically arranged, in the appendix. From the very fact that so many words could be offered as opposites to certain

given words, it will be seen how valuable a standardized set would be. In the various tables that follow a score for accuracy is given in terms of the per cent. which the score given to the individual in question was of the score he would have received had every opposite written by him been rated as worth 2 credits. Thus a record of five opposites valued as 2, 0, 1, 1, 2 respectively is scored 6/10 or 60 per cent.

First, to compare the various blanks used. Columbia freshmen have not been put through this test. Barnard freshmen have usually taken the "barbarous" blank, though 14 were given "vertical I." "Barbarous" took 166 seconds on the average or 8.74 seconds per word compared with 105 seconds, or 5.25 seconds per word for "vertical I"; the scores for accuracy were (average) 69 per cent. and 72 per cent. respectively. The short-term practise group who also worked with each blank, and by the same method, took 141 seconds, or 7.42 seconds per word for "barbarous," and 89 seconds, or 4.45 seconds per word for "vertical I." Their average scores were 69 per cent. and 71 per cent. Thus the difference in time taken shows that the "barbarous" blank is more difficult than "vertical I." The average score for "barbarous" is also lower than that for any other blank, as may be seen from Tables II and III. An easier blank, such as "serious" or "day" would probably be more suitable for this type of subjects.

TABLE II

SPEED AND ACCURACY IN WRITING OPPOSITES

	"BARBAROUS"			"VERTICAL I"			"VERTICAL II"		
	Time required in seconds for 19 words	Seconds required per word	Av. per cent. of max. credit obtained	Time required in seconds for 20 words	Seconds per word	Av. per cent.	Time required in seconds for 20 words	Seconds required per word	Av. per cent.
Freshmen	166	8.74	69	105	5.25	72			
Seniors	93	4.89	69				96.2	4.81	71
Short-term ...	141	7.42	69	89	4.45	71			

So far as these blanks reveal differences in maturity, there is a decided improvement in speed with more mature subjects; the freshmen take a longer time than the short-term group at their first trial with both the difficult blanks, and considerably longer than the seniors. The accuracy is practically the same for all these three groups on the same blanks. Looking also at Table III, all the records from the short-term group are poorer than even the first record of the more mature long-term group for "vertical II" which is a

fairly difficult blank, though the easier blank "day" seems too easy to show differences in the groups of subjects. In this table all the records are reduced to the amount done in 30 seconds, and the accuracy score to percentage, whether the test was by amount-limit or time-limit method, and no matter what the blank.

To compare differences in method, a group of Barnard seniors were given "vertical II" by the amount-limit method, and a group of Teachers College women students the same blank by the time-limit method, with scarcely any difference in the results, though what there was, was in favor of the time-limit method, as will be seen by Table III. These two groups were of about the same maturity, but again with the slight difference in favor of the Teachers College students, so that either this factor, or that of difference in method may be responsible for the very slight difference in the figures.

TABLE III

SPEED AND ACCURACY IN WRITING THE OPPOSITES OF GIVEN WORDS

Speed is measured by the number of seconds required per word. Accuracy is measured by the average per cent. of the maximum credit that was obtained.

	"BARBAROUS" TEST		"VERTICAL I" TEST		"VERTICAL II" TEST		"SERIOUS" TEST		"DAY" TEST	
	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy
Amount limit										
Freshmen	8.74	69	5.25	72						
Seniors	4.89	69			4.81	71				
Short term	7.42	69	4.45	71						
Time limit										
Instructed					4.62	73			2.36	93
Short term 1st					4.48	70			2.21	91
Short term last					4.55	75			2.03	94
Long term 1st					3.23	91	3.13	86	2.50	94
Long term Average					2.48	88	2.22	88	2.19	95
Long term 10th trial					2.17	89	1.76	90	2.07	94

To test the effect of practise, the short-term group were given six different tests, the "day" being repeated after six weeks, giving 7 trials in all with the time-limit of 30 seconds, also "vertical II" once with a time-limit of 30 seconds. The Columbia blanks were given on the fifth day by the amount-limit method, so that a total of 10 trials was made by this group of subjects.

Since the "day" test when repeated after practise with "good," "great," "vertical," and "right" shows so little gain the practise effect is very slight, and the test continues to be an association test rather than a series of specially trained responses.

Even special practise with the same blank shows rather slow im-

provement. The long-term group used three blanks only, "day," "serious," and "vertical II." After the first two trials these were used in rotation till it was evident that the easy "day" blank had been memorized. The other two were used ten times each, on alternate days, and beginning alternately at the top and the bottom of the column. There was, of course, a gain in speed, the time per word being reduced from 3.23 to 2.17 and from 3.13 to 1.76 in the 10 trials, but the rate is still much above that for writing the numbers from one to twenty or other familiar series.

Comparing this test with the *first idea* in rapidity, it will be seen that this form of controlled association does take slightly longer with subjects practised with both tests.

TABLE IV

SECONDS REQUIRED PER WORD TO WRITE (1) The First Idea Called up by a Printed Word, (2) A Series of Words Started by a Printed Word, and (3) The Opposites of the Words of the "Day" Blank

	(1)	(2)	(3)
Time limit			
Instructed group	2.68	1.31	2.36
Short-term group	1.91		2.11
Long-term group	1.92		2.19

Other controlled-association tests used in comparison with this were: for the "instructed" group, two in number, the *preceding letter,* and *complete the word;* for the "long-term" group, six in number, these two and also the *subject predicate, difference between, Ebbinghaus combination,* and *addition;* for the "short-term" group, the first five given above, a different set of *addition and subtraction, noun and adjective, nonsense words,* and one or two *nonsense sentences, genus species, multiplication.* They will be taken up in that order.

Except where otherwise stated, these were always given by the time-limit method.

Preceding Letter.

The series of stimulus letters is as follows:

f
k
s
p
w
l
e
r
d

o
v
j
n
t
h

The time-limit was 15 seconds. The subjects were told to "write beside each letter the letter which precedes it in the alphabet," oral examples being given by two letters. With 197 subjects, one trial, the average number written was 5.5 letters, a clear mode of 5, a range of from 0 to 12 and an average deviation from the mode of 1.6. One letter thus required 2.73 seconds (Av.) or 3 seconds (Mode). Introspective evidence shows that this is a peculiarly difficult test to start right in spite of the preliminary oral practise. Old habit asserts itself to such an extent that many subjects are unable to react at all without mentally repeating the whole of the alphabet up to the test letter. Others try to repeat it backwards; others to make use of visual imagery. If this is the first test given in an hour's work on various tests, it seems particularly bad. When it is the sixth or seventh test given, the average on three different occasions with small groups, making 36 subjects in all, was 6.1 letters in the 15 seconds, or 2.46 seconds per letter, with an A.D. of 1.2.

The short-term group used it three times with an average of 7.3, the first day's average, 5.6, deviating by -1.7, the last by $+1.0$, showing a very decided practise effect for so few trials. The long-term group made averages of 7.3 letters or 2.05 seconds per letter, 6.3, or 2.05 seconds per letter, 8.6, or 1.74 seconds per letter, and 9.3, or 1.61 seconds per letter, in their first four trials. They were also very variable throughout the entire 20 trials. This test then seems to be a specially bad one.

Complete the Word.

The form of the test was as follows:

1. ri
2. bon
3. mil
4. la
5. flo
6. chi
7. dr
8. fas
9. sk
10. bra
11. med
12. bus
13. spo
14. gam
15. an
16. che
17. chu
18. we
19. rec
20. par
21. chap

Fifteen seconds was allowed. Eight subjects used it three times, and the three subjects ten times, beginning with the first or second column or at the end, after which they made ten more trials with fresh sets.

In a first trial it is very noticeable that a subject may think of long words in the beginning, and continue to think of them even when shorter words are completed in the spelling out of the word actually written, as "ri" suggesting "ribbon" when "rib" would suffice, or when cognates would be shorter, such as rite for ritual. At the same time it is introspectively an easier test than the *first idea,* because, in the first place, the subject seems to be less suspicious of what may be demanded of him, and feels more free to write down what he has actually thought of; in the second place, parts of words seem to be more suggestive of whole words than one word is of another, perhaps for two reasons; first the conditions are more like ordinary reading, second the motor or auditory imagery or perhaps the incipient movements of the speech organs seem to perform the task of completion automatically, while all the judgment has to do is to acquiesce. With both this and the absolutely free association test, the factor of long words may increase the time taken through the mere mechanics of writing. The statistical results will favor those who think of short words as well as the rapid thinkers.

For the "instructed" group of 37 subjects the average number of words completed in 15 seconds was 8 (1.88 seconds per word), with a range of from 3 to 15, and an A.D. of 2.8.

TABLE V

NUMBER OF WORDS COMPLETED IN 15 SECONDS

	No. of subj. Men	No. of subj. Women	No. of trials	Av. No. written Men	Av. No. written Women	A. D.	Sec. req. per word Men	Sec. req. per word Women
Instructed group	19	18	1	8.2	7.7	2.8	1.83	1.94
Short-term group (using the same blank):								
1st					9.5			1.58
average		7	3		9.1	2.0		1.65
last					11.4			1.31
Long-term group (using different blanks):								
1st					9.3			1.61
average		3	10		10.5	.8		1.43
last					11.8			1.27

The short-term practise group in three trials made an average of 9.1 words completed or 1.65 seconds per word, with a range of from 4 to 15 and an A.D. of 2.

The long-term practise group averaged 10.6 words in 15 seconds or 1.42 seconds per word in their first trial. After 10 trials with the

same blank, improvement being very rapid, 10 more trials were made, with two or three from the original blank introduced into each set. The average was then 10.5, ranging from 9.3 on the eleventh day to 11.8 on the twentieth, showing a slight practise effect. Had the word beginnings been absolutely new, the practise effect would presumably have been still less.

Six of the short-term practise group later took this test orally by the amount-limit method. Eight trials were made with different lists. In this way it could be seen how a poor record is made by the influence of some one combination which halts a subject unduly long rather than by slowness in general. One list seemed easy for all subjects, but no one list was hard for all subjects; one or two exceptionally poor records occurred with every list. The combination "um" halted three subjects a comparatively long time. One subject made the worst record 7 times out of the 8, though in the written test by the time-limit method she had been one of the best subjects. Introspectively, all preferred the oral method. Compared with other tests, *completing words* is less disturbing than the *first idea*, but less definite than the *opposites*.

Subject-predicate.

As a test this is not in common use, so that the blanks were prepared in round handwriting, which may have retarded the speed somewhat as compared with the *first idea* and *opposites* tests, which were printed. Mimeographed sets were later used for the short-term practise group.

SUBJECT-PREDICATE LISTS

convenes	matriculates	stings	brays	confesses
butts	scratches	parries	steals	lubricates
explodes	earns	waxes	preaches	hatches
hops	bleats	prescribes	plays	disperses
sucks	illuminates	swims	arrests	reverberates
plants	paints	enlists	lectures	hoards
chases	flies	buys	flashes	smoulders
alleviates	experiments	quacks	rings	ordains
extinguishes	strikes	applauds	fights	nourishes
re-acts	reaps	sews	condemns	sneers
ebbs	cackles	navigates	graduates	performs
composes	inherits	freezes	burns	sells
shoots	learns	riots	drives	amputates
bites	blows	sues	cleanses	neighs
stitches	testifies	disbands	crows	rotates
trumps	owes	governs	calculates	fades
shines	adjourns	roars	haunts	bets
hammers	sings	occurs	melts	tolls
marries	sacrifices	raves	limps	foretells
trots	flows	surrenders	withers	barks

Subjects were warned not to supply a subject by forming a noun in "er" from the verb such as "sing*er*" sings, nor by using indefinite words as "man," "boy," but to supply the definite agent such as "bird." Two or three examples were illustrated. One hundred verbs were made up in ten sets of ten, each being used twice for the long term of practise, and once each on typewritten sheets for the short term of practise. Unfortunately for strict comparison they were not given in the same order for the short practise as for the long. The scoring for accuracy was done as for the *opposites* test, giving 2 for the best choice, 1 for a poorer one, 0 for a poor one.

TABLE VI

N. = number of subjects written to fit given predicates in 20 seconds.
Acc. = per cent. of maximum credits obtained.

Order given	1		2		3		4		5		6		7		8		9		10	
Tests	confesses		ebbs		cackles		navigates		brays		convenes		graduates		performs		stings		matricula	
Subjects	N.	Acc.	N.	Acc.	N.	Acc.	N.	Acc.	N.	Acc.	N.	Acc.	N.	Acc.	N.	Acc.	N.	Acc.	N.	Ac
Bu.	10	75	9	100	9	89	6	92	8	100	8	99	6	92	9	100	4	75	10	
Gr.	2	25	6	100	7	71	2	100	5	100	5	90	6	100	7	64	4	75	10	
J.	4	63	6	67	7	71	6	83	8	88	6	50	7	86	4	100	5	30	7	
L.	5	70	6	100	7	79	7	71	5	100	4	63	5	80	7	64	6	92	7	
M.	5	30	6	100	5	60	3	33	5	100	4	100	3	100	8	88	5	50	7	
Ba.	10	65	9	44	9	33	8	75	9	89	7	86	8	63	10	55	10	30	10	
Bf.									8	100	5	80	7	100	8	100	8	69	8	1
Averages .	6		7		7.3		5.3		6.8		5.5		6.0		7.5		6		8.4	
Medians ..		64		100		71		79		100		86		92		88		69		

TABLE VII

N. = number of subjects written in 20 seconds to fit given predicates.
Acc. = per cent. of maximum credits obtained.

	First trials 1-10		Second trials 11-20	
	N. Av.	Acc. Median	N. Av.	Acc. Median
performs	5.6	100	9.1	100
stings	7.0	94	7.1	93
matriculates	6.6	93	7.1	100
ebbs	7.0	94	8.0	94
brays	8.3	95	8.8	100
cackles	7.6	94	7.3	100
convenes	5.6	100	8.0	88
navigates	7.0	81	8.6	94
graduates	6.3	93	7.1	100
confesses	8.8	100	8.6	95
Average	7.0	95	8.0	96

The results for the short-term group are shown in Table VI. The practise effect is apparently very slight, the last five tests being only a trifle better in speed or accuracy. Further tests are, however, needed to separate the influence of differences of the tests in diffi-

culty from that of practise, and from that of the chance variations in the subjects.

The results for the long-term group are summarized in Table VII. The practise effect of ten trials, including one of the same blank, is in general to increase the speed only by a seventh, leaving the accuracy uninfluenced.

The time required in these tests is about the same as that in the difficult "vertical" opposite test.

The "Difference Between."

The form of the test used is as follows:

Answer these questions as quickly and as well as you can.

1. What is the difference between *grab* and *take?*
2. What is the difference between *eat* and *devour?*
3. What is the difference between a *stream* and a *river?*
4. What is the difference between a *wagon* and a *cart?*
5. What is the difference between *sorry* and *sad?*
6. What is the difference between *naughty* and *bad?*
7. What is the difference between *homely* and *ugly?*
8. What is the difference between *right* and *correct?*

Other lists used were:

II

confess, reveal
confine, limit
colleague, partner
bend, curve
resistance, opposition
deceive, mislead
adrift, afloat
extend, increase

III

above, over
demonstrate, illustrate
deluge, flood
guardian, keeper
merry, gay
bring, fetch
heavy, weighty
innocent, harmless

IV

show, indicate
watch, observe
trial, test
contract, bargain
peace, repose
clear, obvious
cleanse, purify
classify, arrange

V

get, provide
win, gain
pair, two
parcel, bundle
womanish, feminine
put, place
boat, ship
clever, talented

VI

chuckle, giggle
honest, honorable
procure, obtain
haste, hurry
crayon, chalk
antagonist, opponent
puff, swell
abrupt, blunt

VII

walk, march
ignore, overlook
corpse, carcass
early, soon
allude, refer
drag, pull

VIII

walk, march
deceive, mislead
corpse, carcass
colleague, partner
drag, pull
adrift, afloat
try, test
extend, increase

The subjects were told that the quickest way to answer was either to explain one word in terms of the other, or to write 1 = —— 2 = ——, not wasting time by repetition. Notwithstanding this, many to whom it was given used an unnecessary number of words in explanation, thus taking longer to write. From the point of view of time consumed, then, it is not a useful nor a satisfactory test whether given by the time-limit or by the amount-limit method. Not only association and speed of writing enter in, but the ability to profit by the advice in the instructions, and ability to condense—also, of course, linguistic discrimination. This test is, besides, not very easy to score, as the answers may vary considerably.

Blank I was kindly filled in at leisure by one of the professors in the English department. Answers were then compared with these standard answers and each of the eight scored 2, 1 or 0, as in the case of the *opposites* and *subject-predicate* tests. For the remaining blanks, dictionaries and books of synonyms were resorted to for standard answers, or, failing anything sufficiently discriminating there, the experimenter's own judgment of the best answer in the group was followed.

An "instructed" group of about 200 were tested with Blank I, time-limit of 120 seconds. In 49 of these chosen at random the average number of answers written was 4.4, with an A.D. of 1.08 and a range of 2 to 8. The average score for accuracy was 89 per cent. (reliability 1).

The short-term practise group took this test only twice, using Blanks I and VIII. The reason more time was not spent with them on the various blanks was that previous experience with the long-term practise group seemed to indicate that the test was not a valuable one. For the same reason and also because the 49 control cases from the "instructed" group were in terms of time-limit, this group were tested by the amount-limit method. Their record for Blank I was: average time taken 217 seconds, score for accuracy 73 per cent.; for Blank II, 233 seconds, score for accuracy 63 per cent.; for both blanks together, average time taken, 225 seconds, A.D. 25.5, average

score 68 per cent. For them, then, Blank I was easier since they made a better showing with it, although it was the first one given.

An "instructed" group of 49, tested with Blank I, with a time-limit of 120 seconds, averaged 4.4 answers written, A.D., 1.08. The average accuracy was 89 per cent.

The long-term practise group used seven different blanks altogether, each one three times except the last, beginning with the 1st, 3d, or last of the 8 pairs of terms. A time-limit of 60 seconds was allowed. Their average for Blank I was 4.6, score of 66 per cent. The average number written for all 20 trials was 3.2, the first day's average deviating by + 1.4, the last by + .4. The average score for accuracy was 70 per cent., the first day's average deviating by + 6 per cent., the last by + 3 per cent. Thus the difference in the difficulty of the blanks again disguises any practise effect. If the records of the first three trials which were made with Blank I are omitted, the average number written is 2.7, the fourth day's average deviating by — .7, the last by + .9, so that there seems a slight gain in speed. The average score for accuracy is then 77 per cent., the fourth day's average deviating by — 2 per cent., the last by — 4 per cent.

Nothing can be surely inferred from these records save that for them less than 20 seconds sufficed to think of and write out a difference (only 13.1 seconds for Blank I). A much longer time limit should have been given.

On the whole, as will appear when the facts concerning correlations and reliabilities are given, this test, if useful at all, is useful only as a specialized measure of linguistic knowledge and facility in expression. The times 27.3 seconds per difference for 49 subjects using Blank I, 27.1 seconds per difference for 6 subjects using Blanks I and VIII, and 18.8 seconds for 3 subjects using Blanks I–VII, show that an elaborate process of selective thinking is involved.

Ebbinghaus Combination Test.

This test was as follows. For the short-term group certain paragraphs of convenient length, averaging 100 words, were chosen from such varied materials as newspaper reports, scientific articles, essays, novels, narrative poems. These were typewritten, with 10 to 16 words, according to the length of the paragraph, omitted in various places, blank spaces being left in their stead. One such paragraph was placed before the subject, who was instructed to write down an appropriate word for each space. The time taken was noted, and a score was made of the suitability of the words supplied in terms of per cent. of a perfect record. Five of the short-term practise group

took ten such tests, repeating the first paragraph used at the 10th trial three weeks later.

In general, subjects will either skim two thirds to the whole of the paragraph at the outset, going back to fill in the spaces, or they will rush at the first phrase, fill in the first thing that occurs, and get tangled up before the end of the first sentence unless the subject matter is very easy. From one or two such experiences the subject is generally led to adopt the other method.

The short-term group took an average of 103 seconds to complete a paragraph, with an A.D. of 32. Comparing their two trials (three weeks apart) with the same paragraph there was an improvement in average speed from 173 seconds to 71 seconds, the A.D.'s 33 and 6 respectively. Their accuracy rose from 70 per cent. to 80 per cent. or, omitting one subject who seemed very much upset at the first trial, it was 80 per cent. on both occasions.

The long-term group was tested with 20 paragraphs averaging 92 words long, each with ten words omitted; they averaged 80.2 seconds, A.D. 18 seconds. Variations of 10 per cent. or less in the length of the passage caused no appreciable differences in the time required. Variations in the content are very influential. The poetry was difficult for these subjects, the average time for that being 108 seconds. Newspaper reports were easy, the average time for them being only 54.4 seconds. Picking the first trial of each kind of material, and comparing it with the last of each, there was an improvement in speed from an average of 104 seconds to 89 seconds. These figures do not measure practise with surety, owing to possible variations in the difficulty of even the same kind of material. The average accuracy was 87 per cent. with no discoverable practise effect. The paragraphs they used are given in the appendix.

In general it appears that adaptation to the form of problem set by the Ebbinghaus test is likely to count considerably, especially with untrained subjects.

Addition.—The blank used was as follows:

ADDITION EXAMPLES

17	26	27	72	23
42	51	24	14	47
38	47	83	39	86
91	82	19	81	54
54	63	45	26	36
17	42	38	91	36
26	51	47	82	26
27	24	83	19	45
72	14	39	62	63
23	47	86	54	54

41	53	67	78	86
52	67	86	37	32
86	34	23	96	44
23	78	45	72	36
35	19	67	23	68
45	52	19	45	23
13	86	78	67	72
68	23	67	78	36
77	35	23	37	68
86	67	86	96	39

A score of 1 for each column added was given and 0.5 deducted for each wrong figure in an answer. The time limit was 60 seconds. The results as to rate will be discussed in connection with those of the next test. Since these experiments were made, it has been shown by Wells and Thorndike that even so familiar a process is, under test conditions, subject to adaptation and practise effects. In these subjects these effects were shown chiefly or wholly in the speed of the process. The short-term group averaged 16, 19, and 18 columns, and .5, .67, and 1.33 errors in three trials on February 15, March 7, and March 7. The long-term group gained in twenty trials about 20 per cent. in speed but lost somewhat in accuracy, so that their net improvement was 17 per cent.

Addition and Subtraction.

The short-term group used a blank, given on the next page, from the collection prepared by Woodworth and Wells.

The test consists of adding a certain number to each figure in succession in the column, or subtracting it, as directed, and writing down the result. One column was counted as a test, making 25 times that a given number was added or subtracted and the result written. Twelve such tests were made, six times with a time-limit of 40 seconds, six times with a time-limit of 30 seconds. In cases where a subject completed the series in less than the allotted time her time was recorded. The key numbers were 3, 4, 5, 6, 7, 8, each added in one test, subtracted in another. Four tests were made in succession, the order in which they were given being as follows:

I. 7 added, 3 subtracted — 40 sec.; 4 added, 5 subtracted — 30 sec.

II. 5 added, 7 subtracted — 40 sec.; 3 added, 4 subtracted — 30 sec.

III. 6 added, 8 subtracted — 40 sec.; 6 subtracted, 8 added — 30 sec.

64	72	47	30
49	35	43	56
62	51	35	44
57	30	64	31
68	56	49	37
74	44	67	60
53	36	28	71
67	73	46	48
25	63	55	53
40	47	65	61
61	43	70	36
71	66	41	42
33	69	62	34
38	37	25	39
28	39	40	33
65	32	57	73
41	59	26	38
50	31	68	63
42	60	66	58
58	48	27	32
52	54	51	59
70	46	69	52
26	55	29	45
34	27	74	72
45	29	50	54

As we now know through the work of Browne,[44] Stone,[45] and others, the adding and subtracting abilities are two very different things; also some figures are easier to handle than others, a combination such as $9 + 2$ being different from and easier than $2 + 9$. These facts complicate the issue.

However, it seems clear that adaptation to the test does bring about a practise effect in the first few trials. The speed with $+8$ in the last of the twelve tests is for every subject save Ji. greater than for $+7$ in the first of the twelve.

By any rational estimate also the second day's records are above the first in general, and in the case of all but one of the subjects measured. They were so probably for Bu. also.

Using the easiest set of these additions of a 1 place to a 2 place number ($+3$), we find the time per operation to be Bu., .76 second; Gr., .96 second; Ji., 1.04 seconds; Le., 1.43 seconds, and Mo., 1.43

[44] "The Psych. of the Simpler Arithmetical Processes," *Am. J. of Psych.*, **17**, 1906.

[45] "Arithmetical Abilities . . .," *Col. Contr. to Educ.*, **19**, 1908.

TABLE VIII

RESULTS IN THE ADD AND SUBTRACT COLUMNS TEST FROM THE SHORT-TERM PRACTISE GROUP

A = amount done in time limit.
E = errors. T = seconds actually taken.

Column		1	2	3	4	5	6	7	8	9	10	11	12
Operation		+7	−3	+4	−5	+5	−7	+3	−4	+6	−8	−6	+8
Time limit in seconds		40	40	30	30	40	40	30	30	40	40	30	30
Bu.	A	?	?	?	?	25	25	25	25	25	25	25	25
	E												
	T					23	27	19	24	22	34	24	25
Gr.	A	21	?	22	20	25	24	25	17	25	25	25	25
	E	1		1						1			
	T					34		24		34	36	24	25
St.	A	13	22	11	12	25	21	25	21	25	21	20	21
	E		3		1							1	
	T					38		26		34			
Ji.	A	20	?	22	14	16	11	21	14	21	11	13	9
	E									1			
	T												
L.	A	9	21	16	11	18	17	21	16	24	20	16	20
	E									1			
	T												
Mo.	A	18	18	18	13				13	25	14	12	17
	E										1	1	
	T									38			
Ba.	A	19	12	19	15								
	E		1										
	T												
Bf.	A	17	21	25	13								
	E			1									
	T												

seconds; a median of 1.04 and an average of 1.12 seconds. On March 15 the short-term group was tested with 100 mixed examples, such as $9 + 7$, $8 - 3$, $6 - 2$, $5 + 8$, etc., 70 seconds time being given. The results were Bu., 100; Gr., 100; Ji., 69; Le., 63; Mo., 67; Ba., 64; Bf., 63. Le. made 1 and Ba. 2 errors. The median time per operation was thus 1.04 seconds, as for the easiest addition to a 2-place number. The average time was probably .9 second. In adding in columns with 5 two-place numbers, for example, in which about three fourths of the additions are to a two-place number, and in which the number added is more often harder than easier than 3, the results were, after the first trial, an average of .67 second per operation (median .87 second). Although the average especially is perhaps too low because the number of actual conscious operations was probably reduced by grouping in the case of the more rapid workers, the fact remains that the mere writing time for a two-place number may,

especially with slow writers, be greater than the time required to add a one to a two place number without writing. One has only a choice of evils. Column addition permits grouping and so mixes the rate of association with the power to associate three numbers with their sum in one connection. A test in writing additions and subtractions with two place answers measures the rate of mere writing in very rapid computers or very slow writers.

Noun and Adjective.

Two blanks with 20 adjectives on each were arranged as follows:

I

Complete the following sentences, after the model of the first one, that is, by adding to each a noun at the beginning, and a second adjective at the end—the whole to make sense:

The	hill	is	high	and wooded.
		"	soft	"
		"	cold	"
		"	new	"
		"	smooth	"
		"	red	"
		"	round	"
		"	windy	"
		"	clean	"
		"	bent	"
		"	wooden	"
		"	deep	"
		"	empty	"
		"	narrow	"
		"	loose	"
		"	bitter	"
		"	level	"
		"	stale	"
		"	oily	"
		"	heavy	"
		"	woolen	"

II

Complete the following sentences, by adding a subject and an additional adjective, as in the first sentence:

Her	taste	is	refined	and delicate.
		"	portable	"
		"	unexpected	"
		"	ridiculous	"
		"	interesting	"
		"	imported	"
		"	probable	"
		"	tapering	"
		"	dangerous	"
		"	complete	"
		"	unusual	"
		"	metallic	"
		"	spacious	"
		"	painless	"
		"	excessive	"
		"	seasonable	"
		"	desolate	"
		"	frequent	"
		"	distinct	"
		"	select	"
		"	temporary	"

A score of 1 was given for each appropriate word written, making 40 the maximum score for a test. Sometimes an indeterminate adjective such as "nice" or "long" would be written several times in succession, and the possibility of this detracts from the value of the test. One subject wrote the pronoun "it" instead of a noun, as directed, and so made a low scoring; otherwise this seems an easy test, for the average accuracy score was 38, or 95 per cent.

The short-term group took this test four times only, the first time with a time-limit of 120 seconds, the other three times by the amount-

limit method. The average time taken to finish was 135 seconds, A.D. 27, or an average speed per word written of 3.37 seconds. There was a slight practise effect in speed even with so few tests, but none in the accuracy. It was written more slowly than the *opposite* and *subject-predicate* tests, but this may be due to the arrangement of the blank, and the need of an additional movement of the hand. Blank I. is, so far as the records from six subjects go, much easier than Blank II., taking only about three fourths as long with equal precision.

English and Nonsense.

The following blank was used three times, a time-limit of 60 seconds being given for each section with 3 minutes interval between the sections.

A. Mark the (familiar) English words among the following groups of letters:

nop	yas	jeb	cug	pin	wam	hay	bot	hub	kib
max	dug	faw	rab	sid	ven	mar	pid	baw	moy
mud	yim	nam	lan	ram	rox	fub	hon	tey	deb
pow	was	jig	ges	lud	wid	jom	kus	dix	bag
cay	yut	dam	lax	sor	not	har	vim	pab	fon
tus	rit	kay	bir	wep	bow	lix	mur	seg	voy
sir	pex	heg	rum	gid	neg	fim	tip	loy	dut
wut	tox	gem	ruy	gor	vig	jad	kow	ton	sut
tir	hig	med	fox	bep	nis	vun	dow	gax	can
jup	nun	yow	mig	dat	tar	soy	few	lun	taw

B. Mark all groups of letters in the following list that are not (familiar) English words:

men	sar	bet	won	pox	hus	nib	ket	sum	hip
tug	mop	jaw	bux	cub	gas	pay	rib	her	num
vat	nay	gup	bun	fit	keg	sop	yes	com	fur
pum	web	ten	wox	dip	jug	sew	jis	toy	gig
lip	tar	jet	pus	rob	feg	coy	win	kid	gum
pew	mix	lep	sar	job	vap	bid	yeb	den	low
sap	ren	fow	new	red	lug	hod	kin	dot	ses
bip	led	war	his	tid	buy	sex	did	rag	hop
yew	mub	got	tax	put	hen	vot	jar	key	him
fad	tub	nor	fix	pem	vow	doy	let	nex	lay

Introspectively it was difficult to take B so soon after A, so that the blank might be cut in two instead of being used as it is. Another difficulty was found in the arrangement of the syllables. There was a tendency to work by vertical columns rather than across the sheet, and section B was confusing for the eye. Either explicit directions should be included, or the syllables printed in even columns.

No one made a perfect record in the time given, but in about all of the "Mark English words" tests, and in some of the "Mark nonsense words" tests the entire blank was gone over within the time, the rest of the time being spent in looking back for omissions. Since, moreover, there were many of both omissions and errors, the measurement of the time of the process is not feasible.

The second test is much harder. The requirement in it of equating time, errors and omissions in the case of almost every subject is troublesome. This difficulty exists to a less degree with the "Mark English words" test.

The amount of improvement due to familiarization with the plan of the test would not apparently be so great as to be very troublesome. When the same blank was used twice, as here, the change of the third over the first trial was for the *marking nonsense* words about 25 per cent. more words correctly marked, and about 30 per cent. fewer words wrongly marked, with a slight increase in omissions.

The remaining three tests were not given each sufficiently often to allow discussion of any practise effect. They were included for purposes of comparison and correlation when taking one or two trials; so that the "short-term group" becomes, to all intents and purposes, nothing more than an "instructed" group in those tests, except for their general experience of test conditions.

B. Relative Value of these Tests

The question of the variability and correlation of these association tests will now be taken up.

The resemblance between an individual's average ability in the *first idea, day opposite, vertical opposite, preceding letter* and *complete the word* tests combined, and his ability in each of these tests separately, was calculated in order to discover the extent to which each single test is significant of the more general ability. This resemblance was calculated both from the percentage of unlike-signed pairs, and also by the Pearson coefficient of correlation.

In the case of these and all correlations to follow, the reader will understand that I am not measuring the correlations between the true abilities which would be found from an infinite number of trials with each test, but only the correlations between the measures got from 1, 2, 3, or 4 trials, as the case may be. The question is not of the significance of certain traits in human nature, but only of certain previously defined tests of those traits.

It will be understood also that other results, mostly from only

10 and in some cases only 6 individuals, are very unreliable. They are however much more reliable than mere opinions.

The performances of the 36 individuals in the "instructed" group were thus correlated with the following results:

TABLE IX

		cos πU	r	(Closest correlation = 1)
Average of these five tests and	First idea	.749	.623	2
	Day opposite	.844	.671	1
	Vertical opposite	.509	.615	3
	Preceding letter	.368	.484	5
	Complete the word	.425	.607	4

Thus by both methods the *easy opposites* seems to be the best test so far as it measures the element common to all these tests on association. By both methods also the *preceding letter* seems the poorest.

Next were used the results (in the first two trials) of the ten individuals in both the long-term group and the short-term group in the following tests: *first idea, vertical opposite, day opposite, preceding letter, complete the word, free association, subject-predicate, difference between, addition, Ebbinghaus combination.*

Again each test was correlated with the average for all, with the following results.

TABLE X

	cos πU	r	(Closest = 1)
First idea	.22	.39	8
"Vertical"	.92	.48	3
"Day"	.79	.71	1
Preceding letter	.81	.42	4
Complete word	.37	.09	9
Free association	—.13	.11	10
Subject-predicate	.37	.47	6–7
Difference between	.64	.23	6–7
Ebbinghaus combination	.66	.67	2
Addition	.79	.39	5

The two methods do not agree so well this time, but again the easy list of opposites correlates high. The *preceding letter* correlates rather low by the Pearson coefficient method, high by the percentage of like-signed pairs. As this latter method takes account only of number of cases of difference whereas r is affected as well by the amounts of difference, it is obvious that a few cases of wide divergence from the average, or in other words a subject making an unusually low record in a certain test, will bring about the discrepancy between the two methods. On examining the original data this is precisely what is found: one subject usually far below the

average made a very good record at the second trial, and one of the very best subjects made the lowest record of anybody at this *preceding letter* test. The Pearson coefficient is greatly affected by these records, and is correspondingly low; by the percentage method their influence is only slightly felt.

Complete-the-word, which was low for the instructed group is also low for these two groups, extremely so by the Pearson coefficient. The other test with very low correlation, the *free association* has inverse relationship by the percentage of like-signed pairs. This means that although the majority of subjects reacted differently in this test from their average reaction in association tests, yet their individual records differ only slightly from each other—the A.D. for this test being very low.

The *Ebbinghaus Combination* test correlates fairly closely by both methods.

The *Free association* test correlates so slightly probably because, as was shown, it is largely a test of the rate of writing for many subjects.

The value of each test of association has been discussed from two standpoints thus far, that of significance measured by highest correlation with the average of all tests in the series and that of least disturbance by practise. A third standard would be that of ascertaining for each test the unreliability of any given number of trials. Where possible this has been measured in the case of: (1) the first four or five records of each member of the short-term practise group, and (2) the first five and sometimes the last five records of each member of the long-term practise group. The average results of (1) and of (2) are presented in the following table in percentage statements. The higher the figure the greater the unreliability of a single trial and vice versa. To this table is added a column to give the number of trials that would be needed to reduce the unreliability to 1 per cent., and a column to give the consequent time it would take to get such reliable information about a person's ability in that test, using as a basis for this calculation the average time taken in an amount-limit test, the time allowed in a time-limit test.

Such determinations are difficult because of the practise effect, and the difference in difficulty of different blanks of the same series. From the gross differences found in an individual's trials, one must, in order to get an approximate measure of how much difference is due to chance variations in the individual, eliminate these two added causes of difference. This can be done only approximately and by more or less arbitrary criteria.

In tests involving differences in quality as well as rate of achieve-

ment there is the further difficulty that one performance may differ from another in quality and in speed or vice versa. The reliability of the test as a whole as a measure of efficiency in the function in question can then be determined only after the combination of the measures for quality and speed into a single measure.

The method taken may be shown best by an example. The records of the three long-term subjects in the " day " opposite test were:

TABLE XI

N.		W.		F.		Av.	
Amount	Quality	Amount	Quality	Amount	Quality	Amount	Quality
13	25	11	21	12	22	12	22.6
15	29	12.5	25	13	26	13.5	26.6
15	29	13	25	13	26	13.6	26.6
17	31	14	26	14	26	15	27.6
15.5	29	14	26	14	27	14.5	27.3

Since the quality was substantially equal throughout for each individual, the reliability may be measured from the differences in the amount score alone. Since, as will be shown in a later section, individuals cluster around a central tendency in respect to changes in the rate of improvement, the general practise effect shown in the average column may be applied to each individual. That general effect smoothed may be taken as 12.5, 13.5, 14, 14.5, 15. So it may be assumed without great inaccuracy that, apart from the chance variations of the subject, the records would have been approximately—

N.	W.	F.
13.5	11.5	11.5
14.5	12.5	12.5
15	13	13
15.5	13.5	13.5
16	14	14

The deviation of the single trials due to the person's varying condition are then for

	N.	W.	F.
	.5	.5	.5
	.5	0	.5
	0	0	0
	1.5	.5	.5
	.5	0	0
A.D.	.6	.2	.3
In per cent. of Av. Amt.	4.0	1.5	2.3

So far as these three subjects go, the probable average divergence of the result obtained from a single trial with the " day " test from the probable true result is then 2.9 per cent. of the former's amount.

To show the reliability of these estimates of reliability themselves, the results from all the short-term and from the long-term subjects are given separately.

TABLE XII

RELATIVE PRECISION OF ASSOCIATION TESTS

Test	No. of Seconds for 1 Trial.	Probable Average Divergence of the Result Obtained from 1 Trial from the Probable True Result, in Per Cents. of the Former				Approximate No. of Trials Necessary to Measure a Person with an Average Divergence of 1 Per Cent.	Approximate Time of Tests so to Measure a Person
		Short Term Data	Long Term Data				
			Early Trials	Late Trials	Combined Estimate		
						49	12¼ min.
Easy opposites [day, good, great, high]	30	6.9	2.9		5	25	12½ "
Hard opposites [vertical, serious]	30		7.4		7.5	56	28 "
Addition [of 5 two place numbers] ..	60	6.0	6.5	5.1	6	36	36 "
Preceding letter	15	10.0	12.4	18.1	13	169	42 "
Complete the word .	15	12.6	8.8	11.2	11	121	30 "

The facts in the case of the *subject-predicate, add and subtract columns, mark nonsense* and *English words* are too intricate to allow even an approximate estimate. So also with *difference between, Ebbinghaus combination, noun and adjective,* and *free association* starting from one given word, though these four are all apparently very much more unreliable than those listed. It appears then that for freedom from ambiguity, significance as a symptom of the condition of the association processes in general, freedom from disturbance by adaptation to the test shown in great early practise effect, and reliability, the best single written test of these is one in giving *easily thought of opposites.* In administering it, skipping should be allowed.

2. TESTS ON MEMORY

A. Descriptive

Along with these tests on association another group of tests on memory was given. Four memory tests are given to the freshmen, the *auditory figures, visual figures, logical memory* and *retrospective memory.* The method of giving them is as follows. For the *auditory figures,* each series of 8 numerals is read aloud at a rate of about 2 per second, after which the subject writes them down " in the order given." In *visual figures,* corresponding sets of 8 numerals are shown one at a time at the same rate. These numerals (Willson's black gummed) are mounted on cards, held in the hand and exposed

by turning them singly to face the subject. In *logical memory*, a passage—to be quoted later—is read to the subjects who then write as much of it as they can. Attempt is made to give the thought completely, and the words where possible. In *retrospective memory*, the subjects are asked to reproduce a line 5 cm. long which they drew as a perception-of-size test at the beginning of the hour, also to " do with it as they did before."

Other visual and auditory tests were used with the practise groups; a few other paragraphs were used though no other change made in the logical memory test; but no other " retrospective " memory test at all similar to this was devised.

The classification into " auditory, visual " and the like may well seem misleading, as it by no means implies that auditory stimuli are remembered in auditory terms, nor, more usually, that visual stimuli will not be translated by the subject into auditory terms. No warning is given to the freshmen with regard to this, and observation shows that the great majority of them do repeat orally the numerals presented visually. Any comparison of tests, then, does not signify a comparison of kinds of memory, but of varied stimuli or material, and varied ways of presenting material. On the report sheet sent to the freshmen care is taken to say " numerals heard," and " numerals seen "; but here, for brevity's sake, the more usual designation of auditory, visual, etc., will be adhered to, with the understanding that the words refer to stimuli, not to memory terms. For convenience sake also, the tests with auditory stimuli are discussed first, those with visual stimuli later, though the related words might possibly be classified as a logical memory test.

Auditory Figures.—Experience with this familiar test as given to the freshmen shows that most of them group the 8 numerals in two groups of four. Enquiry reveals that many depend upon a memory after-image for the last four, and memorize the first group only. The average number correctly remembered is 7.6 for the men, 6.7 for the women. This test is thus too easy, many of the individuals obtaining perfect scores.

The chief difficulty in comparing people's work on memory lies in the variable methods of scoring, especially with regard to transpositions. If the order is 76431528, and a subject writes 7463 . . ., some experimenters call it two errors because both the 4 and the 6 are in the wrong places; other experimenters call it one error because by making one change—by " lifting " the 6 over the 4, it is corrected. The latter method seems preferable. Supposing a subject were to write 87643152, eight errors would be scored by the first method since each numeral is misplaced; by the latter method only

one error is scored, since one change would set all right. Also, a misplacement error would be rated more nearly as an omission. A subject writing 76–31528 would be scored one error for omitting the 4, but two if he places it before the 6, by the first method; in either case he is scored just one error by the latter method, putting misplacements and omissions on an equal basis.

In the work to be reported on therefore, the second method was used, only that a positive score was used instead of counting the errors. Thus each numeral given correctly was scored 1/2, and if it was in the right place—interpreting this as relative place not absolute place—it was scored 1/2 more. This modification has the advantage of being rapid to use in determining the score, especially of the different kinds of material used in the tests. It is also much easier and can be used more rapidly than the Spearman "foot-rule" method, or the modification recommended by Whipple ("Manual," p. 266). If it is too cumbersome when it comes to calculating correlations, the figures can be very quickly read off as numbers of errors.

According to this method the average freshmen scores would be, as before, 7.6 for the men, 6.7 for the women.

To the "instructed" group of eighteen subjects, two sets of ten numerals were given, with an average score of 7.2 figures remembered for the men, A.D. .75; and 6.1 for the women, A.D. .85. This agrees with the superiority shown by the men over the women in the freshmen results, though showing lower scores.

The short-term group made six trials with ten numerals at a time, with an average score of 8.8 numerals remembered, A.D. .7. The series of 10 was long enough to measure all in this group. No practise effect was observable.

The long-term group made twenty trials with ten numerals at a time. One subject made only four errors in the whole series, her memory span for this being evidently greater than ten; in consequence her records were not used in estimating practise. For the other two subjects the average score was 9.55, the first day's average deviating by —.55, the last by —.5, or taking the first two and the last two trials, the deviation at first was —.45, and at last +.2.

For these two subjects also the list of 10 was not long enough to measure the practise effect accurately, there being numerous perfect scores. Their records were, in order (in errors):

N.	1	2	1	0	1		2	1	0	0	1		0	0	1	1	0		1	1	0	0	2
F.	3	1	1	2	3		0	1	0	0	2		1	0	0	1	0		2	2	0	0	0

Two other auditory tests were used, (1) series of fifteen *related words*, and (2) *mixed* series of unrelated units, including besides

LISTS OF RELATED WORDS

I	II	III	IV
College	See	Book	Holiday
course	sensation	author	excursion
grade	perception	style	boat
graduate	interpret	classic	train
senior	illusion	literature	ticket
dues	cortex	essay	early
money	hemisphere	poem	seat
purse	ganglion	rhyme	hot
lost	dendrite	meter	window
advertise	branch	scan	draught
reward	conduct	quantity	cold
deceive	intercept	Latin	bronchitis
angry	numb	translate	doctor
threaten	injury	language	medicine
blows	paralyze	accent	cure

V	VI	VII	VIII
Noise	Sunset	Time	Black
cat	dusk	test	negro
baby	lamp	write	Africa
child	table	quickly	Congo
kindergarten	play	maze	Leopold
child-study	deal	difference	rubber
psychology	lead	sorting	cruel
Thorndike	queen	color	atrocity
chickens	trump	forms	remonstrate
monkeys	short	remember	America
bananas	partner	auditory	Rockefeller
fruit	trick	score	millions
skin	point	improve	oil
slice	rubber	average	monopoly
supper	stop	twenty	trusts

IX	X	XI	XII
Picture	Child	Sunday	Finance
photograph	teacher	rest	stocks
pose	rude	church	rise
recognize	naughty	sing	fortune
because	punish	choir	invest
older	sorry	organist	dividends
friend	forgive	training	railroad
together	better	abroad	anthracite
travel	promise	Germany	Phoebe
foreign	broken	Berlin	advertisement
steamer	hardened	university	magazine
seasick	discourage	philosophy	story
improve	report	research	read
turbine	trouble	valuable	hammock
Cunard	consult	publish	trees

XIII	XIV	XV	XVI
Dog	Sky	Paper	Teach
kind	cloud	envelope	physics
terrier	raining	write	experiment
rats	wet	letter	light
hunt	spoilt	parents	refraction
catch	new	away	angle
trap	expensive	seaside	measure
poison	money	sands	survey
antidote	draw	bathe	instrument
doctor	bank	swim	careful
ambulance	cashier	deep	understand
policeman	dishonest	cramp	accurate
Irish	abscond	drowning	rely
Murphy	scandal	revive	promote
milk	newspaper	thankful	successful

words, numerals, letters of the alphabet and sounds such as clapping the hands, tapping, ringing a bell, shuffling the feet, whistling, etc., the necessary movements being out of sight of the subjects.

The short-term group made five trials using series I., II., III., IV., and VI. Besides scoring in the manner described, note was kept of whether the errors were those of omission or misplacement, or whether extra words were put in. At first sight it would seem best to handle this score by keeping it in terms of errors made; but as the score is given for the right words in the right order, additional words practically counted as errors. From the point of view of interest in individual differences, however, it was felt worth while to keep track of the number and occasion of additional words; also to note whether any one list seemed more tempting to the imagination than others. In a total of 30 records, eight of them had extra words, one subject supplying them three times. She remembered the greatest number of words correctly. The subject with the lowest score put in extra words twice. Every subject misplaced some words, the one with the best score doing so most often. The average score was 8.9 words A.D. 2.5. There was no practise effect discernible.

The long-term group in a total of twenty trials made an average score of 12.6 words, A.D. 1.5. The first two trials deviated by —.75, the last two by —.15, but there seemed no certainty of practise effect. The lists of 15 words were just long enough to measure the most capable of these subjects; toward the end of practise a list of 16 would be better for regular use. No particular list seemed specially liable to error. The subject with the highest and least variable record wrote the fewest extra words, and made six perfect records.

Both of the other subjects showed considerable variation, one having five perfect records, but misplacements 50 per cent. of the time, the other having no perfect records, and only three free from extra words or misplacements. The one with the greatest number of misplacements also wrote the greatest number of extra words. The subject who had so good a record with the auditory numerals was not the best in this test.

Auditory Mixed.—The object in giving this test was to present material absolutely disconnected, yet with each of the units in the list having its own meaning. Even with nonsense syllables some fanciful connections are usually made, so that it was not supposed that artificial associations could be entirely avoided; nevertheless by introspection there seemed to be very few of them in this case. There is some difficulty in presenting nonsense syllables orally, but with this incongruous yet senseful material there is less danger of errors in hearing on the part of the subjects. The tendency to groupings of four was broken up somewhat by the introduction of the various sounds or noises (shown in the list by italics). By introspection this test proved difficult and irritating to those accustomed to the other material.

The lists used were as follows:

(1)	(2)	(3)
Carriage	Distance	Oo
F	as	but
adversary	*whistle*	16
preach	flag	resting
stamp with foot	require	*clucking noise*
lamp	38	organ
never	other	3
ring a bell	harper	spring
K	*clap hands*	W
green	H	matches

(4)	(5)	(6)
And	Monstrous	99
20	(*jingle keys*)	monotone
ring a bell	X	*scrape with foot*
wall paper	Symphony	alphabet
stampede	*tap with pencil*	tomahawk
tap with finger	she	*jingle keys*
M	*whistle*	asleep
symmetry	bugle	purple
stamp with foot	typewriter	*tap, or clap*
56	ice-cream	because

The short-term group made only 2 trials, with an average score

of 8.15, A.D. .45. The long-term group made 20 trials, with an average score of 9.2 of the ten remembered, A.D. .35. The detailed results were, in order (in terms of errors)—

N.	3	5	2	1	1	2	2	1	2	3	2	1	1	2	1	2	1	1	0	2
W.	0	2	2	2	2	1	1	1	2	1	3	1	1	1	0	3	1	1	1	2
F.	2	2	2	2	3	1	1	0	3	1	2	1	2	0	1	3	2	1	1	2

There was no practise effect discoverable. The subject who was so very competent with the auditory figures was also the best in this test. The misplacements were unfortunately not noted, so that no comparison can be made in this respect with the *related words.*

Visual Figures

Three sets of eight numerals are shown serially to the freshmen. No apparatus is used, and some little practise is required on the part of the experimenter to expose the cards regularly and at a convenient angle. As said before, no warning is given about not repeating to one's self orally what is shown. The men remember 6.9 correctly on the average, the women 5.7.

Two of these sets were used with the "instructed" group. The men made an average score of 5.85, the women of 5.15, again agreeing with the freshmen results in the superiority of the men's record over the women's, though showing lower scoring for both men and women than in the case of the freshmen. The percentages would be 73 and 64.

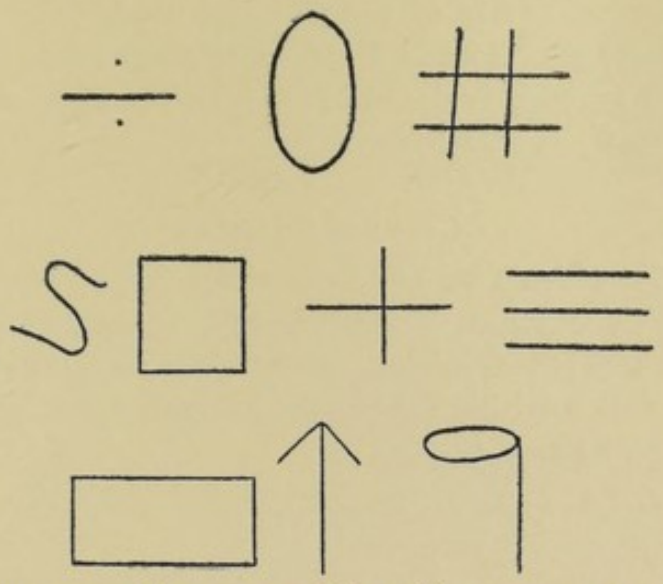
The short-term group made 5 trials with sets of 8 numerals; their average score was 7.5, A.D. 0.5. Series of 8 are thus too short for an adequate measure of visual as well as auditory memory.

The long-term group made 20 trials with sets of 10 numerals. For the first four trials cards were used as for the freshmen. After this, as a screen with a slit was in use for other visual material it was used for the numerals also. This screen was a very simple affair of pasteboard with a 2-inch square opening in the middle. The visual stimuli were written or drawn with charcoal on a long strip of cardboard which was pushed along behind the screen, allowing one second for the exposure of each unit in the series. By reversing the strip, one series could be used as two different tests on different days. Sixteen trials were made with this, making twenty in all. Even series of 10 numerals are too short for adequate measurement of these subjects, perfect records being made frequently after the first three trials.

Their average score was 9.4, A.D. .5, the range from 8 to 10. The first day's average deviated by — .1, the last by + .8.

Other visual tests were: *grouped forms, serial forms, grouped objects, serial objects, forms recognized.*

Grouped Forms.—Five different sets were used, one of which was as follows:



These forms were drawn roughly with crayon on a small blackboard which could be turned and exposed to view for 10 seconds, then turned away again. The short-term group made only two trials with sets 2 and 4. Their average score was 5.4 forms, A.D. .9. The long-term group made 10 trials, average score 8.15 forms, A.D. .1.0. The first day's trial deviated by —1.35, the last by +.35. It had been intended to make 20 trials with this as with the others; but very soon the question arose whether it was not much easier to look at a group of 10 for 10 seconds than to see 10 units one at a time for one second each, in the same way that the numerals are shown, with no chance of looking twice at any one of them. It was decided to compare the grouped with the serial method, both for forms and objects, though cutting down the number of trials to 10 each, for this group of subjects.

Serial Forms.—The cardboard screen and strip, as described before, were used in this test. The sets of forms were similar to those used in the *grouped forms* test. Two of them are here reproduced.

The short-term group made 4 trials, of which the average score was 6.5, A.D. .75. The averages of the successive trials were 5.66, 5.83, 7.07, 7.43 showing a greater gain for them in this test than in the other immediate memory tests. Probably this is due to the initial comparative unfamiliarity of the material used.

The long-term group made 11 trials, average score 7.95, A.D. .95. The first day's average deviated by —1.95, the last by +1.7, showing a very great practise effect.

Grouped Objects

Ten familiar objects chosen from about 25 in daily use, such as a watch, box of matches, bunch of keys, spool, envelope, pack of cards, books, scissors, fish-hook, soap, were arranged in the same groupings as that used for the grouped forms, a row of three, a row of four, a row of three, thus,—

X X X
X X X X
X X X

on a small table behind a screen. At the signal the screen was raised for 10 seconds. The subjects then wrote down the names of the things seen, grouping the names as the objects had been grouped.

Only the long-term group practised with this test, their average score in ten trials being 8.85. The first day's trial deviated by —1.25, the last by —.1. On the fifth and eighth trials, perfect scores were made, however, by all three subjects.

Serial Objects

In this test, the same sort of objects were picked up one at a time and shown for one second each above the screen.

The long-term group in ten trials made an average score of 9.3, the first day's average deviating by —.3, the last by +.1.

So far then as serial grouped method is concerned there seems, by examination of the accompanying table,

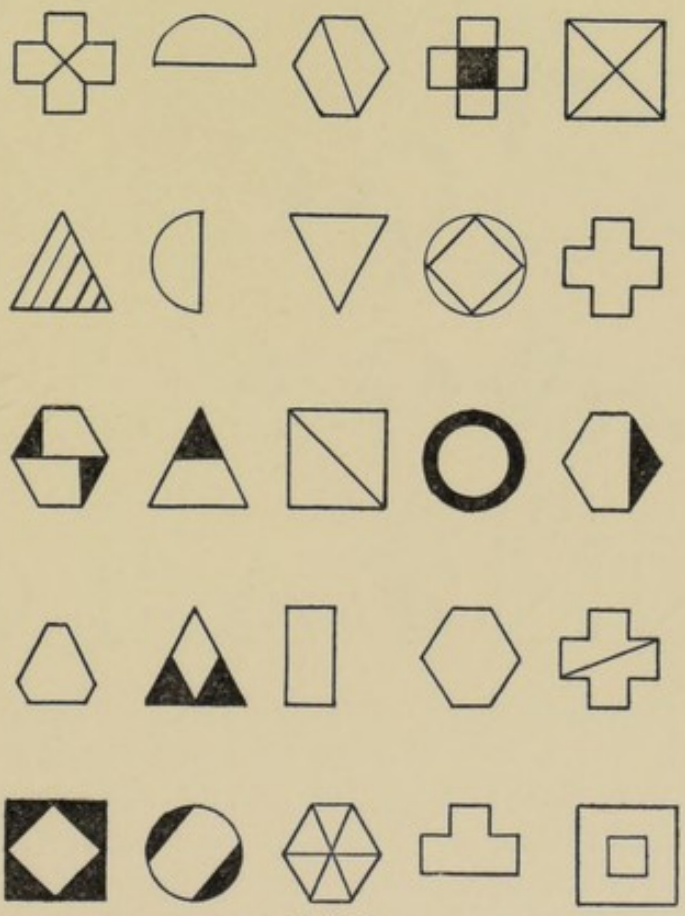
	Serial	Grouped	
Short-term	6.85 (4 trials)	5.4 (2 trials)	Forms
	6.0 (first 2 trials)		
Long-term	7.95	8.15	
	8.3	8.75	Objects

to be a slight balance in favor of the serial method, probably because this is the familiar method used for numerals, and in auditory stimuli. Introspectively, the long-term group found the *grouped*

forms easier than the *serial forms*. The reason is, perhaps, that with the latter method the second of exposure is not always sufficient for the recognition of some of the forms, whereas when grouped, the total 10 seconds can be distributed in the most economical manner, the eyes pausing longer, or returning to those forms not so readily apperceived. In the case of objects shown, this factor of apperception scarcely entered in, as each object was readily recognized, and mentally named in its one-second exposure. A slightly higher score was made on the average for objects shown serially than shown grouped.

Forms Recognized

The blanks used in this test are reproduced on this and the three following pages.



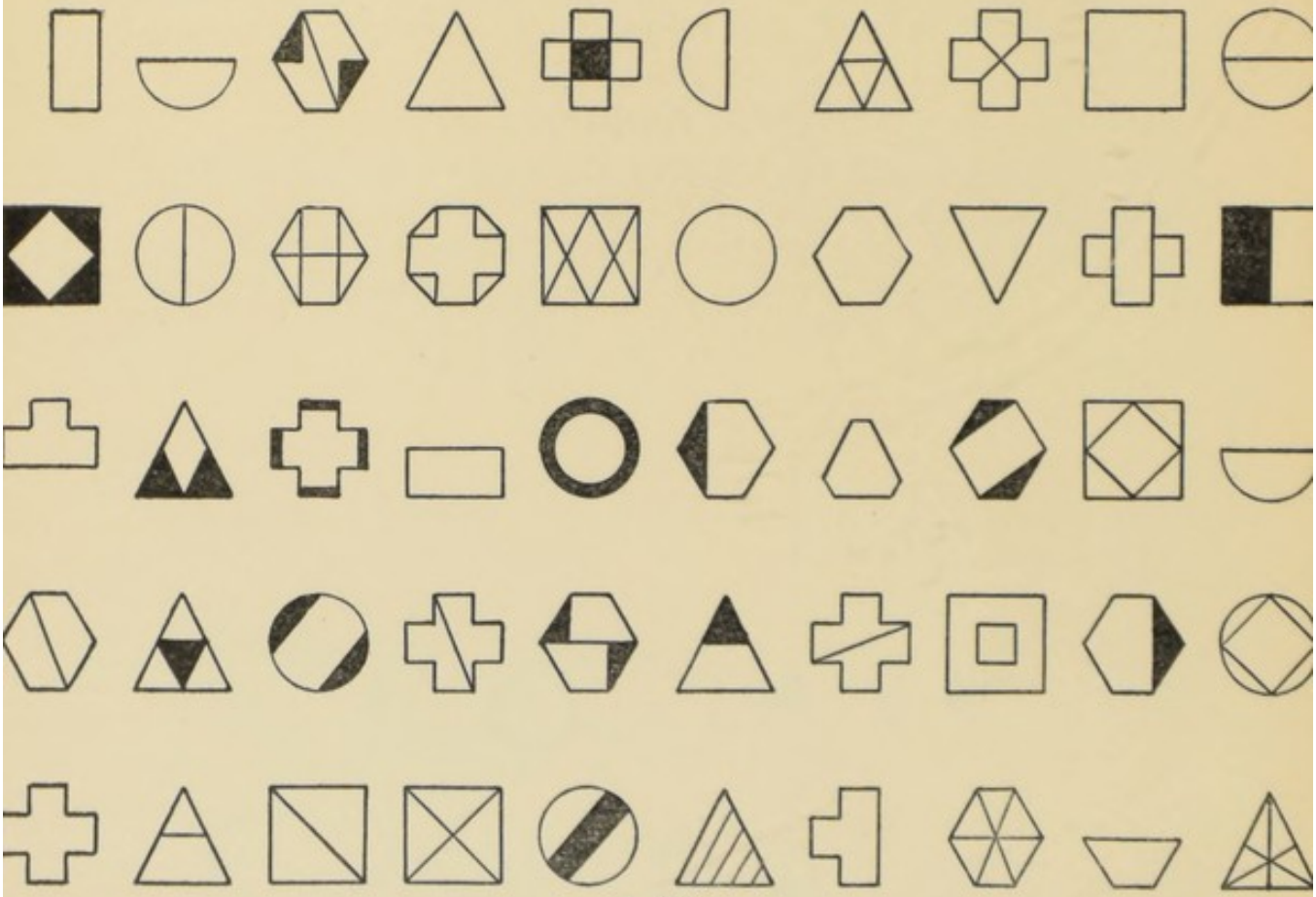
1 (*a*)

The subject is given the small sheet with instructions to study it in any way preferred till at the end of 60 seconds he is given another sheet on which he is to mark as quickly as possible all the forms he remembers having seen on the first sheet. It will be noticed that on (1) 24 can be marked, on (2) only 18.

The time taken to mark the second sheet is noted, also the total number marked, and the number correctly marked.

Set (1) was given to the Barnard freshmen of the class of 1912. The average time taken by 49 of them was 66 seconds, A.D. 16.2, with 15.6 correctly marked, A.D. 2.3, and 5 wrongly marked.

Six members of the short-term group and the most rapid worker in the long-term group made one trial with this set. Their average time was 81 seconds, or, not counting N., 88 seconds, A.D. 22.5 with 15 correctly marked and 2 wrongly marked. These subjects made trial



1 (*b*)

also with (2), where their average time was 115 seconds, A.D. 33, with 9.5 correctly marked, A.D. 1.3, and 3.5 wrongly marked. It is much more difficult than set (1).

The attempt thus to measure memory by a combination of the amount recalled, the quickness with which it is recalled, and the errors made, should be carried on with better material. The results obtained here are of value only for measurements of the significance of this particular test by its correlations.

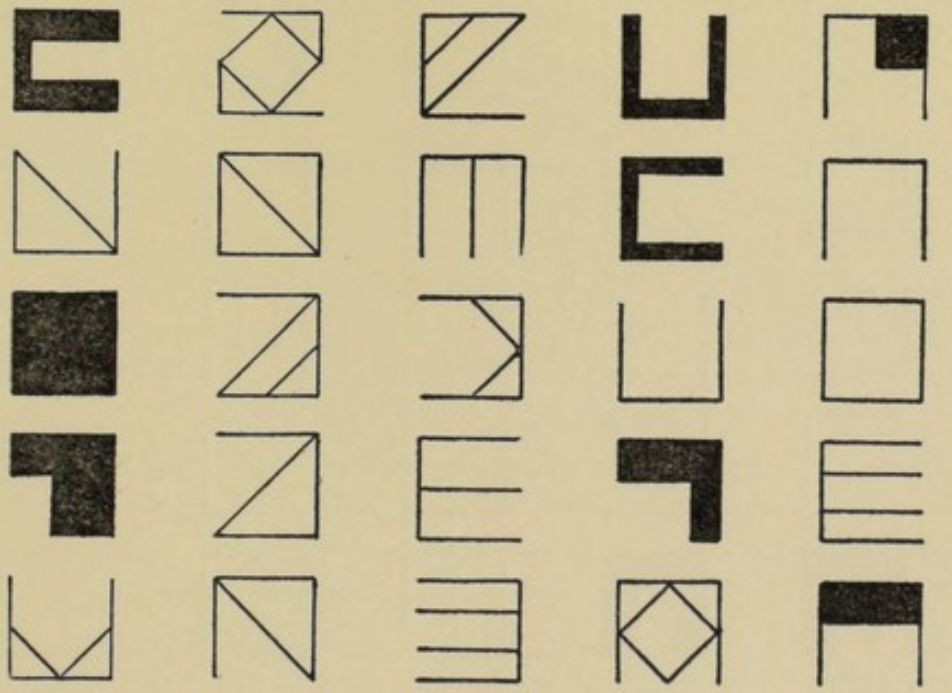
Two other memory tests were given, the *logical memory* and the *retrospective memory*.

Logical Memory. A paragraph is read aloud to the subjects who then write out as much as they remember of it, stress being laid upon

the matter rather than on the words remembered. To the freshmen the following paragraph (I.) is read:

I

Tests such as we are now making are of value both for the advancement of science and for the information of the student who is being tested. It is of importance for science to learn how people differ and on what factors these differences depend. If we can disentangle the complex influences of heredity and environment we may be able to apply our knowledge to guide human development. Then it is well for each of us to know in what way he differs from others. We may thus in some cases correct defects and develop aptitudes which we might otherwise neglect.



2 (*a*)

The men remember 44.5 per cent. of the ideas contained in it on the average, the women 51.2 per cent.

The short-term group made four trials, once with this paragraph, and once with each of three others II., III., and IV.

Other Passages Used

II

Could the young but realize how soon they will become mere walking bundles of habits, they would give more heed to their conduct while in the plastic state. We are spinning our own fates, good or evil, and never to be undone. Every smallest stroke of virtue or of vice leaves its never so little scar.

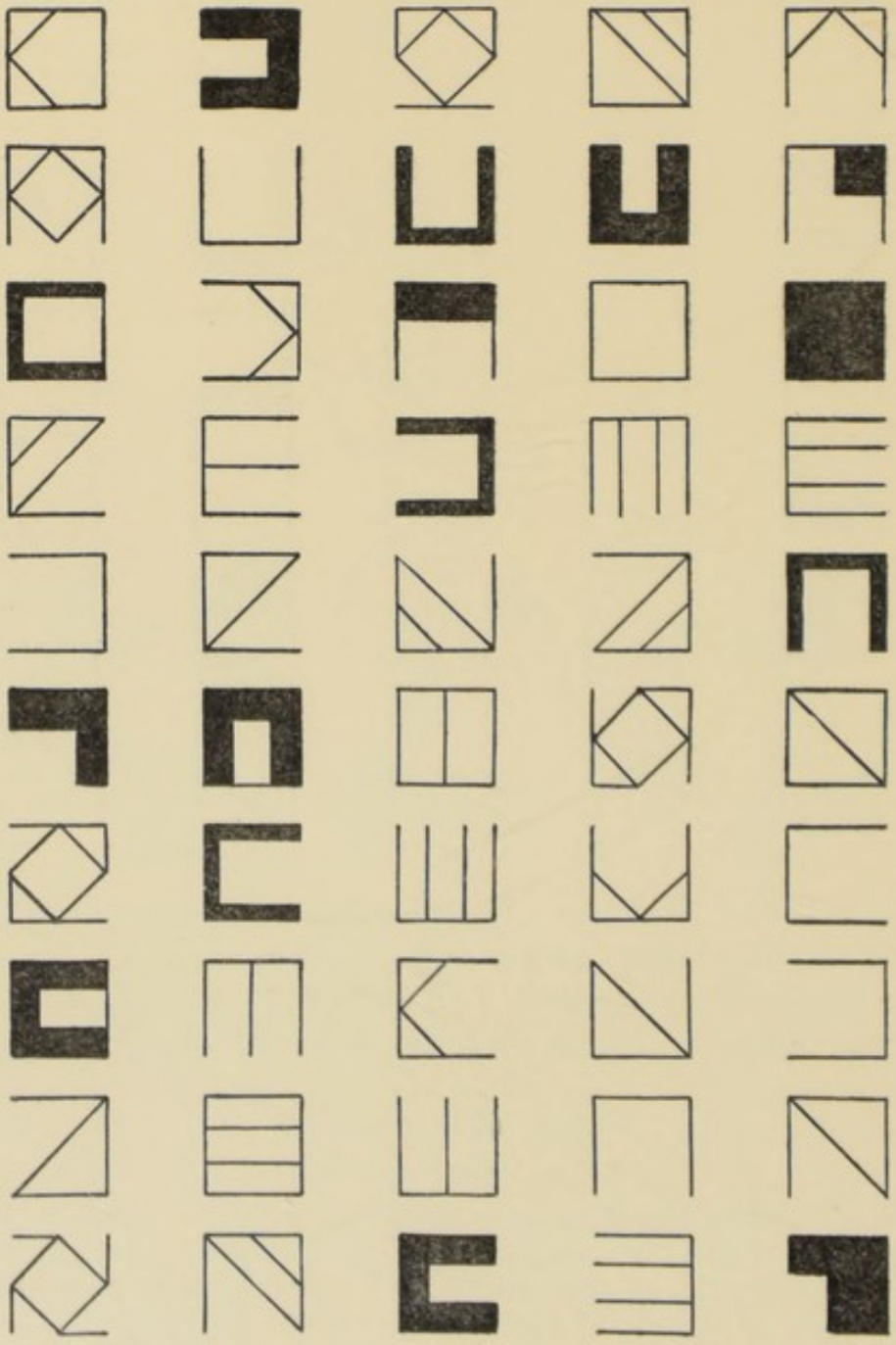
III

Measures of the variability of the individual measures are of two sorts: measures of the averaging type and measures of the percentile type.

The mean square deviation equals the square root of the average of the squares of the deviations of the individual measures from their average, median, or mode.

IV

The abstract scheme of successive predications, extended indefinitely, with all the possibilities of substitution which it involves, is thus an immutable system of truth which flows from the very structure and form of our thinking. If any real terms ever do fit into such a scheme they will obey its laws.



2 (*b*)

The average percentage remembered was 39.1; for paragraph I. alone it was 49 per cent., slightly lower than was the case with the

Barnard freshmen. These tests were given primarily as a means of estimating the significance of so called "logical memory," and no data on the effect of practise were secured.

TABLE XIII

INDIVIDUAL CREDITS FOR MEMORY PASSAGES

Graded on a scale of ten

	I	II	III	IV
Bu.	7.0	7.0	5.5	6.0
Gr.	3.5	5.5	2.0	3.0
St.	3.0			
J.		1.0	3.5	1.0
L.	5.0	6.5	1.5	3.0
M.	6.0	5.5	1.5	3.5
Ba.		6.0	1.5	3.5

Retrospective Memory.—Instead of the test given the freshmen, which consists of reproducing a line the same length as one seen and reproduced an hour previously, the long-term group made ten trials in eight of which they were asked to reproduce the list of 15 related words given as an auditory test on the previous day. On that occasion the list had of course been read, written more or less correctly and then re-read for the subjects' satisfaction in their performance, so that there had been three repetitions of the list, two of them correctly, followed by an interval of about twenty-four hours. At the third and seventh trials other material was used. Once they were asked to reproduce a paragraph used the day before in a "complete the paragraph test," and once to give the ten kinds of objects used in a "naming 100 objects" test—yet to be described. It would be interesting to prolong and vary this test indefinitely, as individuals differ so much in their ability to recall different kinds of things after different intervals, and so many human interests depend upon the accuracy and length of retention; but as the object here was merely to discover any tendency to practise effect in such mature subjects, and as time and opportunity were lacking for more prolonged series, only these ten trials were made. The score was 9.9 on the average, with no practise effect discernible.

B. Relative Value of these Tests on Memory

On the whole, there is no evidence that in any of these tests of immediate memory, a first trial measures a markedly different process from later trials after the subject is adapted to the form of the test. No great difference can exist, or it would show itself in the work of the short-term group. With the possible exception of *serial forms,*

there is no test in which the second trial shows any greater proportionate improvement over the first than the fourth or fifth shows over the third or fourth. Indeed, in almost every case it is among the records of the long-term group that evidence of the existence of any practise effect must be sought.

The tests rank in respect to susceptibility to practise as follows:

Very slight, not discernible in these cases	*Auditory mixed.*
	Serial objects.
	Retrospective.
Slight (less than 10 per cent. in 20 trials)	*Auditory figures.*
	Auditory words.
	Visual figures.
Considerable	*Grouped objects.*
	Grouped forms.
Most	*Serial forms.*

Certain correlations of these various tests on memory have been computed.

First of all, taking the short-term and long-term groups together, the average of the first three records of each subject in the following tests were compared, each test with the average for all six tests: *auditory figures, related words, auditory mixed, visual figures, grouped forms, serial forms.* In calculating this set of correlations the deviations of each subject in the short-term group from the average of her own group were taken, not from the average for the ten subjects treated as one group.

Next, the records of the "instructed group" with *auditory figures* and *visual figures*—18 cases, two trials for each—were correlated; also the same tests for the short- and long-term group, as above. Similarly nine subjects' records with *auditory figures* and *related words,* and five subjects' records with *related words* and *logical memory.*

Third, all auditory tests, viz., *auditory figures, related words,* and *mixed series,* were averaged, and each test correlated with the average of all, using again the average of the first three records of both short- and long-term groups.

Fourth, using 10 subjects as above, the correlation of *grouped* and *serial forms* was computed.

Last, *visual figures* was compared with *forms recognized* using the records of the 49 freshmen, and also those of the short-term group. The latter test was also compared with *grouped forms,* a supposedly similar test.

All these results are presented in the following table, where in addition to the Pearson coefficient, the rougher correlation by the

method of unlike-signed pairs is given wherever justified by the number of cases available.

It will be understood that these correlations are to measure the significance of three (or two, as noted) trials of a given test, not the true relation between an individual's total ability in one trait and his ability in another. The reader is again reminded that the results commonly from only ten subjects are only very coarse approximations, but are nevertheless by so much better than nothing.

TABLE XIV

		cos πU	*r*	No. of Cases
1. Average of these six tests and	Auditory figures	.31	.51	10
	Related words	.93	.64	10
	Mixed series	.31	.05	10
	Visual figures	?	?	10
	Grouped forms	.95	.91	10
	Serial forms	.31	.45	10
2. Auditory figures and Visual figures		0	.21	18
Auditory figures and Visual figures		0	.17	10
Auditory figures and Related words			.12	9
Logical memory and Related words			.55	5
3. Average of these three tests and	Auditory figures	.48	.69	10
	Related words	.93	.58	9
	Mixed series	.93	.64	10
4. Grouped forms and Serial forms		.81	.76	10
5. Forms recognized, and Visual figures		.03	.37	49
Forms recognized, and Visual figures			—.13	6
Forms recognized, and Grouped forms			.26	6

In the first set of correlations, with varied material and including auditory and visual tests it would be surprising to find high correlations. *Grouped forms* stands out conspicuously therefore as a typical test—in so far as it measures whatever element may be common to all these six tests. *Related words* comes next by both methods of correlation, while *visual figures* is actually an inverse relationship.

In the second set it is seen that *auditory* and *visual figures* have a very low correlation, none by the percentage of unlike-signed pairs. Clark Wissler, who differentiates between numerals correctly given and those correctly placed, found correlations of .29 and .39 respectively.

The correlation of *auditory figures* and *related words* is, however, still lower, though too much can not be argued from the records of only 9 subjects. The very few records for *related words* and *logical memory* similarly cautious against too great emphasis on the higher correlation found there, though this is certainly more what might be

expected. The unreliability of these two Pearson coefficients is (P.E. r true—r obtained) .021 and .184 respectively.

In the third set, it is interesting to see that all the correlations of the auditory group are fairly high, and that *auditory figures* come out better than *related words* reckoning the Pearson coefficient only, though in the first set this was not the case. Even the *mixed series* correlates well with the average of the group, and the coefficient is higher than that of *logical memory* and *related words* (in the second set), rather unexpectedly.

Summing up this work on memory from the point of view of intercorrelations, *auditory figures* and *related words* seem tests fairly typical of any presented to the ear. *Grouped forms* seems distinctly typical as, taken all through, its correlations are high.

As to the question of the relative precision of the different tests of memory, making a reasonable allowance for practise effect, where such exists, the unreliability of single trials with the tests described are as shown in Table XV. The unreliability of a test with visual figures can not be properly estimated. The series of eight were, as has been stated, too short, and the series of ten was for the long-term group too short toward the end of practise. From the early trials of these three subjects the average divergence of the result from a single trial from the true result may be estimated as from 5 to 7 per cent. according to how the probable course of practise is estimated.

TABLE XV

RELATIVE PRECISION OF MEMORY TESTS

Test	Most Probable Average Divergence of the Result Obtained from 1 Trial from the Probable True Result, in per cents. of the Former				Approximate No. of Trials Necessary to Measure a Person with an Average Divergence of 1 per cent.
	Short Term Data	Long Term Data			
		Early Trials	Late Trials	Combined Records	
Auditory words	18.0	13.1	12.8	14.6	213
Auditory mixed		4.3	3.5	3.9	15
Visual grouped forms		14.6	12.1	13.3	177
Visual serial forms	9.7	9.9	13.6	11.1	123
Visual grouped objects		5.4	10.8	8.1	65
Visual serial objects		3.1	4.0	3.5	12
Visual figures		6.7			(45)

So far as the data go, *Auditory mixed series, Visual serial objects* and probably *Visual figures* (with a long enough series) have decided advantages from the point of view of precision over the other tests. *Auditory figures* was, as given, too easy a test to measure the subjects and therefore could not be included in this list.

If a choice of tests were to be made therefore, a good test, correlating with other auditory tests and not much subject to practise with mature subjects, and requiring few trials for a fair degree of precision is *Auditory figures. Related words* is good except for the lack of precision accentuated by the fact that any selected list of words with its varied appeal to different types of subjects would be less simple than numerals with their greater similarity of associations.

In spite of its susceptibility to practise and the greater number of trials required to give a fair degree of precision *Grouped forms* is suggested as the best visual test for three reasons: (1) it is significant of memory in general; (2) subjects have slight tendency to repeat the name of the form, so that it appeals merely to the eye better than do numerals or objects; (3) it is equally easy if not easier to give than *Visual figures,* requiring less dexterity in manipulation. Standard groups could easily be drawn or printed on cardboard, say two feet six inches square, and thus used for small groups as well as for individual work.

These tests complement one the other and would together make an easily given, easily scored and fairly significant and precise test.

3. Tests on Perception

A. Descriptive

The A Test.—The following blank, here reduced in size, is used with the freshmen.

OYKFIUDBHTAGDAACDIXAMRPAGQZTAACVAOWLYX
WABBTHJJANEEFAAMEAACBSVSKALLPHANRNPKAZF
YRQAQEAXJUDFOIMWZSAUCGVAOABMAYDYAAZJDAL
JACINEVBGAOFHARPVEJCTQZAPJLEIQWNAHRBUIAS
SNZMWAAAWHACAXHXQAXTDPUTYGSKGRKVLGKIM
FUOFAAKYFGTMBLYZIJAAVAUAACXDTVDACJSIUFMO
TXWAMQEAKHAOPXZWCAIRBRZNSOQAQLMDGUSGB
AKNAAPLPAAAHYOAEKLNVFARJAEHNPWIBAYAQRK
UPDSHAAQGGHTAMZAQGMTPNURQNXIJEOWYCREJD
UOLJCCAKSZAUAFERFAWAFZAWXBAAAVHAMBATAD
KVSTVNAPLILAOXYSJUOVYIVPAAPSDNLKRQAAOJLE
GAAQYEMPAZNTIBXGAIMRUSAWZAZWXAMXBDXAJZ
ECNABAHGDVSVFTCLAYKUKCWAFRWHTQYAFAAAOH

There are 100 A's on it, and the directions are to mark as quickly as possible all the A's. Since several A's occur together more than once it might be better to tell them to mark each A.

The men take 100 seconds on the average, the women 87.3 seconds, agreeing with the general conclusion that women are quicker with this sort of test—noticing details—than are men. The general ex-

perience is that all the A's are not marked by either the men or the women, so that when using these figures comparatively, *i. e.*, when 60 A's are scored in 60 seconds for the men, and 68.7 for the women, it must be understood that they are only approximately correct, are in fact a little too high.

In testing this test, the following blanks were used. No. 2 has also 100 A's; No. 3, 50 of each of the letters A, B, K, S.

SET No. 2

GAAQYEMPAZNTIBXGAIMRUSAWZAZWXAMXBDXAJZ
ECNABAHGDVSVFTCLAYKUKCWAFRWHTQYAFAAAOH
UOLJCCAKSZAUAFERFAWAFZAWXBAAAVHAMBATAD
KVSTVNAPLILAOXYSJUOVYIVPAAPSDNLKRQAAOJLE
AKNAAPLPAAAHYOAEKLNVFARJAEHNPWIBAYAQRK
UPDSHAAQGGHTAMZAQGMTPNURQNXIJEOWYCREJD
TXWAMQEAKHAOPXZWCAIRBRZNSOQAQLMDGUSGB
FUOFAAKYFGTMBLYZIJAAVAUAACXDTVDACJSIUFMO
SNZMWAAAWHACAXHXQAXTDPUTYGSKGRKVLGKIM
JACINEVBGAOFHARPVEJCTQZAPJLEIQWNAHRBUIAS
YRQAQEAXJUDFOIMWZSAUCGVAOABMAYDYAAZJDAL
OYKFIUDBHTAGDAACDIXAMRPAGQZTAACVAOWLYX
WABBTHJJANEEFAAMEAACBSVSKALLPHANRNPKAZF

No. 3

GWBTBVKIKSCSAUEBCIWVABZSMDUBKLWHKHYCGYGK
NANNCBVBSAKOIUPEKCXVGSTVRIWYBYGKHAZLPBYO
XAPYEXXHUFSBVDYDIAZLRSATZAZVFCOFSAIPTDOK
BBISKAKHXDYIUZRHVRZYSCIGECPOFKBICBMGFSDC
YHSRMVBLYICKZBMXFVBBIKUCBZLOGLVKGFMOATUN
SHOFHXIMKUXLDZKMRYRLVUWWKYEUVECSOUWBADEX
ALUAKRMSFTGXWLVGAOWBTPODXBNSFSFSWSDRSMPO
KBRIGAXZBZACKFBBEVWCGSWBMFEMXXOKRDIWGGBL
BTPNSKBACVTCSSRKUBURUDMZEWIZFESTMZEBWAFI
BKSGYHSLSFABTLTIUDXGAKROZYKOBHEAALPMLLKC
GVCWKKPTUYUGSTSSDWNKSIEICSNBTVADKANTKKPB
UXGTSOSUZPNBKRBAFDYFOVYBMPSOMBUOPMEGKKTA
COWVFXATSVAPAKYVAHNFXSBDAZYDCFDPPKNPHAMM
XUNKDXSRAAMDVOPECXRKTLHAXVKSHYWEWMMNNHBR
SLSOZFBZGRRIIHKRLEKHEZRGSCYKUIPSLECKYNDA
UGKLLEMAXFYERKWZYSNTTUAVSNAAMNWSAODFWAEH
WBNSPAKBBAOAHPHBHNRDELDLMPWZTAIORTSKLBAZ
HNBKXPSNXAZHNIPHFGTE

The disturbing effect of adaptation and practise with this test is very slight. The short-term group using blank 2 required .783 second per A marked in their first trial of 45 seconds and .869 second per A marked in a second trial of 60 seconds. The long-term group using blank 1 required .643 second per A marked in the first, and .636 second per A in a second trial, each of 60 seconds.

An "instructed" group of eleven subjects who marked A, B and K in order in three successive trials with blank 3, took only nine tenths as long per K as per A; but the same proportionate time was taken when K was given, as the first to be marked, to one group of 18 and A to another group. The difference was therefore probably largely due to the greater ease of marking K.

To determine the relative difficulty of finding A, K, B and S on No. 3 blank, four similar groups of 19 subjects were tested, each group marking a different letter. A time limit of 105 seconds ($1\frac{3}{4}$ minutes) was allowed to mark the 50 letters. The results were as follows:

TABLE XVI

Blank	Letter	Time	Av. Marked	A.D.	No. of Cases
No. 3	A	105	41.3	5.1	19
No. 3	B	105	40.0	5.2	19
No. 3	K	105	37.5	3.9	19
No. 3	S	105	44.6	5.1	19

The time was possibly too long to measure all adequately in the case of the letter S.

The short-term group gave the following results which, in view of the probability that practise effect is very slight, may be used to estimate the relative difficulty.

TABLE XVII

Letter	Method	Time in Sec.	Av. Marked	A.D.	Sec. per Letter
S	Time limit	40	26.0	8.0	1.54
S	Time limit	30	18.0	5.0	1.67
	(Three other trials intervening)				
B	Amount limit	Av. 117	47	17 sec.	2.49
K	Amount limit	Av. 112	43.5	12.3 sec.	2.61
A	Time limit (not reached)	90	50	?	1.80
A	Time limit	60	31	5.3	1.94

K is a little harder than B as before, and S is easier than A by about the same proportion as before. A and S can not properly be compared with B and K since the announcement of a time-limit seems to have a stimulating effect.

An "instructed" group of eleven subjects in a 60 second test with the order ABKS gave averages marked of 30.1, 32.7, 27.0, and 37.1 respectively, or 2.0, 1.83, 2.22, and 1.62 seconds per letter marked. These figures where the practise effect for A in comparison with S is reversed confirm the others.

Concerning the influence of the time-limit versus amount-limit method the following records show that the former does seem to act

as a suggestion to greater efficiency. Those subjects who with amount-limit required more than 105 seconds, often completed the blank with that time limit, making as high scores for accuracy as with the longer time. The facts are:

TABLE XVIII

	Time Limit 105 Fifth Test Letters Marked	Amount Limit Eighth Test Time	Amount Limit Eighth Test Letters marked	Time Limit 105 Thirteenth Test Letters Marked	
Gr.	40	149	47	48	Marking B
L.	45	125	46	49	
M.	47	117	46	50	
Ba.	48	127	47	47	
	Sixth Test	**Ninth Test**		**Fourteenth Test**	
Gr.	39	111	41	41	Marking K
J.	49	110	50	46	
M.	42	134	45	41	
Ba.	37	127	38	37	

a — t Test.—The blank is as follows: parts A and B are generally used for separate tests.

(A)

A.

Dire tengo antipatia senores; esto seria necedad, porque hombre vale siempre tanto como otro hombre. Todas clases hombres merito; resumidas cuentas, sulpa suya vizxonde; pero dire sobrina puede contar dote viente cinco duros menos, tengo apartado; pardiez tamado trabajo atesorar-los para enriquecer estrano. Vizconde rico. Mios, quiero ganado sudor frente salga familia; suyo, pertenence, tendran. Conozco marido pueda convenirle Isabel; Carlos, sobrino. Donde muchacho honrado, mejor indole, juicioso, valiente? Quieres sobrino. Esposo parece natural, pero. Pero, pero, diablos, objeciones hacer. Posible quedandonow solos siempre hacer oposicion. Solo delante hentes eres ministerial. Pues, sidens siempre plan, dicho antes, porque hace tiempo notade cose aflige cierto. Sabes cuante quiero Carlos; consuelo apoyo; despues persona quiero mundo. Como eres buene amable, quieres porque, darme gusto, pero quisiera. Palabra cuesta trabajo; parece sino teines miedo agasajarle, manifestarle carino. Veces tratas cumplimiento veces senor. Probare; ejemplo pudiendo abandonar case negocios, deseaba hubiese acompanado viaje; preferiste sola sobrina doncella. Quise contradecir, pero para sentimento, para tambien. Voto gasta palabra, dice frases, dice; pero alla adentros quiere. Mientras estado malo, puesto dirigir casa; pardiez aunque carrera, hacia mejor; cabo tiene sobre ventaja poca edad, actividad zelo, pues para contigo digo. Siempre ordenes; dejaria matar alcanzarte billete para opera para baile. Necsitamos para felices; algo estrano, desconocido. Esta resuelto; supuesto hemos hablado esto, mismo, preciso empieces darle conocer nuestros planes. Quien mejor. Opone nunca deseos, sera facil nadie persuadirle. Probare menos, preciso sino creere tienes interes decidido proteger vizconde. Pudieras creer siempre inclinado senores cabra tira monte. Pero tengo nada ellos esposo tienes siempre pensativo siempre trists. Diablos tiene Carlos acercate tiene hablarte. Holo parece sacado letargo tengo algunas instrucciones cajero marcha dentro poco. Para empresa piensa usted establecer Habana.

Precisamente bonita especulacion bien manejada sobre todo. Espero poro tengo entre manos etro proyecto interesa aqui estabamos ocupando pienso. Eres porque

(B)

B.

quieres porque e tragas defensa peligro lugar huir mujer, harto debil duda pero algun desgracia tuviese luchar sentimientos semejantes tuyos, lejos ceder ellos cobardemente moriria pero triunfaria. Tendras menos valor tendre darte lecciones valor energia. Vamos, Carlos, amigo creeme sentimiento, profundo razon pueda subyugar, desgracia grande pueda soportar vencer nuestro corazon. Ofrezco apoyo eres creo sequiras consejos. Bied, hable usted. Quiere casarte Isabel. Isabel, prima imposible; quiere otro, vizconde amigo. Preciso persuadirselo hare otros partidos habra jamas para jurado nada espero pero conservare siempre entero este amor ella ignora unos juramentes recibido. Enhorabuena otro medio asequarara tranquilidad, uya destino ofrecido aleja Madrid, preciso aceptarle. Privarme presencia felicidad hecho usted para consejo especie embargo preciso seguirle solo puedes conservar amistad elige. Jamas caballero crei usted digno consejos dejo usted abandouado mismo nada tango decirle Carlos aleja, echa mirade salir Dona mira; suspira sale. Porque inquieta partida desterremos para siempre memoria quiero puedo presente temo; ausente, echo menos, verle sonrojo, nombre hace temblar. Embargo nunca dicho debiera ignorario Dios Dame fuerzas para resistir.

Subjects are told to mark every word that contains both an *a* and a *t*. If they look doubtful, examples are given of words such as *cat* which should be marked, and *paper* which should not. Even so, experience shows that further directions are often necessary even for educated adults. Some subjects mark the letters *a* and *t* in the word rather than the word; others do not mark a word unless the *a* precedes the *t*, others unless the *a* and *t* are together. A sample line with a judicious mixture of words correctly marked might be printed on the blank, and subjects told to look at it for a minute before the signal to begin is given. Those subjects who hit soon upon the device of looking for the rarer and projecting letter *t* first and then to see if there is an *a* as well, make better scores than the others. This method might be more easily suggested if the directions said "both a *t* and an *a*." Other letter combinations might be better.

Two "instructed" groups using the first part with a time-limit of 45 seconds marked, one an average of 11 words correctly, A.D. 2.5, the other an average of 10.2 words, A.D. 1.7. There was an average of 1.4 omissions for the second group, the greatest number being made by those below the average score.

The short-term group improved from 9.3 to 13.3 words correctly marked in their second test with the first division of the blank and from 7.5 to 10.7 words marked in the second test with the second division. Thus even over an interval of one or more weeks the acquaintance with the form of the test or the special blank or both has an effect of over 40 per cent. gain. The long-term group taking the

two divisions alternately gained in days 3 and 4 9.5 per cent. over days 1 and 2. In 20 days they improved from 15.6 and 9.6 words marked for the two divisions to 20.0 and 15. Apparently much of the improvement of the short-term group was due to familiarity with the form of the test rather than with the special blank.

Misspelling.—The blanks used are as follows:

(A)

MARK EVERY WORD THAT IS NOT SPELLED CORRECTLY

1. On the 3d of September, 1832, inteligence was broght to the collecter of Tinnevelly that som wildd eliphants had appeared in the neighborhod. A hunting party was imediately formed, and a large number of nattive hunters were engaged. We left the tents, on horsback, at half-past sevin o'clock in the morrning and rode thre miles to an open spote, flanked on one sid bye Rice-fields, and on the other by a jungle.

2. After waiting som time, Captain B——— and myself walked acros the riee fields to the shad of a tree. There we herd the trumpett of an elephant; we reshed acros the rice-fields up to our knes in mud, but all in vaiu, thogh we came upon the trak of one of the animels, and then ran five or six hundredd yards iutoo the jungle.

3. After varius false allarms, aud vane endevors to discuvor the obgects of our chace, the colector went into the jungle, and Captin B——— and myself into bed of the stream' where we had sen the traks; and here it was evedent the elaphents had passed to and fro. Disapointed and impasient, we allmost determened to giv up the chace and go home; but shots fird just before us reanimated us, aud we proceded, and found the collecter had just firred twicce.

4. Of we went throuh forest, over ravin, and through strems, till att last, at the top of the ravine, the elephants were seen. This was a momant of excitment! We wer all scatered. The collector had taken the midle path; Captain B———, some huntsmen, and myself took to the feft; and the other hunters scrabled down that to the rite. At this momunt I did not see enything but after advanceing a few yards, the hugh hed ef an elephunt shaking abuve the jungle, withen ten yards of us, burst sudenly upon my view.

5. Captain B——— ande a hunter justt befor me; we al fired at the same moment, and in so dirrect a line that the percussion-cap of my gun hitt the hunter, whome I thougt at first I had shoot. This acident, thogh it prouved slight, troubled me a litle. The grate excitement ocasioned by seeing, for the first tim, a wild best at liberty and in a state of natur, product a sensation of hop and fear that was intens.

(B)

MARK EVERY MISSPELLED WORD

I percieved, about four years ago, a large spiider in one korner of my room, makeing its web; and through the maid frequentely leveled her fatale brom against the lobors of the little anemal, I had the good fortoone then to prevente its distrucsion, and, I may say, it mor than paid me by the intertainement it aforded.

In thre days the weeb was, with encredable diligence, compleeted; nor could I avod thinkeing that the insect seemed to exult in its new abode. It often treversd it round, and exsamined the strenth of every part of it, retierd into its whole, and came out very ferquently. The first inemy, however, it had to inconter was

another and much larger spidur, which, having no web of its owne, and haveing probibly hexausted all its stock in former labors of this kind, came to invaide the prouperty of its nieghbore.

Soon a terreble encounter ensooed, in which the invader seemed to have the victorie, and the laborius spider was obleeged to take refug in its hole. Upon this I perceived the victer useing every art to draw the enemey from his strongholde. He seemed to go of, but quicklie returned, and, when he found all arts vane, began to dimoilish the new web withoute mercy. This broght on another battle, and contary to my expextations, the laborious spider became conckeror, and fairly killed his antagonist.

Nou in pieceable possession of what was justely its own, it awated three days with the uttmoste impatients, repairing the breeches of its web, and taking no sustenance that I could perceive. Ate last, houever, a large blue fly fell into the snaire, and strugled hard to get lose. The spider gave it leeve to intangle itself as much as possible, but it seemed to be to strong for the cobwebe.

I must own I was grately serprised when I saw the spider imediately sally out, and in lese than a minite wheave a new nett around its capthive, by wich the moshun of its wings was stoped, and, when it was fairely hampered in this maner, it was siezed and druged into the houle.

In this manner it lifed, in a precarious staite, and Natcher seemed to have fited it for such a life, for upon a singl fly it subsested for a weak. I put a waspe into the neat, but the spider sit it free.

To a class of 183 members blank B was given. In 30 seconds the average number marked was 18.3 at the first trial, A.D. 4.5, and 18.2 at the second trial, A.D. 3.4, when beginning at the third paragraph. There was a total of 34 errors in the first trial, 63 in the second. There were also 156 omissions in the first trial, 160 in the second, the mode being 1 both for errors and omissions, the average omission 2.8.

The short-term group made four trials with each blank beginning with the first and third paragraphs alternately, 8 tests in all. Their average on the A blank in a time limit of 30 seconds was 18.2; for the B blank, 18.8, or 19.6 for the first paragraph, 18.0 for the third.

The effect of practise and adaptation was as follows: the record with the two divisions of blank A in the first two sets was 13.1 words marked, 3.1 omissions for A1 and 18.8 words, 4 omissions for A2. In the seventh and eighth tests it was 17.7 words, 5.1 omissions, and 23.4 words, 6.1 omissions. If one word is deducted for each omission the individual scores become:

TABLE XIX

	First and Second Trial:		Repeated after Four Other Tests:	
	Blank A1	Blank A2	Blank A1	Blank A2
Bu.	7	19	15	16
Gr.	12	20	6	15
Ji.	8	14	6	4
Le.	5	9	15	17
Mo.	8	17	5	17
Ba.	16	26	23	27
Bf.	12	17	18	24
Average	9.7	17.4	11.1	17.1

The long-term group made 20 trials all with B blank, beginning at different trials with the first, second, third or fourth paragraphs. In a time-limit of 30 seconds their average was 28.4 correctly marked. For the first paragraph it was 30.5, for the third 23.9, with a very slight practise discernible which is here probably traceable to acquaintance with the blank. From the first four trials to the last four the change was only from 26.5 words to 28.8 and from 2.2 to 1.8 omissions. These blanks should be revised to make each of even difficulty throughout, and to make sure that the A and B blanks are of equal difficulty. The following table shows their present defects and also gives an approximate idea of the time required to find and mark a misspelled word such as these.

TABLE XX

	A Blank	B Blank First	B Blank Third		Seconds per Word
Class of 183		18.3	18.2	correctly marked	1.64
Instructed		16.0		correctly marked	1.87
Short-term	18.2	19.6	18.0	correctly marked	1.61
Long-term (first)		29.3	22.6	correctly marked	1.16
Long-term (average) .		30.5	23.9	correctly marked	1.11

At the end of the 20 trials, each of the three subjects completed the blank, *i. e.*, the amount-limit method was used. Two subjects were slower by this method, the third quicker than she was on the average by the time-limit method. This one subject, who was the most rapid in this test, did not with the amount-limit method exceed her maximum speed with the time-limit method. The following table will make this clear.

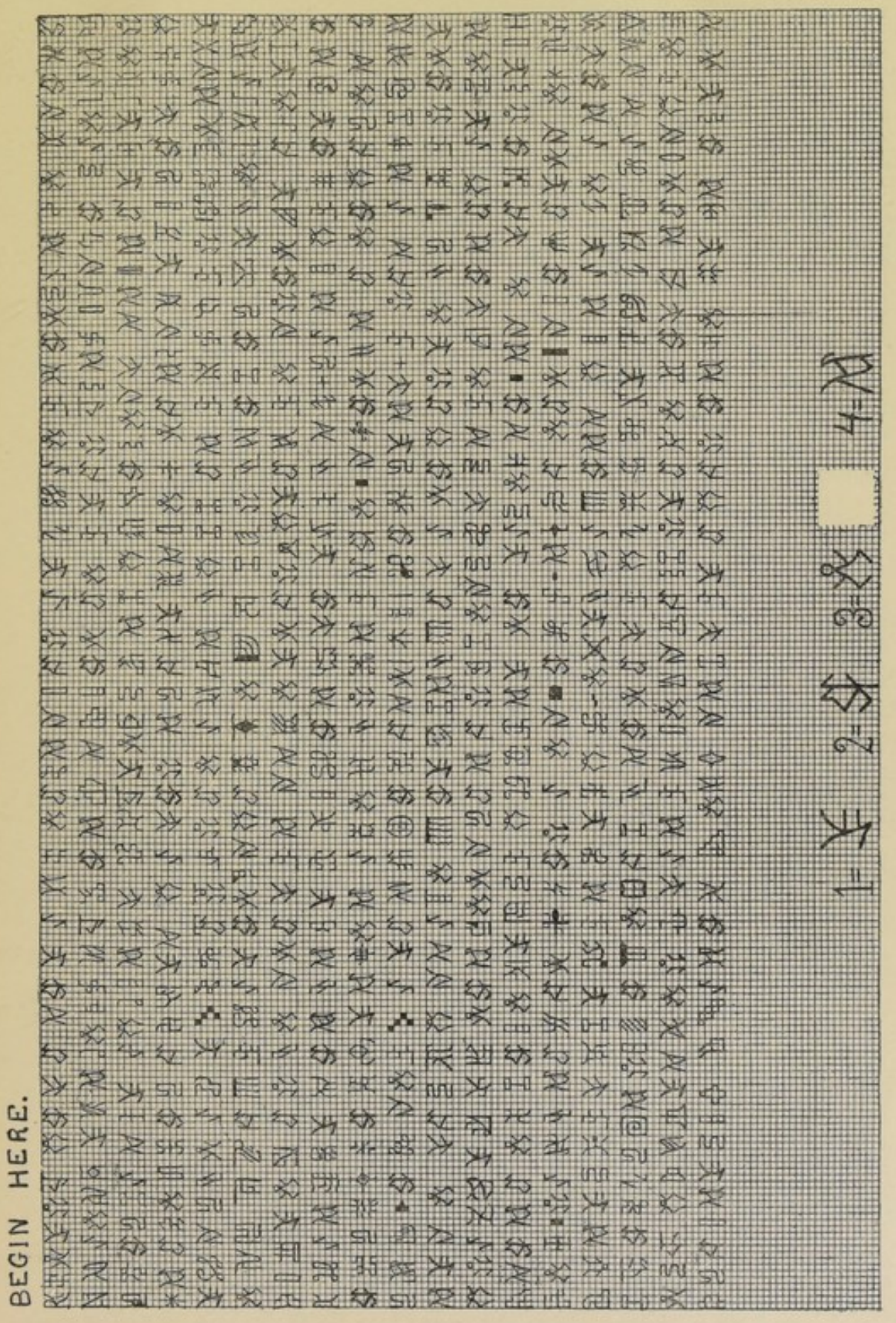
TABLE XXI

MISSPELLING TEST

	Subject	Time	Right	Wrong	Omitted	$R-(W+O)$
Record in last four tests, Blank B, beginning at ¶ 1, 2, 3, 4, 30 sec. each	N.	120	108	1	6	101
	W.	120	111	0	10	101
	F.	120	124	0	6	118
Record in amount-limit test	N.	118	92	1	7	84
	W.	130	94	1	5	88
	F.	93	98	0	1	97

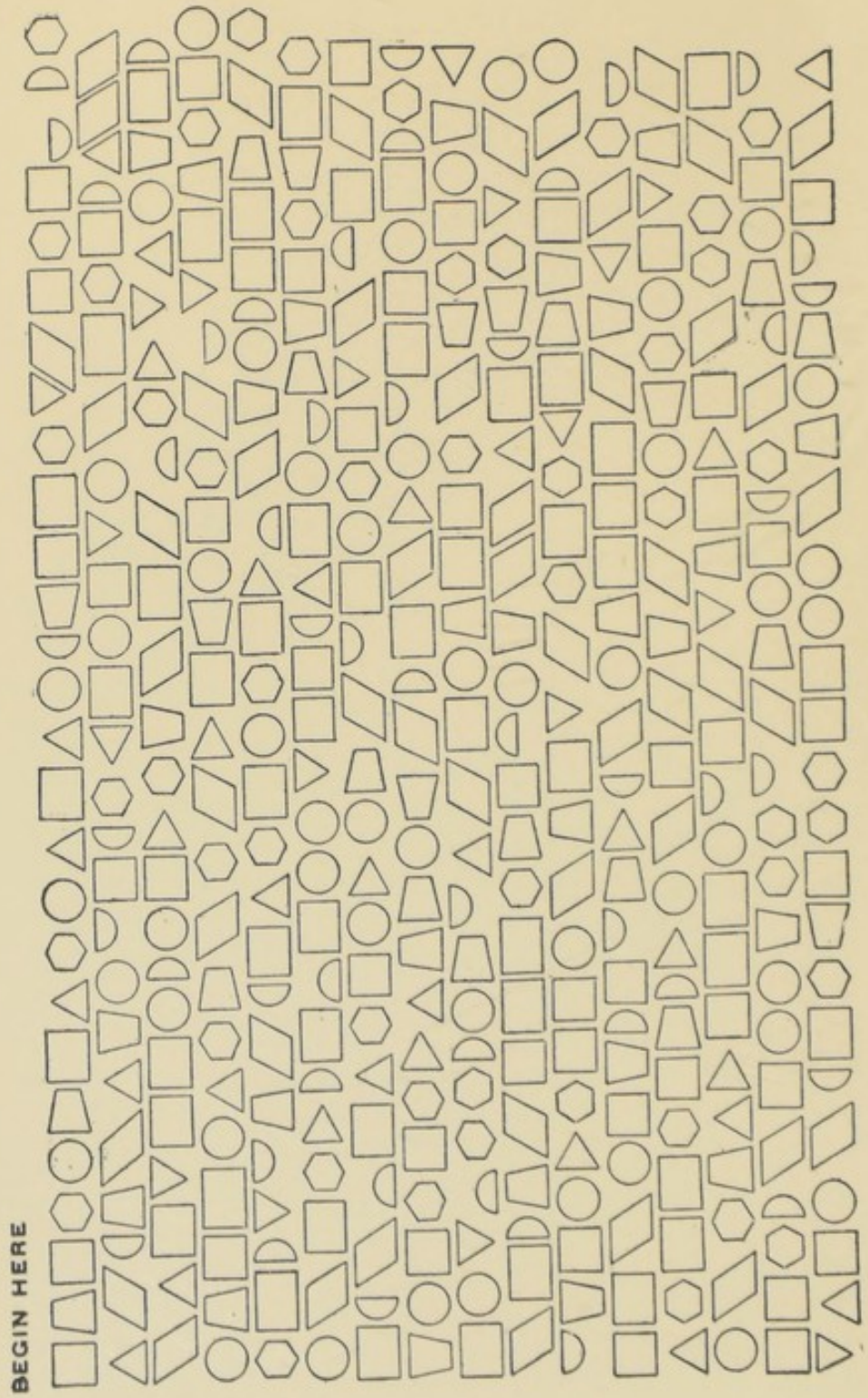
N. lost approximately ..... 15 per cent.
W. lost approximately ..... 13 per cent.
F. gained approximately ..... 6 per cent.
Approximate average loss by amount-limit ..... 7 per cent.

Perception of Forms.—The two blanks used were as follows:



No. 1

No. 2 is very convenient as it has eight different geometrical forms of which there are 50 each on the sheet: it is thus to some de-



gree comparable with "A" blank No. 3. The square and rectangle may, however, be easily confused, and for that reason were not used. No. 1 has four forms of which there are but 50 each; but in the first place this blank is exceedingly trying for the eyes, and in the second

place forms No. 1 and 3 are not easily and rapidly distinguishable from other forms that appear fairly often. The long-term group had to use this, however, as at the time of their practise the other blank had not been prepared.

Blank No. 2 was given to the "instructed" group with directions to mark every triangle. The time limit was 60 seconds. The average number marked was 35.2, A.D. 5.8, or 1.71 seconds per triangle.

The short-term group made two trials marking the trapezoid in each case. The time limit was 70 seconds. The average number marked was 39.3 (A.D. 3.1) in the first and 41.4 (A.D. 3.9) in the second trial.

Tests were made also with five other forms, but as the subjects after completing all the lines looked back to seek omissions, instead of reporting themselves as having finished, the records are not usable to estimate either practise effect or the difference in difficulty of the forms. The circle and semi-circle are proved to be much easier than the trapezoid, since within 60 seconds the blank was completed by all for the circle (Av. No. marked 48.3) and by three out of seven for the semicircle (Av. No. marked 42.4, Median 41). The last measure is valid, so that we may assume the trapezoid to be approximately a sixth harder to locate than the semicircle on this blank.

This group made also two trials with blank I. They were told to study the selected pattern at the bottom of the sheet on the word "go," till the signal "now," when they were to mark as rapidly as possible every one exactly like it till the signal "stop." Five seconds was allowed for the study, 55 seconds for the marking. With form 1, their average was 13 marked, with form 2 it was 10.6.

The long-term group made 20 trials with blank I following the directions given above. As they took the different forms in rotation they had only five trials with each form. The average for any form was 19.4, the first four trials' average deviating by — 3.2, the last four by + 2.8.

This and the *a — t test* gain from repetition with the same blank far more than do the *A test* and *misspelled word* tests. The gain would appear therefore to be due more to becoming accustomed to a novel problem in identification rather than to partial memorizing of the positions on the blank. The latter should have been most influential in the *A test* when repeated 20 times with just the same arrangement of objects to be marked.

On examining the records to see if one form benefited more than another, it was seen undoubtedly that form 2, subjectively the easiest, benefited most, and form 4 next. The average number marked in the five trials with each was respectively, 24.6, 22.2, 18.4,

12.3. Thus No. 4 proved the most difficult. Errors and omissions were not counted on this blank, as it was judged that its difficulty put it on altogether a different plane from the *A*, *a* — *t*, and *misspelled words* tests.

At this point some note may be taken of the speed attained in these tests. The process required is so similar in all of them—to look for some special thing, and mark it when seen, that more uniformity in speed might be expected than was found among the association tests. One test classified under association requires this same process of checking rather than writing words or parts of words, and the consideration of speed in that was deferred for comparison with these tests. It was the marking of nonsense syllables and English words out of a mixed list. For purposes of comparison, all are reduced to the time required to find and mark one object of the specified sort. The conditions of the surroundings of the object must be kept in mind in considering these figures.

TABLE XXII

SCORES IN EARLY TRIALS

	Sec per Unit Found and Marked			
	Short Term	Long Term	Various Instructed	
100 A's amongst 400 other letters	.83	.64	1.06	.94*
50 A's amongst 650 other letters	1.87		2.00	
50 B's amongst 650 other letters			1.83	
50 K's amongst 650 other letters			2.22	
50 S's amongst 650 other letters	1.61		1.62	
50 triangles amongst 350 other forms			1.71	
50 trapezoids amongst 350 other forms	1.74			
50 semicircles amongst 350 other forms	1.44			
Misspelled words amongst 300 other words	1.61	1.16	1.87	
25 nonsense syllables amongst 75 confusion words	3.10			

* Columbia and Barnard students.

From the difference found in marking A's, it is evident that the arrangement of the blank itself and the possible number of units to be examined is one of the largest factors in the rate of marking.

Another test commonly classified under perception tests, though totally different from all so far described, is that known as "perception of size." The freshmen are given a sheet of paper bearing a 5-cm. line, which is placed to their left, also a blank sheet of paper. They are asked to draw a line the same length as the standard without moving the papers or measuring in any way, then to bisect the line drawn, then to erect a perpendicular the length of the line. Columbia freshmen are also asked to bisect the right-hand angle.

The men make an average error of 2.4 mm. in drawing the first line, the women 3.7 mm.

The records of three graduate men students who made 50 trials in five sets of ten each of drawing a line equal to a standard line, were examined. These three were chosen at random from a class of eleven. The average errors in 50 trials were respectively 2.3 mm., 3.7 mm., and 1.8 mm. *A* changed from 1.5 for the first group to 4.5 in the last, mainly on account of developing a positive constant error. *B* changed from 1.6 to 5.5, also because of a large positive constant error. *C* changed from .7 to 1.0, his larger average for the total series being influenced by a negative constant error in the fourth group.

The short-term practise group made ten trials of each of the four processes required of the freshmen, after taking the test as a whole once. Unlike the method in other tests, they made all ten trials of one process at one sitting, as the three subjects *A*, *B*, and *C* had done.

The results were, in terms of error:

	Av.	A.D.
Line	3.4 mm.	1.8
Vertical	5.7 mm.	3.9
Bisect line	1.5 mm.	1.0
Bisect angle	3.2°	1.7

As might be expected from the illusion involved in erecting the perpendicular, the largest error is found there, and is a negative constant error. The average for drawing the line equal to the standard is very near that of the Barnard freshmen. No subject did equally well in all four processes; in fact the one with the least error in drawing the line made the greatest in bisecting the line, and another who made the least error in bisection of the line made the greatest in erecting the perpendicular.

No practise effect was discernible in the ten trials, and since the tendency of a rather longer practise is to confirm a constant error, the earlier trials may perhaps give more accurate results, though they may not reveal individual differences in habituation.

B. Relative Value of these Tests on Perception

There can be no question that in freedom from ambiguity due to measuring, in early trials, a combination of ability to perceive objects and ability to get used to the form of a test the *A test* and geometrical *forms* test are markedly superior to the *a — t* and the *hieroglyph* tests. There is some uncertainty with respect to the *misspelled words* test, but it is at least probable that the first trial with it is largely influenced by a person's ability to set his mind to

the novel task. It is unnecessary to repeat details here as it will appear that for other reasons the *misspelled words* is an undesirable test.

The question of the significance of these tests of perception as shown by their correlations was next studied.

First of all the performances of the eighteen instructed subjects were compared in the four tests, *A*, *a* — *t*, *triangle* or *perception of forms* and *misspelled words*. Each test was compared with the average for all four.

The coefficients are:

TABLE XXIII (*a*)

	Tests	Cos πU	*r*	Av.	Order
Average of these four tests and	Perception of geometrical forms	.90	.65	.78	1
	A	.34	.82	.58	4
	a — *t*	.81	.49	.65	3
	Misspelled words	.64	.85	.75	2

Next the first two trials of the short-term practise group were compared in seven tests—*a* — *t*, *e* — *r*, *A*, *misspelling*, *perception of forms* (2 blanks), *perception of size*, each with the average of all. The results are:

TABLE XXIII (*b*)

	Tests	cos πU	*r*	Av.	Order
Average of these seven tests and	Perception of geometrical forms	.90	.83	.87	1
	Forms 1 and 2 (hieroglyphs)	.48	.16	.32	
	A	.61	.65	.63	3
	a — *t*	.90	.72	.81	2
	Misspelled words	0	— .35	— .18	4
	e — *r*	.90	.57	.74	2
	Perception of size	.22	.54	.38	

Next, the performances of the long-term group were compared in the four perception tests with which they practised. For this all the 20 records for each subject were averaged. As there were four forms in the *perception of forms*, and two parts to the *a* — *t* blank it was all the more advisable to avoid making any selection from the total number of trials. It should be noted that this group used different blanks in the case of the *A* test, and *perception of forms* from those used by the other two groups, also that in the *A* test these subjects reached something presumably near the physiological limit.

The correlations were:

			Order
Average of these four tests and	Forms 1, 2, 3, 4 (hieroglyphs)	$r = .87$	3
	A	$r = .88$	2
	a — *t*	$r = .98$	1
	Misspelled words	$r = .79$	4

It appears that even so few as two tests of approximately a minute with the *A*, *a* — *t* or geometrical *forms* tests are significant of an individual's ability in visual perception. Amongst these three tests there is little choice. The geometrical *forms* test is perhaps the most typical of the general function in question, but both the *A* or the *a* — *t* are satisfactory in this respect.

The precision of the otherwise desirable tests of perception was measured, as for the association and memory tests, in terms of the average divergence of the result obtained from a single trial from the individual's true total ability, and the amount is expressed, as before, in per cent. of the former.

TABLE XXIV

RELATIVE PRECISION OF PERCEPTION TESTS

Test	Time in Seconds	Probable Average Divergence of the Result Obtained from 1 Trial from the Probable True Result, in Per Cent. of the Former: Short Term	Long Term, Early
A (Blanks 1 and 2)	60	5.4	2.8
S on blank 3	35	5	
a—*t*	45	7	4.6
Misspelled words	30	10	5.4
Forms (trapezoid)	70	4	

Here again, marking letters, marking words containing certain letters, and marking geometrical forms are all fairly satisfactory with little to choose among them. On the whole perhaps the *A test* and *geometrical forms* used together would be the best. The latter has the advantage of being uninfluenced by habituation to any one visual alphabet, and is therefore adaptable to more kinds of people, *e. g.*, young children or members of different racial groups.

4. TESTS ON DISCRIMINATION

A. Descriptive

Another test given the freshmen is that of *naming 100 colors* as quickly as possible. 100 1 cm. squares of 10 different colors are arranged in chance order on a white ground. Care is taken that the students have a ready name for each color there before beginning the test; then they are asked to read off—or name—all the colors there as rapidly as possible, while the time taken is noted. A name like "old rose," preferred by some students to "pink," makes an appreciable delay, so that it might be better to have 10 indisputable shades, or even briefer names assigned in print to a sample row.

The men take 85 seconds on the average (P.E. 14) to read the 100 colors, and the women 67.2 seconds. Here, as in the marking 100 A's, the women are quicker than the men.

The short-term group made 6 trials with this test individually. Their average time on the first trial was 56 seconds; for the total series it was 53.1 seconds, with A.D. 9.9. In half the cases there was a slight practise effect discernible. The A.D. of the successive averages was only 1.2. The successive averages were 56, 54, 51.5, 51.7, 51.8, and 53.

The long-term group made, as usual, 20 trials, using a rather smaller piece of apparatus. Their average time was 46.7 seconds, the first trial's average deviating by + 16, the last by — 4. The greatest gain was made from the first to the second trial. The first six averages were 62.7, 49.6, 50.8, 48.1, 50.9, and 46.6. It was interesting to note that the most rapid talker was considerably the slowest at the beginning of this test, though by the twentieth trial she had caught up with the second quickest. The one who did the best seemed to acquire her speed principally by careful economy of breath. On three occasions she read the 100 colors in 36 seconds.

At the end of the 20 trials each was asked to read off 100 color names without discrimination; that is, to move eyes and hand in pointing as before but to use the same word 100 times. The respective times taken for this were 37.5, 33, and 31 seconds, as compared with 44, 44, and 40 seconds at the 20th trial. The average extra time needed for discrimination beyond the mechanics of the test was therefore at the end 8.2 seconds.

Naming Forms.

Along with this test it was thought that comparison of forms and objects might be made, as similar material was being used in the memory and perception tests. Accordingly 100 squares were filled with 10 each of 10 different forms in chance order. These forms were star, cross, square, oblong, spiral, circle, "dots" (three dots spaced to form an equilateral triangle), oval, line, and triangle, and were drawn in ink or stamped from rubber type in black on a white ground. The whole resulting square was only four inches. Only the long-term group practised with this test. In 20 trials the average time taken was 53.3 seconds, the first day's average deviating by + 16.7, the last by — 5.3. Again the greatest gain was made from the first to the second trial. The first six averages were 70.0, 58.5, 59.2, 58.0, 57.6, 54.8. More errors in naming were made with this than with *naming colors*, though very few all told, a total of 9 for one subject, 6 for another, 4 for the other. Introspectively, these errors

are not due to faulty recognition but to difficulty in saying the right word; in the rapid enunciation the speech channel got blocked, or the "tongue twisted" as we say commonly, so that a circle would be called spiral, the subject being conscious of the error at the time of making it. Just here a question arises: the freshmen make slips in naming the colors too, and the directions should include advice about going on in spite of mistakes recognized as soon as made, or going back to correct them. Otherwise a considerable difference occurs in the time taken. The Barnard freshmen are told to go on usually, but in spite of this some conscientious students go back. Individual differences come out rather well on this point but escape the measuring rod of the statistician.

To return to the long-term group—the same subject was quickest in these two tests, but the other two changed rank. In neither of these two tests could there presumably have been any memory aid, as on successive trials the apparatus was turned round and the reading begun from a different corner.

Naming Objects.

A third test was devised, that of naming 100 objects. Owing to the trouble involved in collecting these and setting them out on a small table, four readings were made on the same day by each subject for five separate days, instead of one a day. They began at a different corner for each reading, however. The objects included keys, spoons, nails, screws, corks, pencils, books, tumblers, hairpins, spools, paper, matches, candles, checkers, picture-hooks ("hangers"), boxes, bottles, flowers, leaves, berries—all small but familiar objects, arranged again in chance order in 10 rows of 10. Introspectively this was a harder test, the space taken up in three dimensions seeming to confuse the subjects. The average time taken was 56.2 seconds, the first trial's average deviating by + 8.4, the last by — 1.3. The greatest gain was made from the first four readings to the next four, not from the first, to the second, nor was there any marked improvement from the first to the second reading on any one day. The first eight averages were—64.6, 61.3, 65.1, 59.9, 54.3, 53.9, 53.1, and 52.3. It may be therefore that the particular combination and arrangement of the objects on the first day was more difficult to read off than on any other day; or else that the new, strange feeling persisted through all four readings on the first day, but disappeared on the second occasion when four readings were to be made.

B. Relative Value of these Tests on Discrimination

First the correlation of these tests was examined.

Again all 20 records for each subject were utilized, as any selection of records seemed to measure the effect of practise at different stages.

The results were:

TABLE XXV

Average of these three tests and	Naming colors	$r = .67$
	Naming forms	$r = .99$
	Naming objects	$r = .96$
Naming colors and objects		$r = .45$
Naming forms and objects		$r = .93$
Naming colors and forms		$r = .73$

From this it would seem that *naming colors* is unlike the other two tests devised, as it does not correlate so closely with the average for the three as do the other two, nor are its intercorrelations close. *Naming forms* seems more a typical test in so far as it measures an ability common to these three tests. These relationships persist through "trial correlations" of selected records.

Unfortunately there were no records available from the "instructed" group to give greater weight to these correlations.

All three of these tests are of the same general degree of precision, color naming being somewhat the best. It is noteworthy that the individual variation of daily trials is so great in so simple a performance. The facts follow in Table XXVI.

TABLE XXVI

Test	Average Divergence of the Rate Found in One Trial from the Individual's True Rate. In Per Cent. of the Former			Time Per Trial in Seconds	Probable Number of Trials Required to Reduce the Unreliability to 1 Per Cent.
	Short-term Group	Long-term Group			
		Early Trials	Late Trials		
Name colors ...	3.8	6.6	5.0	50	26
Name forms ...		6.8	5.1	53	35
Name objects ..		4.6	8.3	56	42

Introspectively, naming objects is most unlike the other two tests; it is certainly the most awkward to use. In the memory tests, objects seemed to have the advantage over forms, but there, of course, there was no question of speed in making the test, and as mental speech was a distinct help in remembering, objects stood a better chance with their definite names than did unnamed forms. It could be wished that perception of colors had also been used, to make comparison possible between colors and forms in the two processes

of checking and naming, though the supposition would be that unless the colors were unequivocably distinguished some students might suspect it as a test of artistic taste or ability to match shades.

From experience with these tests it is suggested that *names of forms* would be less indefinite to read off than are those of colors; and as colors are apt to fade, the forms test has a slight advantage. The forms test is as easy to administer, is almost or quite as desirable from the point of view of susceptibility to practise and unreliability, and is perhaps more significant of the process of naming in general.

5. Discrimination and Motor Tests

A. Descriptive

Another allied series of discrimination tests was practised by the long-term group, but they are discussed separately as they involved a different motor reaction. The series included sorting ordinary playing cards by suit, similar sized cards by number, and small objects by size, color, or shape, making five tests in all. Similar tests have been devised before and used in such studies as Bergström's.[46]

Sorting Cards.—An ordinary pack of cards was well shuffled, and then, held face up, dealt out into four piles according to suit, the subjects choosing their own positions for the piles. Before making the first trial, each subject dealt a pack into four piles without discrimination of suit, as one deals when playing a game; the respective times taken in this preliminary trial were 17 seconds, 17.2, and 19, as against 26.4, 39.2, and 28.2 for the first trial with discrimination. Thus, the average extra time needed for the discrimination process was 13.5 seconds. The average time taken through the 20 trials was 26.5 seconds, the first day's average deviating by + 4.8 seconds, the last by — 2.7. Near the beginning there was no marked improvement; the greatest change occurred between the eighth and ninth trials. The slowest subject made a total of eleven errors, the quickest two, the other one none. On four days two trials were made in succession, and of the twelve records, there were five where the second trial took less time than the first.

Sorting by Number.—Compared with this was a test in which 60 cards—10 each of 6 different numerals, were to be sorted into 6 piles. These sets were selected from the complete pack of 150 used in playing "Flinch," care being taken not to confuse the eye by including 5's, 3's, and 8's in the same set of 60. Different sets were

[46] *Am. J. Psy.*, **6**, 24.

used on different days. On ten occasions the subjects knew beforehand what numbers to expect; on ten, they had to find out as they dealt. As before, they were at liberty to place their piles as they wished, but in this test the cards were held face down.

The average time for the 20 trials was 58.4 seconds, the first day's average deviating by +7.4, the last by —4.6. The greatest improvement occurred near the beginning, between the second and third trials. Comparing the ten trials when the numbers were known beforehand with those when they were not, there was an average difference of 2 seconds in favor of knowing them.

At the end of the 20 trials each subject dealt the 60 cards into 6 piles without discrimination. The times taken were respectively 24, 26, 25 seconds, as compared with 55, 55 and 51 at the 20th trial. The average extra time needed for discrimination was then 28.8 seconds.

Comparing the two tests—with the more familiar material, an easier manipulation and a narrower choice, a card was handled in .51 of a second on the average. With numbers, an additional movement, and six instead of four, possibilities, in .97 of a second. Eliminating the discrimination, before practise the playing cards were handled at the rate of one in .34 of a second; with the additional movement and after practise, the numbered cards at the rate of one in .42 of a second. This extra time is probably taken up by the turning of the cards. Unfortunately, trials by both methods with each kind of material were not made to make this point decisive. There is also the possibility that the pack of "Flinch" cards was less easy to handle than any of the three ordinary packs of cards.

The subjects held the same relative rank for speed in these two tests.

For the other three tests small objects such as pieces of thick cardboard, checkers, buttons, marbles, kindergarten beads, chess pawns, "halma" men, ping-pong balls, candle-ends, small spools and children's alphabet blocks were used. Three sets of 60 objects each were made up from this assortment, one to be sorted by size, another by color, the third by shape. In *sorting by size,* the objects were all discs, but varied in color as well as in thickness and diameter. In *sorting by color,* all sizes and shapes were included, and in *sorting by shape,* all sizes and colors.

The 60 objects were contained in a cardboard box; from this they were to be sorted into six smaller cardboard (shoe) boxes placed in a row. The subjects were at liberty as in the card sorting test to distribute as they wished rather than to memorize the experimenter's choice of the position of the different kinds of material. Usually the

three tests were taken one after the other with about two minutes' interval. The order was varied from day to day to equalize the interference effect. On the first day, each subject had the benefit of watching the other two do two of the tests, herself going through the third test in their presence before they did it. Otherwise these trials were made alone.

The general experience with these tests was that the subjects did not take any object that was nearest and then place it in the right box, but tried to get all 10 of one kind of object before beginning on another kind. This was not invariable however, as there was also a tendency to handle the largest objects first whatever they might be. No restrictions were put upon the subjects except that the objects were to be handled one at a time. This ruled out an ingenious device of one subject, of leaving the thinnest and flattest till the last and then pouring out all 10 at once straight from one box into the other. Careful observation showed that the training of the left hand played no small part in the gain in speed.

Sorting by Size.—The average time taken was 31.5 seconds, the first day's average deviating by + 4.3, the last by + 1.7. The best record was made on the 18th trial. In all 60 cases there were but five errors.

Sorting by Color.—The colors were black, white, red, blue, green, and yellow. The average time taken was 33.5 seconds, the first day's average deviating by + 7.0, the last by + 2.0. The greatest improvement came between the second and third trials. The best score was at the 16th trial.

The most rapid worker made eight errors, the other two five each. Thus there was greater inaccuracy with the color discrimination than with the size.

Sorting by Shape.—The shapes were—cube, sphere, cylinder, disc, flat-square, and halma man (resembling a chess-pawn, but only three fourths inch high). The average time taken was 47.5 seconds, the first day's average deviating by + 10.4, the last by — 6.7. For the first nine trials the improvement was very irregular (av. 51.4, A.D. 3.7), but from the tenth trial on it was much more regular (av. 44.4, A.D. 2.1). The best score was the 20th. The most rapid worker made 14 errors, the next 12, the slowest 8.

Sorting by Size was least influenced by adaptation and practise, *sorting by color* next, while *sorting by shape*, though irregular in its course, showed a gain of from 25 to 30 per cent. in twenty trials.

This and also the time per unit of the process is shown by Table XXVII.

TABLE XXVII

AVERAGE TIME OF THREE SUBJECTS IN SUCCESSIVE DAILY TRIALS WITH THE SORTING TEST

Time Required Per Unit Sorted, in Seconds

Playing Cards Held Face Up, Into 4 Piles, by Suit	Cards with Large Numbers Held Face Down, Into 6 Piles, by Varying Number		Sorting 60 Objects		
	Number Known Beforehand	Number Unknown Beforehand	By Size Into 6 Boxes	By Color Into 6 Boxes	By Shape Into 6 Boxes
.60	1.10		.60	.68	.98
.60		1.09	.57	.64	.87
.58	1.00		.53	.56	.98
.62		1.02	.55	.54	.88
.58		.98	.52	.52	.74
.56		.93	.52	.58	.85
.59		1.07	.54	.57	.82
.53		.96	.55	.55	.76
.47		.99	.55	.54	.84
.48		1.03	.45	.53	.73
.44	1.01		.47	.51	.74
.43		.96	.51	.52	.72
.49	.94		.49	.55	.80
.48	.97		.54	.58	.81
.46		.93	.50	.53	.78
.45	.96		.54	.52	.77
.46	.92		.51	.58	.72
.47	.89		.49	.56	.72
.43	.93		.55	.58	.68
.46	.90		.55	.60	.68

Comparing all three tests, the same subject was quickest in all of them, and was also the second quickest in the two card sorting tests. Neither of the other two kept the same rank throughout. In the average time taken, it would have been expected that *sorting by size* might be different from the others, as there was not quite the same variety in the material, and the objects were slightly more tiresome to handle. However, the average times for *size* and *color* are about the same, 32 and 34 seconds, while that of *shape* was considerably longer, 47 seconds. Introspectively, *sorting by shape* was the most difficult, perhaps the least familiar way of regarding things.

B. Relative Value of these Discrimination-motor Tests

These various "discrimination-motor" tests were correlated, using as before all available records from the three subjects of the long-term group. The results were as follows:

TABLE XXVIII

Average of these three tests and	Sorting objects by shape	$r = .68$
	Sorting objects by color	$r = .98$
	Sorting objects by size	$r = .99$
By shape and by color		$r = .54$
By shape and by size		$r = .55$
By size and by color		$r = .98$
Sorting cards by number and by suit		$r = .96$

From this it appears that sorting by shape is most unlike the other tests, agreeing with the introspective evidence and the observer's notes at the time; otherwise, all the correlations are close. If however we include the two tests with cards and correlate each of the five with the average of all five sorting tests, sorting by shape is found to be the best representative. One individual who was the slowest in sorting objects by *size* and *color* and in the second place in sorting objects by *shape* was the most rapid in both tests with cards and the correlations became:

TABLE XXIX

Average of these five tests and	Sorting objects by shape	.99
	Sorting objects by color	.52
	Sorting objects by size	.61
	Sorting cards by suits	.63
	Sorting cards by 6 numbers	.43

The measurements of relative precision on the basis of early and late trials of the three subjects show, as with the naming 100 colors, shapes, and objects, a large variation due to accidental causes including those which differentiate one day's condition from another. Even so simple a process repeated 60 times needs apparently from 10 to 50 trials, or from 8 to 30 minutes to measure a person within 1 per cent. *Sorting by size* is especially variable, and *sorting by number* least so. The facts are as given in Table XXX.

TABLE XXX

PRECISION OF SORTING TESTS

	Test	Probable Average Divergence of the Result Obtained from One Trial from the Probable True Ability. (3 Individuals) First Five Trials: In Seconds	First Five Trials: As Per Cent. of the Time Required by Individual	Last Five Trials: In Seconds	Last Five Trials: As Per Cent. of the Time Required by Individual	Approximate Average Time Necessary to Sort the 60 (52 in Case of D)	Approximate Number of Trials Needed to Reduce the Average Divergence to 1 Per Cent.	Approximate Time in Minutes Necessary to Reduce the Average Divergence to 1 Per Cent.
A	By size (60 objects) ...	3.0	8.6	2.9	10.3	31	88	45.5
B	By shape (60 objects) ..	4.4	8.3	1.4	3.3	47	34	26.5
C	By color (60 objects) ..	2.0	5.5	2.7	8.3	33	48	26.5
D	By suit (52 cards)	2.1	6.0	1.5	6.6	26	40	17.3
E	By number (60 cards) ..	2.0	3.3	1.5	2.8	58	9	7.7

From these facts, and from experience with the tests it is suggested that *sorting* small objects by *color* is a good test. It is less confusing than *sorting by shape*, yet can be varied more than *sorting by size*. In sorting cards one is confronted with the very unequal abilities people possess in their manual dexterity owing to previous experience; in using objects, the extra trouble in providing them is offset by the greater equality in experience of subjects at the start. Otherwise, pictures, words, figures, geometrical forms, material in great variety can be prepared on cards.

6. Tests for Speed and Accuracy of Movements

A. Descriptive

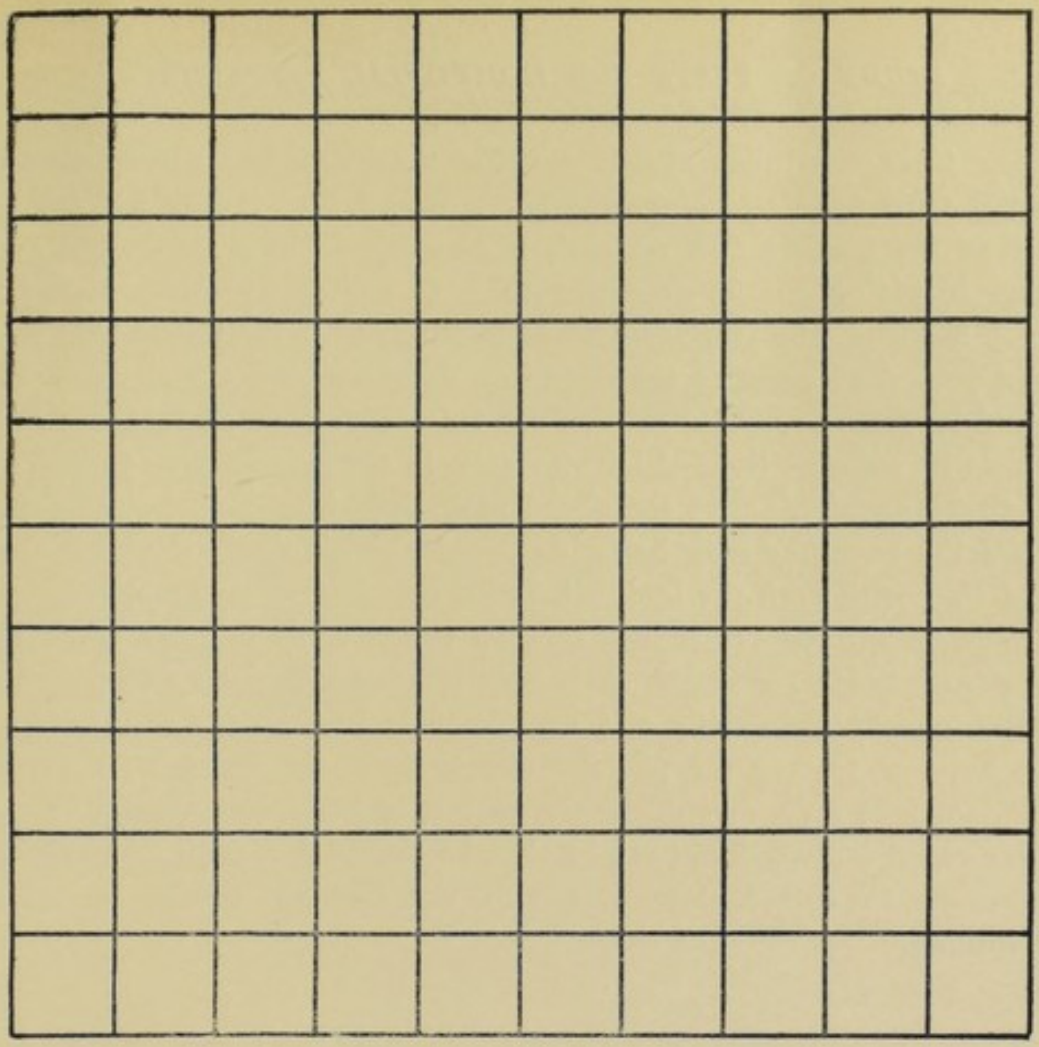
To the freshmen is given the following blank with directions, for the first half, to place a dot in each square as rapidly as possible.

The average time taken by the men is 34 seconds, P.E. 4; by the women 30.8 seconds.

In the second half of the test the subjects are required to strike each dot. The average times taken are 49 seconds by the men, 45.5 by the women. The average error in accuracy has been measured only for the men; with them it is .8 mm.

Trials of this by the short-term group were not sufficiently numerous to develop a practise effect, but only to give a basis for correlation with other tests. Their average speed in the first half was the same as the freshmen's, though given by the time-limit method. This might suggest that an easy test such as this, where speed is the only thing emphasized, could be given by either method without suffering in rate. In the second part of the test, the short-term group worked proportionately slower than the freshmen, making an average of 59 hits in 30 seconds (or needing 50 seconds to complete the test). Three fifths of these were not separated from the dot to be struck so that their average deviation from the mark might be called the radius of the pencil mark plus the radius of the printed dot (the latter is about .25 mm.). But the dot is often a very short dash and its radius or width varies so that such measurements are hardly of value. Wissler, who computed the average error of .8 mm. for the freshmen does not state how he computed it.

More attention was given by the short-term group to the various forms of *maze* tests that have been prepared. Of these the following five were used, known respectively as the *curved*, *straight*, *combined*, *black*, and *spiral*. The instructed and long-term groups used only the *curved*. The directions in each case were to draw a line between the two lines without touching either, working as quickly as pos-

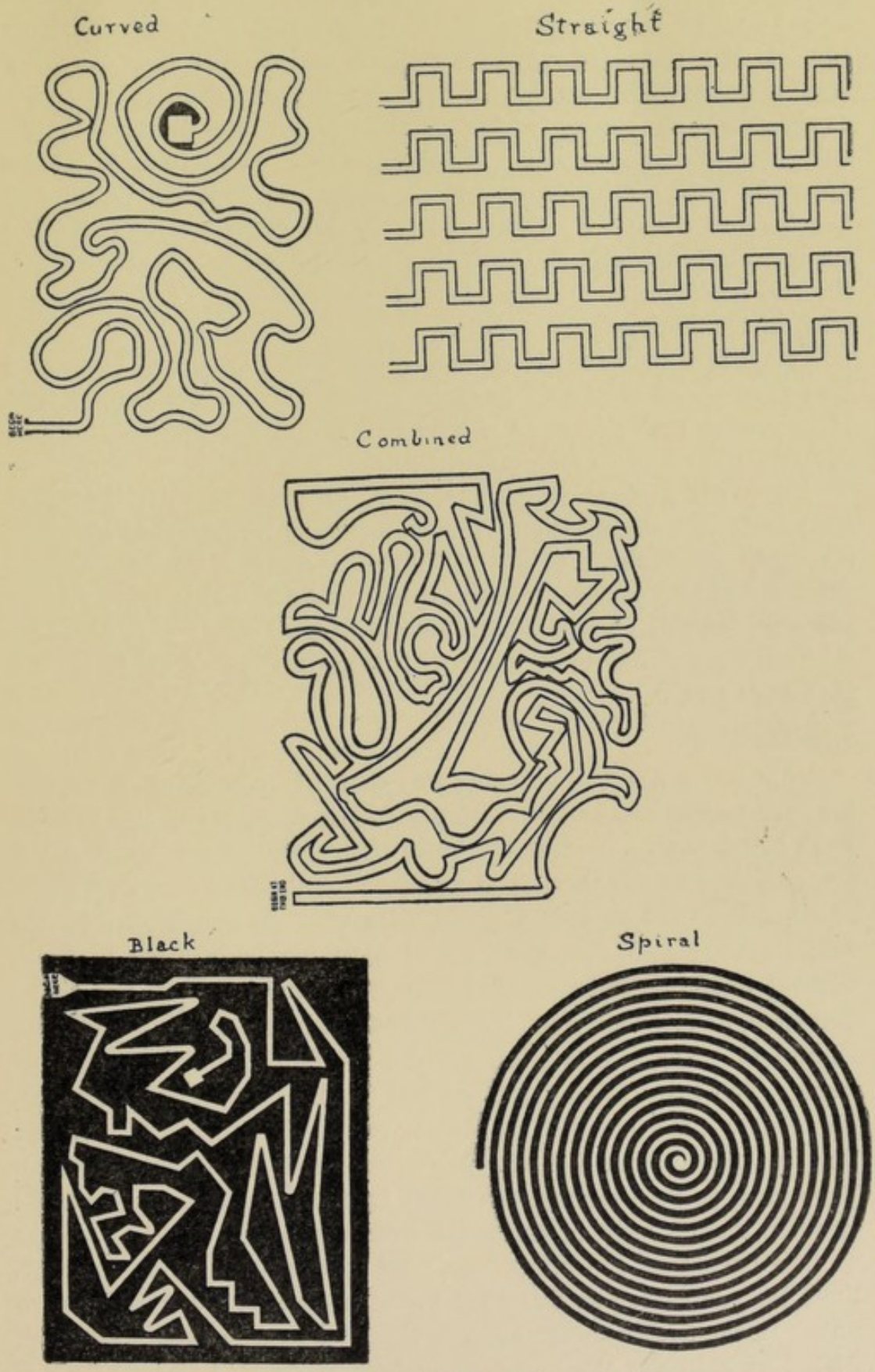


sible. Care was taken also to see that the blank was placed always in the same position before the subject, and that it was not moved during the tracing. In general, most subjects in a single test pay more attention to the accuracy than to the speed; with repeated tests, however, the emphasis tends to shift, with the result that in a long period of practise the accuracy decreases for a while and the speed increases very considerably. Once conscious of this, the subjects will redirect their chief attention to the accuracy so that after 20 to 24 days' practise the speed may have increased but slightly, while the accuracy may have improved a great deal. Having realized this, with both the instructed and the short-term-practise group —who, it will be remembered, were tested some months after the long-term group, although their results have here been noted first— the emphasis was chiefly and continuously laid on the accuracy, in the hope of getting the practise effects shown in speed, with errors constantly at zero, or sufficiently near it to be almost negligible. A more rapid improvement might thus be looked for, with unwavering attention to one factor, and also the scoring would be much simplified.

Curved Maze.

The instructed group used this as a time-limit test. In 60 seconds they traced (omitting one subject who completed the blank, but with 26 touches) 41.4 per cent. on the average, with 2.9 touches. The short-term group made three trials with this. The first two were amount-limit tests, with an average time taken of 169.5 seconds. The third trial was meant as a time-limit test and so announced, but all the subjects except one finished before the 165 seconds limit set. As in the *cancellation* test then and in the *first-idea* test, the announcement of time limit spurred on most of the subjects to work faster. Taking the three tests together, the average number of touches were 1, 3 and 1.

The long-term group made 20 trials with this as a time-limit test, using 60 seconds. The average amount traced was 76 per cent., the first day's average deviating by — 7, the last by + 1.6. The average number of touches was 11.3. In these subjects no steady improvement was noticed. N in the first five trials paid most attention to speed, with an average of 16 touches. In the next four trials, with more attention to accuracy the average number of touches dropped to 8, while the speed very slightly decreased. After this, her records were not so markedly irregular. W was most ambitious to complete the maze within the 60 seconds at least once. For this reason she began on the ninth day to spurt, succeeding on the thirteenth day



N. B. These are reduced to $\frac{2}{3}$ actual size.

in finishing. During this spurt her number of touches rose from an average of 12 to an average of 19, after which they dropped back again to 12. The third subject was slower and steadier than the other two. Finding, however, by the fifth day that she did not get so far as the others, she attempted for two days to put on speed with the result that her average number of touches rose from 6.5 to 15.5. Thereafter she paid most attention to accuracy and kept the number of her touches down. As these spurts by the three subjects did not occur simultaneously, the resulting average curve scarcely reveals the real conditions. On the whole there was a gain of 10 or 15 per cent. in the 20 days.

It appears then that if subjects work with the *curved maze* at a very high speed they gain perhaps one half of one per cent. a day. If they work with care so as to have only one or two touches they can increase their speed much more than that per day.

From these observations therefore, in practise with the other maze tests with the short-term group, accuracy was strongly and continuously emphasized, to see (1) if when errors were kept at zero there would be a practise effect in speed, and also (2) if there was an optimum time discoverable which could be used as a standard whenever such maze was to be used with large groups of subjects as a time-limit test.

Straight Maze.

This maze has two advantages—that of permitting a regular familiar movement, and that of presenting units easily measurable. Each blank can be used as the basis of five separate trials, and was twice so used by the short-term group. For the first five, time limits of 60, 50, 40, 30 and 30 seconds were set. At the beginning the subjects were told that they would have plenty of time to finish without touching, later on that they would have a little less. The first trial, of eight subjects two did not finish and two made touches (2 and 1). The second trial, one did not finish, and one made one touch. The third trial, three did not finish and one made a touch. The fourth time, six did not finish, two made touches (1 and 1). The last time, three did not finish, two made touches (2 and 1). Thus no gain in accuracy was made by the increase from 30 to 60 seconds, though most of the extra time was used.

The next time the blank was used it was given as an amount-limit test, or rather as five such tests, as each line was taken as a unit. In the five trials the average times taken were 29.3, 27.3, 27.9, 24.1, 23.5 seconds; the average numbers of touches were .4, .9, .1, .3, and .7.

The *combined maze* and *black maze* were used each only once with the short-term group by the amount-limit method. The average time taken for the combined maze was 294 seconds, A.D. 13; the touches were 2, 3, 5, 6, 12, and 13. The average time taken for the black maze was 202 seconds; the touches were 0, 0, 0, 1, 2, 2, 3.

The *spiral maze* was designed to provide another regular movement and one more natural perhaps, than the *straight*.

Endeavors were made to practise this keeping the touches at zero, and it was also hoped to practise with and without turning the paper, with wrist and with free-arm movements, beginning from the outside and from the center; but after a few trials this hope was given up, as all the subjects complained so much of eye-strain involved, and the unpleasant after images.

The average times taken in successive trials were 360, 360, 298, and 316 seconds. The average number of touches was in the first trial 2.3; in the second 2.8; in the third 2.4; in the last 2.0. The time taken would alone show how tiring to the eyes this might be, staring at a heavy black spiral for over five minutes, and following the pencil point round dizzyingly. The number of touches was very low all through with one glaring exception when one subject decreased her time from 475 to 288 seconds and increased her touches from 2 to 13. In 27 records there were 6 of zero touches, 5 of 1, and 6 of 2.

Of the tests tried none are injuriously susceptible to adaptation to the task and practise. The *straight* maze is the easiest to score. The *spiral* is too much a test of ability to stand eye-strain. It would also be the easiest to use if the rate of the subjects was to be controlled so as to compare individuals in accuracy alone.

B. Relative Value of these Motor Tests

The data serviceable for correlation are given in Table XXXI. Having two records for each test, one of amount done, the other of number of touches in the case of a time-limit test—one of time taken,

TABLE XXXI

	Curved Maze Av. of 3 Trials		Straight 5 Lines		Black 1 Trial		Spiral Av. of 4 Trials	
Subject	Time	Touches	Time	Touches	Time	Touches	Time	Touches
Bu.	142	1.3	145	3	207	2	341	3.0
Gr.	136	3.7	147	4	224	0	315	3.5
J.	177	3.7	146	1	227	1	310	2.0
L.	182	1.0	112	0	225	0	359	.5
M.	147	.7	128	2	195	0	324	1.0
Ba.	126	0	125	4	154	3	397	2.3
Bf.	128	0	119	2	175	0	302	1.3

one of number of touches in the case of completing the maze—the resulting score must be arbitrarily determined, if a single measure for efficiency is to be used for correlations.

As a fairly just method 5 seconds per touch has been added.

The Pearson coefficients are then,

TABLE XXXII

Average of all four tests approximately equal weight in determining the average being given to each.	Curved maze	.60
	Straight maze	.49
	Black maze	.76
	Spiral maze	.29

The tests of rate of putting dots in the squares and of hitting the dots showed little or no correlation with each other or with these maze tests.

In estimating the relative precision of these tests of motor control two methods have been used. First, each individual's several trials have been expressed as deviations from the probable result, in view of the practise effect which he would have shown apart from other variations than those due to the general tendency to improve with practise. This is the result hitherto employed. Second, each individual's several trials have been expressed as deviations from the average score of all the group on that day, and then the average deviation of these deviations has been computed.

The following will illustrate the second method. The five successive trials with the *straight maze*, gave, as average times for the seven subjects, 29.3, 27.3, 27.9, 24.1, and 23.5. L, whose times were 30, 22, 25, 18, and 17 deviated by $+.7$, -5.3, -2.9, -9.9, and -7.1. The deviations of these latter from their central tendency (-4.9) were 5.6, .4, 2.0, 5.0, and 2.2, averaging over three seconds, or 13 per cent. of L's average time.

With the first method in the case of the short-term group additions were made to the time to compensate for the touches. With the second, no account was kept of touches. The results are given in per cents of the time taken. The probable average divergences of the score in one record from the individual's true ability are for the *curved, spiral,* and *straight* mazes in order 10, 6, and 6 per cent. by the first method, and 7, 9, and 9 by the second. Early trials of the *curved* maze with the three long-term subjects showed by the first method a corresponding figure of 7.3. Remembering the relative lengths of the time required it will be seen that the *straight* maze has a great advantage over the *curved* maze and a still greater advantage over the *spiral.*

Comparing all five maze tests as to the time taken to complete

with no touches, it is found that the *curved* and the *straight* take about equal time, 156 and 155 seconds respectively, the *black* takes somewhat longer—199 seconds, the *combined* 327 or more, and the *spiral* longest of all, 364 seconds. From the point of view of discomfort the *spiral* and the *black* are hardest on the eyes, and even the *combined* becomes somewhat dazzling when over five minutes is spent following its windings. For a short, convenient test either the *curved* then, or the *straight* maze might be used. This last has, as before mentioned, advantages of regularity of movement and ease of measurement, but to offset this, it may be suggestive of jerky, discrete movements by its very angularity; also the units are very small.

From all these indications the choice would lie between the *straight* for its convenience and precision, the *black* and the *curved* for their higher correlation. Of these two the first has also some disadvantages, already mentioned, which the others have not, and since the *black* is somewhat trying to the eyes and takes longer, the choice would rest upon the *curved* maze as a suitable and convenient second motor test. It would probably keep its present advantages and gain others if arranged in a series of straight lines each repeating some simple series of curves. The spiral maze has no merits.

7. Miscellaneous Tests

A. Descriptive

Six of the short-term group spent some time practising seven other tests that are usually given the freshmen, viz: perception of force of movement, with the monochord, the æsthesiometer and the algometer, all of which test perception in some form; each also practised 40 to 80 times with reaction time, 10 to 15 times with the dynamometer and 5 times with the spring ergometer, all three tests of movement in various ways. This work was done not so much to find out anything about each test when practised as to get a basis for intercorrelations when there was more than one trial of each—which is all the freshmen take—and to get a basis of comparison with some of the other tests already described.

With some few tests records of long practise were also available from two subjects who were making some cross-education experiments.

Perception of Force of Movement.—This is as often considered a test for perception of weight, or perception of distance. As described by Wissler[46a] the test is as follows: "the lift is vertical and the dynamometer gives a pressure of 1 kg. to 10 cm. A mechanical

[46a] *Psy. Rev. Mon. Suppl.*, No. **16**, 1901.

stop is provided at a pressure of 1 kg. to give the student his standard. In making the test he is told to lift the handle to the stop three times and then make ten (more recently five) attempts to lift it to the same height after the operator has removed the stop. Each lift is to be made in about 2 sec., with equal pauses between. A graphic record of the lifts is taken on a kymograph." The errors are afterwards recorded in cm. The men make an average error of 1.44 cm., the women of 1.8 cm.

The apparatus has been criticised on the ground that it is sure to induce a positive constant error because of the impact necessary in the first three trials while getting the standard. Even with directions to the Barnard students to be very careful in the first three trials, this positive error persists; and after even 75 trials with some of the short term group it was not overcome, though the subjects had the benefit of seeing their records after every 15 trials.

In tabulating the results only the average error was considered. Six of the short-term group and one member of the original long-term group made from 9 to 15 groups of 5 trials, and the two other extra subjects made 36 such groups of trials each.

TABLE XXXIII

ERRORS IN CM. MADE IN PERCEPTION OF FORCE OF MOVEMENT

Subject	Av. Error			No. of Groups of Trials
	First	Total	Last	
Ba.	1.06	1.70	.88	13
Bf.	1.52	.85	1.22	13
Bu.	2.12	1.29	.52	12
J.	1.74	.74	.22	15
L.	.32	.97	.44	9
M.	.80	.64	.20	10
N.	.74	.34	.46	10
R.	1.54	.65	.40	36
Wy.	.42	.67	.68	36

From the above table it will be seen that there is a certain amount of practise since the error is reduced in all cases except two. That improvement with practise is slow and irregular may be seen from the single records and even from the averages of the seven subjects for each successive group of five trials, up to ten groups, which were:

1	2	3	4	5	6	7	8	9	10
1.21	1.06	.93	.92	1.28	.73	1.24	1.04	.98	.76

The record is better than the freshmen records.

It might be better to require the subject to make a given number of movements of approximately the force shown him with the stop, each as nearly as possible equal in force to the one just made, and

to use the successive differences as the measure of his efficiency in the test.

With the *monochord,* the freshmen are tested for perception of pitch as follows: The instrument is tuned so that F below middle C is given when the bridge is at 75 cm. The tone F is given twice at an interval of about 2 seconds while the subject's back is turned. The bridge is then shifted and the subject told to find the tone given. The position is recorded. Then the original tone is given as before, and the bridge shifted to the place where it was left by the subject in his first trial; he is told this, and again required to find the tone. The position is recorded. Also, before the test is begun, the subject is shown how to use the instrument.

In general, if a subject is diffident, or slow in moving the bridge, or by chance tries at first tones a long way from the standard, he rapidly gets confused and forgets the original tone. On the other hand, a very good record at the first trial is followed frequently by a very poor one at the second, showing that in addition to memory and celerity in moving the bridge, something is due, with poor subjects, to chance. This seems to be a test of memory of pitch and of general intelligence in using the instrument as much as of perception of pitch.

Among the men 10 per cent. make an error of less than one tenth of a tone, 53 per cent. of one tenth to one tone, and 37 per cent. an error of more than one tone. For the women the corresponding percentages are 17 per cent., 63 per cent., 20 per cent.

TABLE XXXIV

ACCURACY IN PLACING A BRIDGE ON THE MONOCHORD SO AS TO PRODUCE A TONE OF THE SAME PITCH AS A REMEMBERED TONE; IN MILLIMETERS

Subject	Av. Error in mm.	A.D.	Av. Error on 75 Position
Ba.	37.2	26.0	24.6
Bf.	10.7	6.0	7.8
Bu.	7.2	5.0	4.2
J.	31.8	29.7	47.5
L.	9.1	5.0	10.5
M.	24.4	17.0	36.8
Average	20.1		
Average of successive records on 75 cm.	12	20.8	21 36 31 15

With this group of six subjects, after the preliminary trials, eighteen to twenty further trials were given on different days, using ten other standards ranging from 58 cm. to 93 cm. and also the original standard 75 on four more occasions. At their last trial they were asked to move the bridge till the tones on each side of it were of the same pitch, thus eliminating the memory factor. This

was of course done without looking at the instrument, though even so, only two subjects realized that the bridge would have to be in the exact middle. In this last trial the greatest error made by any one was a difference of 3 mm., whereas, as is seen in the table above, only one subject was distinctly good at the test given in the usual way.

The variability from one trial to the next, particularly in the case of those with poor records, completely disguises any practise effect, and emphasizes the need of more than one trial at the original test.

For sensation areas, "the points of the *æsthesiometer* are 2 cm. apart and the instrument is applied longitudinally to the back of the left hand between the bones of the second and third fingers. Five tests are made, the student being touched with one or two points in the order, two, two, one, one, two, and being required to decide in each case whether he was touched with one or with two points." Of the men, 63 per cent. are correct four or five times, of the women 52 per cent.

With six subjects the right and left hands were used alternately with the above series of touches twice each day for three days, twelve tests in all. The total average error for the R. hand was 40.5 per cent., for the L. hand 40.6 per cent., or practically no difference. As this means that they were correct only three times out of five on the average with either hand, they were rather below the Barnard standard. There was no discernible improvement with practise.

The *algometer* used has a pressing surface 1 cm. in diameter which is made of rubber. It is applied with gradually increasing pressure till the student signals that it is felt as disagreeable. Usually there is some little difficulty in making students understand just what is wanted. Some are nervous and afraid of receiving electric shocks, others consider it a test of endurance, particularly if it is given later in the series than the ergometer. With suggestible subjects too the judgment is apt to be based on the rate at which increasing pressure is applied. At the second trial with either hand when an equivalent time has passed the student will frequently signal "stop" though the pressure is only from a half to two thirds of what it was at the first trial.

The averages for the men are: R. hand 5.9 kg.; L. hand 5.6 kg.; for the women, 3.8 kg. and 4.3 kg. respectively.

The short-term group made eight trials with each hand on different days. Two subjects showed considerable difference from the first to the last trials, one changing from 7.25 kg. to 3.5 kg., the

other from 4.7 kg. to 2.5 kg. With the other four there was an average reduction of only .5 kg. The averages for the whole series of trials were: R. hand, 3.7 kg., L. hand 3.4 kg. The averages for the first four successive trials (both hands together) were 4.7, 3.9, 4.6, 3.7. There would thus be no very great advantage in making a first trial merely for adaptation to the test and using the second and later trials as the record. The test doubtless measures an individual's notion of the meaning of "painful" as well as his threshold for pain as he defines it. Even so it is a significant test; the correlation between the first eight and the last eight trials of the same individual is close.

In *reaction-time* the freshmen are tested five times in succession, with the Hipp chronoscope. The average of the five tests for the men is .159 second, for the women, .186 second.

The short-term group and the two extra subjects made from 40 to 75 trials each. Up to 30 trials, the average from each group of five was recorded, as well as each separate trial, after that the average from each group of three trials only. There is apparently a considerable effect from adaptation to the form of the test. The average times for the eight subjects in the first six successive 5-trial groups run 155, 158, 139, 133, 129, 130.5. This is also disturbing since the relative rates assigned to individuals from the first ten trials do not correspond at all perfectly to those assigned from say the next twenty trials. In these eight subjects the deviations were as follows:

TABLE XXXV

DEVIATION OF THE INDIVIDUAL'S AVERAGE REACTION-TIME FROM THE AVERAGE OF THE GROUP'S IN THOUSANDTHS OF A SECOND

Subject	First 10 Trials	Next 20 Trials
Ba.	+ 46.5	+ 20
Bf.	+ 5.5	+ 10
Bu.	+ 5	— 0.5
J.	—12.5	—11
L.	—16.5	—11
M.	— 1	— 6.5
R.	—10	+ 2
Wy.	—17	—12

These give a correlation of less than .09. The records of the first reactions correlate with those of the twenty from the 11th to the 30th by less than .07. It would seem worth while to take 15 reactions, discarding the first five.

With the oval *dynamometer* the freshmen make two trials with each hand in the order R. L.; L. R. The average strength of grip

found is for men, R. hand 36.3 kg.; L. hand 33.5 kg.; for the women, R. hand 25.8 kg.; L. hand 23.6 kg.

The short-term group made, on different days, from nine to sixteen trials, but this series also was not long enough to develop noticeable practise, with one possible exception. Their averages were as follows:

	R.		L.	
	Av.	A.D.	Av.	A.D.
First	21.8	3	19.8	1.8
Average	21.5		19.6	
Last	22.4	2.2	14.8	3.8

In this test a good deal of interest has attached to the question of whether the maximum strength is attained at the first or at the second trial, it being claimed that since a larger percentage of women reach their maximum at first than do men, and that the left or weaker hand in men is more apt to reach its maximum first than the stronger hand, that therefore to do so is a sign of weakness. However this condition goes with all degrees of strength of grip among the freshmen; and experience with repeated sets of trials with even this small group indicates that an individual may vary very much in the relationship of the first two trials. The following table illustrates this:

TABLE XXXVI

	Greater the first		Greater the second		Equal	
	R.	L.	R.	L.	R.	L.
Ba.	2	2	1	3	2	0
Bf.	4	2	0	1	1	2
Bu.	4	2	1	2	0	2
J.	4	3	1	2	0	1
L.	2	2	2	2	1	1
M.	3	2	1	1	0	1
Total	19	13	6	11	4	7

Too much must not then be argued from the comparison of only one set of trials. According to these records a single trial is subject to an average divergence from an individual's true ability of 9.5 per cent. The difference between two single trials would then be subject to an average divergence from the true difference of $\sqrt{9.5^2 + 9.5^2}$ or 13.4 per cent.

Cattell's spring *ergometer* is used for a test of fatigue with the freshmen. The student is shown how to work the instrument with particular attention to the use of only the end of the first finger on the top of the piston. He is instructed to press the piston down as far as possible fifty times without stopping. A rhythm of about

one a second is set by counting aloud at the outset. The reading on the dial for each ten pressures is recorded.

The men's average for the total amount of work done in the 50 pressures is 284.3 kg., the women's 172.9 kg.; the degrees of fatigue are 65 per cent. and 63 per cent. respectively.

The short-term group made five trials with this on different days. Their average amount of work was 267 kg., considerably nearer the men's than the women's average among the freshmen. There was the reverse of a practise effect from trial to trial, the average of the last was 254 kg. The percentage of fatigue likewise increased. With extended practise by the two extra subjects there was a similar falling off for the first eight days; then one of them reached and maintained her original level, and the other reached it and during the last seven days of the twenty-two days' practise, went far beyond it. As the average amount of work done for the first 10 pressures of the series varied scarcely at all, however, what practise effect was present was due to the increased power of endurance. The data for the comparison of these tests were scarcely reliable enough to warrant computing correlations by the Pearson coefficient. In general there seemed to be correlation between reaction time and speed of perception, and to be a slightly closer relation in speed in all the tests than in accuracy.

A summary of the results found in Section II. will be deferred till the end of the study.

III

CHANGES WITH PRACTISE

1. Methods of Measuring such Changes

Before taking up the work of individual differences and the practise curve, it would be well to take up some of the difficulties of interpretation due to the method of constructing such curves. Different units may be taken as the basis, the starting-point may be obscured by the use of percentile values only, and units may be differently equated, perhaps distorted, in different parts of the curve.

First as to the kind of units used.

Curves may be constructed in terms of decrease in error (a time or amount-limit test), decrease in time (amount-limit test), or increase in amount (time-limit test). Or, whether time-limit or amount-limit test, the scores may be reduced to the hundredths of a second required to perform a definite minimum of work such as adding two figures, cancelling one letter, etc. Bair, in his "Practise Curve,"[47] used units both of errors made after a given number of practises, and of number of trials necessary to eliminate all errors. His curves then slope down from left to right. Bryan and Harter[48] in their study of the acquisition of telegraphy used the number of letters tapped per minute. Swift[49] in his experiments with the typewriter used the number of words written during an hour, smoothing the curve by averaging each successive three scores. In later similar work undertaken with Schuyler,[50] two units were used, one of strokes made on the typewriter, one of errors made. His curves then—for no tables are given—show one a rise, the other a slight drop. Coover and Angell[51] in making tests on the vexed question of the general practise effect of special exercise, used variously the number of right judgments before and after training, the decrease in time in 100 reactions, and the similar decrease in errors. Where practise has meant a long period of exercise taken regularly on successive days, the unit may be the average deviation of each

[47] *Mon. Suppl. to Psych. Rev.*, 1902.
[48] *Psych. Rev.*, **4**, 1897, and **6**, 1899.
[49] *Psych. Bull.*, **1**, 1904.
[50] *Psych. Bull.*, **4**, 1907.
[51] *Am. J. Psych.*, **18**, 1907.

day's performances, giving a downward sloping curve for any one individual.

So long as only one individual's curve is being considered, or only the mean curve, the use of such varied units presents little difficulty; but when comparisons are to be made of the curves of learning whether of different subjects in the same test, or those of the same subject in different tests, it becomes important to know whether a different choice of units may show the same performance in two different ways, and whether the units are alike all through the curve. Otherwise, the questions "Does practise increase or decrease differences?" and "Who profit most by practise, those whose initial record is best or poorest?" may receive quite different answers according to the varied statistical treatment of identical facts.

There is considerable divergence of custom. One method has been to keep all scores in gross amounts, basing conclusions directly on them. Examples of this would be Swift's and Schuyler's work already referred to, and Smythe Johnson's experiments on motor education.[52] Let us call this the gross method.

Another method is to turn each score into percentile values of the initial record, or perhaps of the maximum reached before fatigue sets in. Examples of this are Gilbert's work on development of school-children,[53] Oehrn's on the work-curve of 10 subjects,[54] Coover and Angell as already referred to, and Wells in reports before the New York Branch of the American Psychological Association. Let us call this the percentile method.

Another way of expressing percentile values used by Smythe Johnson,[55] and modified by him from Amberg[56] is as follows: The difference between the first and second scores, first and third, and so on, is taken, and the sum of gains so found averaged and expressed in percentage of the first score. This process is repeated with the second score used as basis, again with the third, and so on through the series. Finally, all percentages are averaged. He says "The significance of such percentages is that they give us a true standard for the comparative influence of practise on different individuals" (page 61). That part of Amberg's method which was modified was, instead of averaging the $n-1$ different percentile values, to weight each one, multiplying the first by $n-1$, the second by $n-2$, etc., adding the products and dividing by $(n-1)+$

[52] *Yale Studies*, **6**, 1898.
[53] *Yale Studies*, **2**, 1894.
[54] *Psych. Arbeiten*, **1**, 1896.
[55] *Yale Studies*, **6**, 1898.
[56] *Psych. Arb.*, **1**, 1896.

$(n-2)+(n-3)\cdots 1$. According to Amberg the resulting figure "giebt mithin in möglichst einwandfreier Weise" the average percentile increase by practise for the whole test.

Just to illustrate to what various conclusions one may be led solely from differences in methods of portraying practise data, the following tables and figures were made from five supposititious cases.

In 15 seconds, using as a score units of gross amount, suppose that in seven trials, five subjects scored as follows:

TABLE XXXVII

GROSS AMOUNTS IN SUCCESSIVE TRIALS

Individual								Total Increase Units
A	5	6	7	8	9	10	10	5
B	9	12	16	16	17	17	18	9
C	10	10	10	12	13	14	15	5
D	6	9	11	12	12	15	18	12
E	5	7	9	10	12	14	15	10
Average ...	7.0	8.8	10.7	11.6	12.6	14.0	15.2	8.2
A.D.	2						2.25	

It might be stated then that D improves most, and A and C improve least.

This same table turned into units of time required to do one unit of work, using hundredths of a second as the basis becomes:

TABLE XXXVIII

GROSS TIME FOR WORK UNIT IN SUCCESSIVE TRIALS

Individual	Hundredths of a Second							Total Decrease
A	300	250	214	187	166	150	150	150
B	166	125	93	93	88	88	83	83
C	150	150	150	125	115	107	100	50
D	250	166	136	125	125	100	83	167
E	300	214	166	150	125	107	100	200
Average ..	233	181	155	136	124	110	103	130
A.D.	60						19	

It might be stated now that E improves most and C improves least.

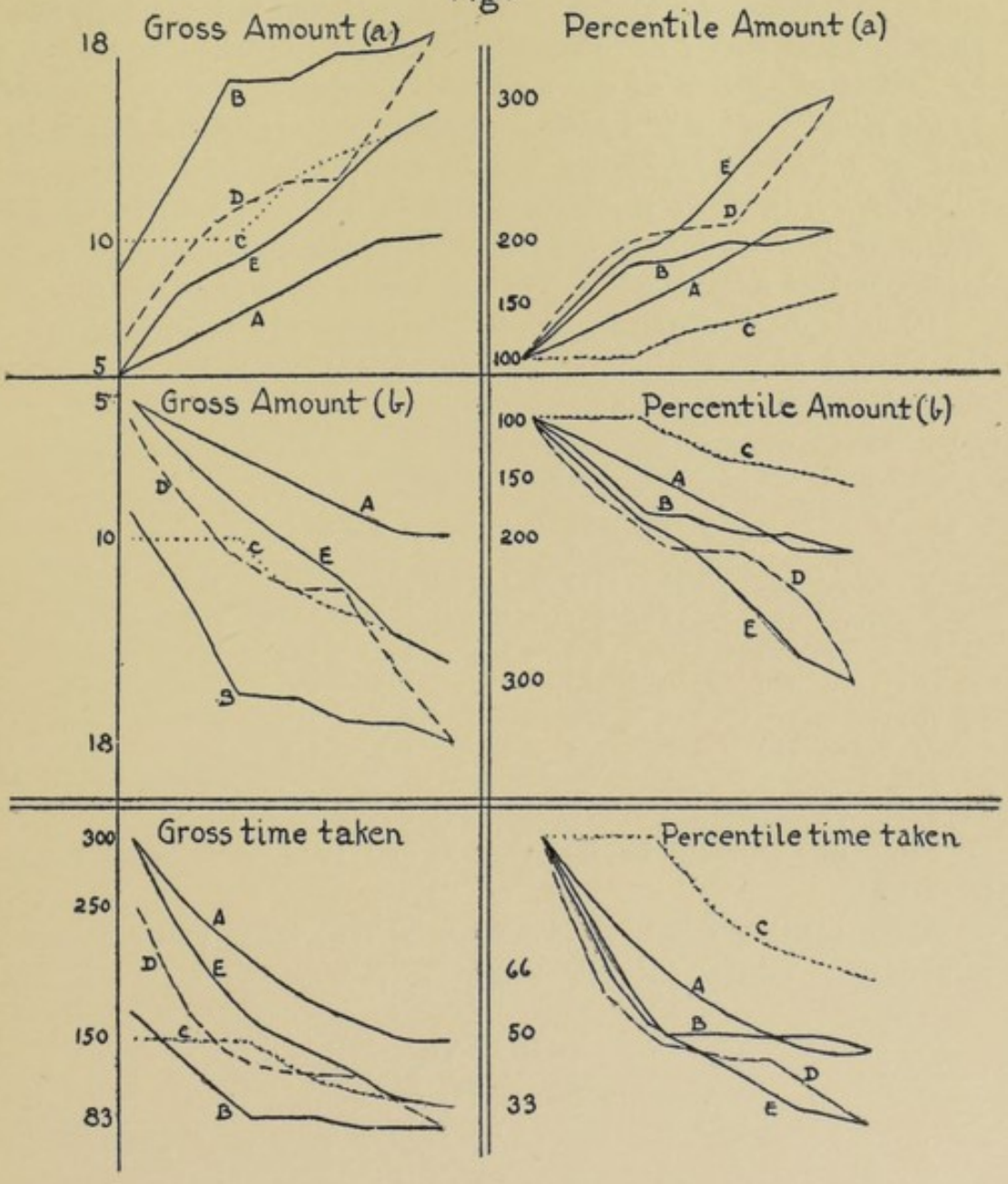
The two sets of curves as plotted* are not strictly comparable, except that the same individuals are alike at the starting point in each, and at the end. Otherwise, in answering the question whether differences are increased or diminished by practise, the curves show graphically that in the first case they apparently are increased, in the second considerably decreased. The tables show the same thing,

* See Fig. 1.

if the A.D. for the first trial is compared with the A.D. for the last, in each table. In the first case there is a slightly greater difference at the end, in the second, there is less.

The inference is then, that the change from the use of one kind of unit to another in expression of one and the same performance makes an appreciable change in its interpretation.

Fig. 1.



Suppose however, as is sometimes the case, it were desirable to compare one individual quantitatively with another, it could be said from the first form of presentation that A and C improve equally, and half as much as does E; and that B improves three quarters as much as D. In the second case it might be said that no

two subjects improve equally though A and D are nearly equal; that A improves three times as much as C, and three quarters as much as E.

Evidently the value of such statements would be conditioned by the nature of the test, for units near the physiological limit would not be equal to those in the lower ranges. In a test such as mental multiplication, the gain of the last few units may be far more difficult than that of the first many. In a cancellation test, the units may possibly be of rather more equal difficulty, conditioned as they are by factors of amount of eye movement necessary, and rejection of wrong stimuli. In a feat such as juggling with balls, the first three or four units may be harder to gain than fifteen such units later. In other words, sharp slants or a plateau may be produced by the nature of the variations in the real value of the units scored as equivalent, so that a "typical" curve for certain work may really exist.

If, as is more customary when individuals are to be compared, the method of percentile values is used, the above table of gross scores becomes:

TABLE XXXIX

PERCENTILE AMOUNTS DONE

								Total Gain
A	100	120	140	160	180	200	200	100
B	100	133	177	177	188	188	200	100
C	100	100	100	120	130	140	150	50
D	100	150	183	200	200	250	300	200
E	100	140	180	200	240	280	300	200
Av.	100	129	156	171	188	212	230	130
A.D.	0	15					56	

From this it could be said that D and E improve most and C least.

Again turning this table into units of time taken and expressed in percentile values of the starting point it becomes:

TABLE XL

PERCENTILE DECREASE IN TIME TAKEN

								Total Improvement Per Cent.
A	100	83	71	62	55	50	50	50
B	100	76	56	56	53	53	50	50
C	100	100	100	83	76	71	66.6	33.3
D	100	66	54	50	50	40	33.3	66.6
E	100	71	55	50	42	36	33.3	66.6
Average	100	79	67	60	55	50	46	
A.D.	0	9.8		9.8			10.8	

As from the preceding table, the conclusion would be that A and B make equal gain, that so do D and E, and that C gains least; but whereas before C's gain was half A's and B's, and one fourth D's and E's, now it looks like one half that of D and E. Again, in each table of percentile values the A.D. tends to increase, and evidently, since in the curves the starting point is a common zero, they inevitably diverge later, and might be interpreted to mean that differences increase by practise.

In general then, this particular use of the method of percentiles must confuse the issue unless each individual's starting point is given, *i. e.*, unless some statement of gross scores is also made.

Working over the original scores given above by both Smythe Johnson's and Amberg's methods, the percentile increase is as follows:

	A	B	C	D	E
Smythe Johnson	23	19	15	38	40
Amberg	32	29	19	53	56

Here the subjects keep the same relative position, though the statements of *how much more* one improved than the other would not be alike in the two cases. E improves most and C least is all that can be said.

Just to put these varying interpretations into strong contrast the following table has been prepared, giving for six ways of expressing the facts very varying answers to the question of relative improvement.

TABLE XLI

IMPROVEMENT OF SEVENTH OVER FIRST PRACTISE PERIOD IN

Individual	Gross Amount Work Units	Gross Time per Work Unit	Percentile Amount Work Units	Percentile Time per Work Unit	By Smythe Johnson	By Amberg
A	5	150	100	50	23	32
B	9	83	100	50	19	29
C	5	50	50	33.3	15	19
D	12	167	200	66.6	38	53
E	10	200	200	66.6	40	56
Av.	8.2	130	130	53.3	27	37.8
Gained most	D	E	D E	D E	E	E
Gained equally	A C	None	D and E A and B	D and E A and B	None	None
Gained least	A C	C	C	C	C	C
Other statements	E gains twice as much as C or A	E gains four times as much as C	E gains twice as much as A and four times as much as C	E gains twice as much as C	E gains between two and three times as much as C	E gains nearly three times as much as C

One more such case will be considered, but, for brevity, instead of the similar four first tables and curves only the first and last scores in gross amount of work done by four subjects in 10 units of time is given, and a set of comparisons worked out as in the table just preceding.

TABLE XLII

IMPROVEMENT OF LAST OVER FIRST PERIOD OF PRACTISE IN

Score First	Score Last	Individual	Gross Amount Work Units	Gross Time per Work Unit	Percentile Amount Work Units	Percentile Time per Work Unit	By Smythe Johnson	By Amberg
20	30	W	10	17	50	66	14.7	17.3
12	20	X	8	33	66	40	19.2	19.6
15	25	Y	10	26	66	40	23	27.7
16	24	Z	8	21	50	33	14.5	17.0
Average			9	24.2	58	45	17.8	20.4
Most gain			W Y	X	X Y	W	Y	Y
Equal gain			W and Y X and Z	None	X and Y W and Z	X and Y	None (W and Z)	None (W and Z)
Least gain			X Z	W	W Z	Z	Z	Z
Other statements			W gains more than Z	W gains less than Z	W gains equally with Z	W gains twice as much as Z	W gains slightly more than Z	W gains slightly more than Z

The conclusion would be that if one wishes to compare one individual with another in rate of improvement, or one individual's performances in two different kinds of tests, any statement based upon a comparison of difference between the last score and the first score will be seriously affected by the kind of units chosen, and may be the more misleading the more definitely comparative they are made. All of these methods alike ignore the actual starting and finishing points which might be useful objective data, and may outrage the sense of fairness by equating units taken from different points of the scale. Thus it seems absurd to call A and C equal because each gains 5 units, since they start and finish at such different points. But to imagine that expressing A's performance as 100 per cent. gain, C's as only 50 per cent. and therefore conclude that A does twice as well as C, may be equally absurd, since it may be no nearer the truth than was the first statement. There is no magic in percentile statements, except it be in blinding people to the actual efficiency of a performance.

Then too, useful information may be obscured by stating merely the amount of *gain* or *loss* whether in gross or percentile statements, information which the full tables would have given and which is of interest; such as, in the first example, that at the start C is much better than E, but after seven periods of practise their performances

are equal, and that A after practise reaches only the point where C started. Also from the second example, W who was best at the start maintains his lead and is best at the finish; X who was poorest at the start was also poorest at the finish. Facts such as these are not brought out by a mere statement of gain, nor by the percentile tables and curves, though they would be by the gross amount tables and curves; yet they are of value in application to everyday tasks where objective norms must hold in speed and accuracy.

At this point examples may well be given of the treatment actually given to practise records—or fatigue. Gilbert[57] argues in favor of the percentile measures thus: "To have expressed the fatigue merely by the difference between the two rates of tapping would not have expressed the truth: *e. g.*, one child who tapped 19 and 15 for the respective periods of 5 seconds lost a great deal more than another who tapped 38 and 34 respectively: each lost 4 taps but the first lost 21 per cent., the second only 11 per cent." His curve shows the average per cent. of loss for each age, which means for eleven-year olds, that children whose records were 30 to 24, 35 to 28, and 25 to 20 were considered equal. Later he says, "The average boy . . . taps 29.4 times in five seconds, the average girl taps 26.9 times, thus tapping 8.5 per cent. slower than boys. The average boy . . . loses 18.1 per cent. by fatigue, the average girl loses 16.6. In other words the boys lose 1.5 per cent. more by fatigue and yet tap 8.5 per cent. faster. This leaves the balance greatly in favor of boys." Elsewhere, however, he does give a table of gross averages.

Wells, in a report read before the New York Branch of the American Psychological Association in 1910 quoted some practise results in two different tests without giving starting points, concluding that as there was 71 per cent. gain in one test and 94 per cent. in the other, there was greater gain in this than in the first. In a published article on practise in free association[58] the curves in that test are plotted on the gross decrease in units of time; but when comparison is made of susceptibility to practise in this test and in two other tests, no gross figures for the others are given at all but only the ratio of the mean of the nineteenth and twentieth days to the mean of the first and second days practise, and the conclusions based on those ratios.

Davis in his studies of cross-education[59] gives no gross gains, only the percentage. The ratio is taken on the basis of the first trial which is called 1; then the result is stated that the left hand gained

[57] *Yale Studies*, **2**, 1894.
[58] *Am. Jour. of Psych.*, **22**, 1911.
[59] *Yale Studies*, **8**, 1900.

more than the right. In earlier work[60] on the same problem he quotes initial and final scores, gross and relative gains, and plots his curves in gross errors.

Woodworth and Thorndike[61] carefully point out that one's interpretation of what equal improvement or indeed proportionate improvement means depends upon what is taken to be the starting point, and they recommend the use of at least two measures of accuracy. They use the gross error, also the ratio of errors after practise to errors before practise, so that improving from 166 to 130 errors or 78 per cent., is considered about equal to improving from 302 to 232 errors, or 77 per cent. Later a statement occurs, "the improvement in—is not equalled in the other functions." Seven years later Thorndike gives this warning:[62] "In estimating individual differences in amount of improvement . . . the ratios listed must not be taken thoughtlessly at their face value. For a person to change from 400 seconds per example to 200 is not necessarily the same amount of improvement as for him or another to change from 200 seconds to 100 seconds. The second is probably an improvement which fewer individuals would be capable of, which the same individual would take longer to attain. . . . To call the two equal as fractions must not lead one to infer any thorough-going equality in the facts which the fractions only partially represent. . . . In fact every measure of improvement by a gross difference or by a ratio must be accompanied by a statement of the initial or final gross actual ability." Such statements are given both in this and in later work,[63] where no conclusions are drawn as to whether one individual improved more or less, especially by how much more or less than another. In presenting a curve which might be representative of the general law of change, whether from the beginning of the test to the end, or between two arbitrarily chosen points each within every individual's compass, it is plotted according to the central tendency of a series of points determined for each individual by the formula

$$\frac{\text{first score—score in question}}{\text{first score—last score}}.$$

But this average or mean curve is characterized as mongrel since changes in the rate of improvement are due "to the action of radically different laws acting on different individuals according to the

[60] *Yale Studies*, **6**, 1898.
[61] *Psych. Rev.*, **8**, 1901.
[62] *Am. Journ. of Psych.*, **19**, 1908.
[63] *Am. Journ. of Psych.*, **21**, 1910.

different physiological changes in them to which the improvement is due."

It would seem then that the answer to the question "How much relative improvement is there, or how much more does one individual improve than another?" can be given only for some arbitrarily chosen definitions of "how much" and "how much more." The nature of the work, the inevitable relativity of the starting points and of the units, and one's preferred method of interpreting statistics will all modify such answer. What must be done is to keep the first factor in mind, to present the second fully, and in more than one way, to be wary and undogmatic as to the third, allowing others to be the same.

There are other questions commonly asked, however, and answered simply from examination of curves plotted according to gross amounts, or somewhat variously by the use of certain formulæ.

For example, it is of great importance in relation to measurements of the relative parts played by heredity and environment in producing the differences between individuals to determine whether, and how far, different amounts of training account for individual differences. The most usual and convenient measurement is of whether and how far equal amounts of practise will reduce individual differences. To make this measurement one might:

1. Examine the average deviations from the average at the first trial, and also after practise, and compare them directly. Then according as one's units of measurement increase in amount or decrease in time or error, so will the deviations in all probability.

2. Use the formula $\frac{\text{A.D.}}{\text{Av.}}$ for both beginning and end, and make comparisons.

3. Use the preferred formula $\frac{\text{A.D.}}{\sqrt{\text{Av.}}}$ and compare.

4. Study the ratio of the range at both the beginning and at the end, by finding in each case the ratio of best to worst, second best to second worst and so on, and comparing each such ratio with the corresponding ratio at the end.

Moreover any of these four methods could be applied not only to the first and last scores, but to averages of the first few and the last few, or the middle, or to each if necessary. Using all four methods on the two examples given, the figures would stand:

TABLE XLIII

FROM EXAMPLE 1

	Gross Amount		Gross Time		Per Cent. Amount			Per Cent. Time		
	First	Last	First	Last	First	Second	Last	First	Second	Last
Average	7	15.2	233	103	100		230	100		46.6
Gross A. D.	2	2.24	60	18.6	0	15	56	0	9.8	10.5
A. D./Av.	29%	14%	25%	18%	0	11%	24%	0	12%	23%
A. D./$\sqrt{\text{Av.}}$	75%	57%	393%	183%	0	132%	369%	0	110%	154%
Worst and Best .	2.00	1.80	.50	.55	0	1.50	2.00	0	.66	.50
Next Worst } Next Best }	1.80	1.20	.55	.83	0	1.16	1.50	0	.85	.66
	or from twice as good to 1.80 times as good		or from half as good to .55 as good		or from 1.50 as good at the second trial to twice as good			or from .66 times as good at the second trial to only half as good		

TABLE XLIV

FROM EXAMPLE 2

	Gross Amount		Gross Time		Per Cent. Amount		Per Cent. Time	
	First	Last	First	Last	First	Last	First	Last
Average	15.7	24.7	65	41	100	139	100	63.3
A. D.	2.2	2.7	9.2	4.5	0	1.9	0	3.3
A. D./Av.	14%	10%	14%	11%	0	13%	0	5%
A. D./$\sqrt{\text{Av.}}$	57%	54%	114%	70%	0	161%	0	42%
Worst and Best ..	1.66	1.50	.60	.66	0	1.16	0	.91
Next Worst } Next Best }	1.06	1.04	.94	.95	0	1.16	0	.91
	or from 1.66 times as good to 1.50 times as good		from .60 as good to .66 as good					

From the tables in gross amounts it would be concluded that individual differences tend to increase with practise; but the terms in which the score is kept, and the method of comparing variations make a great difference in the apparent amount or ratio of that decrease. The last method illustrated needs perhaps a word of caution. In the second column—although the figures increase from .50 to .55 and .60 to .66, this means a decrease in differences of range, as the interpretative readings added for both the first and second columns show. Obviously, in the next two columns by the percentage increase or decrease scoring, individual differences must be shown to increase by practise, since all are made to start equal. The answers to the questions obtained by such methods are then necessarily absurd.

Therefore in using any of these four methods to examine the variability one should again: (1) beware of being misled by the kind

of units used, both at the chosen starting point and at any point in the practise series: (2) prefer gross to percentile measures of the ability in question: (3) remember that only general tendencies are given, not specific comparisons.

Even the fourth method would not make comparisons always between the same pairs of individuals unless they happened to retain their relative position all through the series, since it is engaged in studying the range whoever may be at or near the extremes. But this very point of individual comparisons is also of interest—whether the one who is best at the start is also best after practise even though the curve may have a less sudden slant than that of the worst at the start, and whether those who start with a poor record will still be poor, or the poorest at the end. The fourth method could be modified to answer that, but there are at least two common procedures. One is to compare the position at the start with the total gross gain or percentile gain or both; the other is to rank all individuals at their first trial and at their last trial and compare the rankings.

By the former method, applied to example 1, between ability at the start and gross gain there is correlation of —.32; between ability at the start and percentile gain a correlation of —.55, from which the inference would be that those who start well gain less than those who start poorly.

By the latter method (used by Wimms[64] in his work with schoolboys in various mental tests) correlating by the "foot-rule" method, R = .75.

Even this ranking method has been variously applied. Wimms, for instance, also tabulates the percentage increase of each of his subjects from the first to the last series of tests and ranks his subjects accordingly. He then finds that the two ways of ranking, this, and by numerical difference of absolute achievement in the last series, do not agree.

Oehrn,[65] whom Wimms quotes, after stating that practise has two effects, that of shortening the time for successive groups of trials, and that of reducing each subject's variability in series of such groups, ranked his subjects first in decrease in gross time taken, also in percentage of reduction of variability, and found that the two ways of ranking were not proportional. His correlations are based on the ranking for the time taken. In his work too he introduces another point as the basis of reckoning for the "work-curve," namely the maximum performance of any individual, which he says is a better standard than the starting-point because more constant

[64] *Brit. Journ. of Psych.*, **2**, 1907.

[65] *Psych. Arbeiten*, **1**, 1896.

for each individual. This is rather a novel procedure, which though it may have suited his conditions—continuous mental work for two hours measured every quarter of an hour—would not suit work like Bair's or Bryan and Harter's where the maximum performance was emphatically not a constant.

In general, this ranking method tells precisely what a direct inspection of individual curves would do; but since with large groups it would be inconvenient and confusing to plot all the curves, tables of ranks would be likely to give direct information about relative improvement. If the question were "Are those who are best at the start also best at the finish?" then ranks in initial and final tests would be needed. If the question were "Do those who are best at first improve most or those who are poorest?" then ranks by the initial record and total increase would be needed. The absolute gain would be the more objective record perhaps, but here, at least, so long as gross measures are available, a percentile or proportional gain would not be misleading, and would often give just the practical information required.

Now this tedious elaboration has been based on simple and suppositious records, solely to bring out possible discrepancies in results and conclusions according to the use of one method rather than another. Actual published results could be worked out in the same way and contrasts drawn. That would, however, be beyond the scope of the present investigation.

That the practise or rather the "work-curve" may be complicated beyond easy and rapid inspection, Kraepelin has endeavored to show[66] when he takes the record of one subject in continuous work for two hours and at great length analyzes and plots curves for at least seven factors: practise, fatigue, adaptation (or warming-up period), inclination (or attitude towards work), initial and final spurts, the desire to improve, and recovery by rest. He points out, too, the difference between morning and evening workers, and the effects of a recent meal or period of sleep.

Who would study individual differences as revealed in or affected by practise has no easy task.

2. Results from a Special Series of Tests

So far in this study, the statistics of practise with the short or long term groups have been confined to the starting point, average and finishing point in gross amount for each group, with no comparison of individuals. Too few subjects made up the long-term group to make any extended comparisons worth while, and the larger

[66] *Phil. Studien*, **19**, 1902.

group made too few trials with most tests to do more than indicate the trend of individual curves at the beginning of practise.

Also, the results have been stated as if a typical curve for a test or a group of tests could be determined. But it is a question whether individuals will not differ so much in their improvement with any test as to make the average or mean curve unreliable, or rather representative of nothing. It is also a question whether an individual's improvement in one test will not so parallel his improvement in another as to make his curve typical of him rather than of the kind of work. Or again, a "motor minded" individual might show a different rate of practise in a motor test from one who is an abstract thinker, and different also from his own improvement in another field. In other words is "the practise curve" that of (1) the kind of work, or (2) of the general abilities of an individual, or (3) of special abilities of individuals?

In the hope of getting a little light on this problem, a further set of tests was undertaken with a larger group of subjects, a long period of practise, and with five tests of presumably very different functions.

Supposing tests could be selected with which the subjects had had no previous experience, then if all show slants and plateaus at about the same level of practise judged by time or amount, the curve would be typical of the kind of work. If there is greater resemblance between all curves from one individual than between one individual and other, then the curve is typical of the kind of person rather than of the kind of work. If any one subject's curves in, say two motor, or two mental tests, resembled each other and were unlike the mean curve, but in tests of some other function were like some other individual's curves, then the curve is typical of specialized abilities in individuals. Lastly, if the mean curve for one individual in several tests is indistinguishable from the mean curve of several individuals in one test there would be no evidence one way or the other except that practise must produce the same results in people whatever the work, and so must reduce differences between people.

In order to discover which of the above conditions would prevail, a group of subjects was put through a period of practise for twenty days, excluding Sundays, in November and December of 1909.

The subjects, nine in number (the tenth did not continue sufficiently long for any use to be made of her records) were all women selected from among Teachers College students on the basis of their needing financial help in working through college and so responding to an appeal for subjects. From the group those were used who

could give from one and one half hours a day at the beginning to whatever time the tests took at the end of the period of practise, always at the same time of day. Four distinctly different nationalities were represented, and five different departments in the college. One was constitutionally delicate, two others showed signs of strain and worry, the other six were in good health. One was over forty, one over thirty, the others under twenty-five. Their college standing for the year 1909–10 was also examined, and they themselves were carefully observed for general temperament as revealed during the practise of one test. These facts are tabulated below:

Subject	Nationality	Department	Health	Relative Age	College Standing
C.	American	Mathematics	Delicate	Young	Good
E.	American	Eng. & Dom. Sci.	Tired	Over 30	Very good
Go.	Russian Jewess (German)		Good	Young	Variable
H.	American	English	Good	Young	Poor
Jb.	German	Domestic Art	Good	Young	Fair
Nb.	American	English	Good	Young	Good
P.	American	English	Good	Young	Fair
Sch.	German	German	Good	Over 40	Good
Sa.	Jewess	Physical education	Strained	Young	Fair to good

The tests selected were five in number: one for accuracy and speed in movement, one for sensory discrimination, one for discrimination plus movements, one cancellation or perception test, one purely mental test. The tests were explained orally to the subjects and demonstrated, after which a manuscript book was given to each with the directions for each test written out, and spaces prepared for the required entries. The subjects were asked to select whatever time of day was most convenient for them, and to work always at that time through the whole number of days that the tests lasted. Four of the tests were thus practised independently and always in the same order; but for the discrimination of lifted weights, which test needs of course an observer, each subject came at an appointed hour.

For the first test the *curved maze* already described (see page 87) was used. The directions were as follows:

"1. Place the maze so that the words *begin here* are at the left-hand bottom corner. Do not turn the paper about during the test. See that you have a sharp pencil.

"2. Note the time when you begin: (wait until the second hand of your watch is at 60).

"3. Draw a line between the two lines of the maze without touching either, working as fast as you can.

"4. Note the exact time at which you finish, entering both times in the proper columns opposite.

"5. Write your name on the blank, also the number of the experiment."

The spaces ruled for entry were headed:

Date	Time of Day	Physical Condition	Time at Start	Time at Finish

In this third column they were directed to grade their felt condition from A, excellent, to D, miserable. Thus a check of health and weather could be applied to each subject's performances.

The "purely mental" test consisted of three sums in mental multiplication of a three-place number by a three-place number. The directions were:

"1. Beginning at the middle of this book you will find, under *day 1, 2,* etc., three sums to be multiplied, each 3 figures by 3 figures.

"2. Cover up all but the one to be worked; take note of the time.

"3. Multiply it mentally. Do not write anything at all till you get the final answer, then write that down.

"4. Record for each sum in the appropriate column the time at the beginning and the time at the end. Do not rest more than three minutes between examples."

This wording might have been still more explicit, but the subjects understood that "take note of the time" meant to write it down, and also that the recording was to be for each sum, not after all three were finished. The spaces for entry were headed:

	First Sum		Second Sum		Third Sum	
	Time at Start	Finish	Time at Start	Finish	Time at Start	Finish
Day 1.						
Day 2.						
Etc.						

For the sorting test, Dennison's colored cardboard counters 1½ inches in diameter, $\frac{1}{20}$ of an inch thick were used, and for the "box," the 5-cent size ice-cream carton. The directions were:

"1. In the little bag are 50 counters all of one color; in the box are 50 counters of five different colors. Empty the varied ones into some convenient place, and empty the bagful into the box.

"2. Distribute the 50 from the box at random into five piles. In doing this use one hand only, and pick up only one at a time. Work as rapidly as possible. Do this twice, just for practise in manipulating the counters. Return them to the bag.

"3. Shuffle the 50 mixed colors well, and put them into the box. Time yourself as in the other tests, and sort the 50 into five heaps *according to color,* using the same care in handling as before. Record the time at the finish.

"4. On the 1st, 10th, and 20th days, record also the time before and after *one* distribution of the 50 all of one color."

Spaces were prepared for the entries of time at start and finish each day as before, also for the three additional entries.

For the cancellation test, two copies of each of two back numbers of the *Journal of Philosophy, Psychology, and Scientific Methods* were provided for each subject. From these certain pages were selected which were fairly evenly filled with print, in the hope of getting about the same number of *a's* for each experiment, also about the same number of lines for the eye to traverse. Previous work with this test had shown how soon a blank is memorized, so that it seemed advisable to use more ordinarily available reading matter. Pages of a foreign text would have been still preferable.

The directions were:

"1. Find the pages for the day: be ready to turn over quickly. Note the time.

"2. Mark, on the pages designated every small print *a* you see, going line by line over the two pages. To underline is the quickest method.

"3. Note the time at start and finish as before."

The spaces for entry were headed as before, besides indicating for each day exactly which pages were to be used. A second trial with the same page was made only four times, and then it came at least ten days later than the first trial, so that there was practically no memory of the location of the *a's*. The average total number of *a's* for the daily task was found to be 338, but unfortunately with a large range of from 268 to 410, which complicated the latter calculations very much.

For the *lifted weights test* thirty weights ranging from 40 to 130 grams were prepared. These were unpainted wooden cylindrical boxes containing lead or small shot to make up the required weight. Six of these were used as standards of comparison, a 40, 55, 75, 90, 110, and 130 box, so labelled, and kept apart by themselves to the side of the twenty-four test boxes. Of these, there were nineteen different weights ranging by differences of 5 grams from 40 to 130 grams, and also six duplicates, one each of the 45, 60, 75, 90, 105, and 120 gram weights. It will be noticed that of these duplicates two are identical with two of the standards. By using six standards scattered through the range, and by using steps of five grams it was hoped to make the test easier and therefore likely to be completed more rapidly than if merely one of the extremes had been used as

the sole standard or if very fine discriminations had been necessary (see Thompson's work[67]).

The twenty-four test weights were arranged in three rows of eight, and daily rearranged in a different order with care to avoid strong contrast effects and consequent probable illusions. Secret marks on the side nearest the observer permitted immediate and rapid checking up of the judgments made. For the first two days preliminary experience was allowed in hefting the six standard weights and one or two test weights. Thereafter the subjects began immediately upon the test.

The first box in the nearest row was hefted with the fingers of the right hand, then one of the standards, whichever would be selected as probably the nearest, then the judgment was generally made in terms of grams. However the subjects were free to try another standard if the first was presumably not near the testbox in weight and then to heft the testbox again. In this way emphasis and help were given to making correct judgments. No fixed speed was insisted on, but a check was kept on the total time taken daily for the whole set of twenty-four judgments. Only on three occasions were subjects hurried up, and then when they had exceeded 25 seconds in arriving at a judgment. Otherwise the aim was to leave the subjects as free as possible.

Each subject came 16 times for this test, though as all did not begin on the same day, any particular arrangement of the boxes would not fall on say the fifth trial for everybody. After a certain date too, each subject after having made a judgment was told what the real weight was, in the hope of facilitating practise by this means. Again, this additional means of training did not begin at the same point in the series of 16 tests for each subject. In the curves this point is indicated for each individual by a small cross.

In working up the results, judgments for weights below 60 g. and over 105 g. were not used, in order to avoid the influence of the "end error." The curves then are plotted from the average error in 14 judgments of 10 different weights from the middle of the series, 4 of which were duplicates and 2 of those duplicates identical with 2 of the standards. This leaves a total of 2,016 judgments instead of 3,024.

The method of scoring was to enter immediately the errors in grams, plus or minus. After the date on which the subjects were told the real weights, the last 12 judgments of the 24 were recorded in ink instead of pencil. In this way could be found (1) the average error with each weight for each subject, (2) the constant error for

[67] "The Mental Traits of Sex."

TABLE XLV

AVERAGE AND CONSTANT ERRORS. WEIGHTS TEST

Day	Subject S.		Nb.		Go.		P.		C.		Sch.		J.		H.		E.	
1	+5.7	10.0	+ 2.5	13.2	−11.8	19.6	− 3.2	8.9	−4.6	6.8	+ 2.5	9.6	+5.3	10.3	+5.3	9.6	−6.8	12.1
2	0	6.4	+11.1	11.8	0	12.1	+10.3	10.3	+1.1	10.0	+ 8.2	11.1	+5.7	10.7	+9.2	10.4	+2.1	7.1
3	− .3	5.7	+ 3.9	6.8	− 2.5	10.3	+ 2.9	6.4	−2.5	8.2	+ 8.6	10.7	+2.1	5.7	+3.9	9.6	−1.4	5.0
4	−1.4	3.6	+ 5.0	7.9	− 5.3	12.8	+ 1.8	6.1	− .3	8.2	+13.5	15.0	+3.5	7.1	+3.2	7.5	+2.8	11.4
5	*+2.1*	*3.6*	+ 2.8	5.7	+ 8.6	11.8	+ 2.1	9.3	+2.1	10.7	+ 5.7	8.6	−2.5	9.3	+3.6	9.3	0	9.2
6	+2.5	5.0	+ 4.3	6.4	+14.6	14.6	+ 8.9	9.6	+7.9	9.3	+ 9.3	11.3	− .3	10.7	+1.6	10.4	+5.3	8.5
7	+1.4	4.3	+ 2.1	5.0	+10.7	15.7	+ 2.1	4.3	+3.2	7.7	+ 4.3	5.7	+ *.3*	*6.8*	+6.4	8.9	+7.5	13.9
8	+2.5	3.9	+ 3.9	4.6	+10.7	14.3	+ 2.1	9.6	+3.6	6.1	+ *2.5*	*7.5*	−3.2	3.9	+3.6	7.1	+5.3	11.0
9	+2.5	5.3	− *1.0*	*6.1*	+ *5.7*	*10.3*	+ *6.4*	*8.9*	*+2.9*	*8.2*	+ .3	5.0	−1.1	5.3	*+6.4*	*7.1*	*+6.4*	*10.0*
10	+3.6	5.0	+ 3.9	7.5	+ 1.4	10.7	+ 4.6	4.6	+ .3	7.5	+ .7	7.1	+3.9	5.7	+7.8	7.5	−1.8	5.3
11	+2.1	4.3	+ 5.0	6.4	− 1.8	11.1	+ 2.9	6.1	+7.5	8.9	+ 2.8	5.0	+2.1	6.4	+8.9	8.9	+1.1	5.7
12	+3.2	4.6	+ 6.1	8.2	+ 3.2	8.2	+ 7.1	8.2	+5.0	7.9	+ 2.8	5.3	+4.3	5.7	+5.7	6.4	+3.5	7.8
13	+3.2	3.2	+ 2.8	5.7	+ 1.8	6.1	+ .3	2.5	0	3.9	+ 4.6	7.9	+3.2	7.5	+2.8	4.3	+6.1	9.3
14	+3.9	6.1	+ 2.8	5.7	+ .7	7.8	+ 4.6	6.8	+5.3	6.1	+ 3.9	9.6	+ .7	5.7	+3.6	4.3	− .7	4.3
15	+4.3	5.7	+ 4.3	5.0	+ 3.2	6.1	+ 1.1	2.5	+5.3	5.3	+ 2.1	7.1	+4.3	5.7	+2.5	4.6	−1.4	5.7
16	+3.2	4.6	+ 2.8	5.0	+ 1.1	10.3	+ 1.8	8.1	+2.1	5.7	+ 1.4	4.3	+5.0	6.1	+1.8	5.4	+2.5	7.1

each weight, (3) the average error daily, (4) the constant error daily, (5) the improvement daily during the test.

Below is given the average error for each weight through the whole period of practise:

40	45	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125
2.1	4.4	5.7	6.1	6.3	7.8	7.2	7.1	9.1	7.8	7.6	8.3	8.2	7.9	7.3	5.5	5.3	3.9

From this the influence of the "end error" is clearly visible, though not so far into the series as it had been expected. The weights of 75 and 90 grams do seem to show the benefit of both their identity with the standards and the double practise they received; the 60 grams perhaps shows the double practice benefit, but the same can not be said of the other weight, the 105.

Table XLV gives for each subject for each day the average error and the constant error for the set of 14 judgments. The scores in italics show the first day on which additional help was given by being told the real weight.

In general this shows a slow reduction in the average error for each subject, a tendency to a positive constant error, a disturbance in the constant error on the day of the change in method, and that the greatest fluctuations occurred between the first and second trials.

Eight other individuals also took this test in this form once each. For these control cases are here given also the constant error and the average error for the set of 14 judgments.

TABLE XLVI

Ind.	Const. Error	Av. Error	Rank for Accuracy. 1 = Least Error	Rank for Time. 1 = Least Time
1	+4.3	6.4	4	2
2	+ .7	17.2	8	8
3	+2.1	5.0	1	6
4	—1.4	10.7	7	1
5	— .3	9.7	6	5
6	—1.0	6.1	3	7
7	—1.8	5.4	2	4
8	+9.6	9.6	5	3
	Average.....	8.8		

Compared with the first trial by the practising group, 11.1, the average record for these is somewhat better. The curves as plotted for each of the group of nine subjects from their daily average error are shown in Fig. 2. The dotted line shows the most probable "smoothed" curve. The two individuals most unlike are Go. and Sa. The latter had the benefit of knowledge of the correct weight longer than did the others; she was also the slowest of the nine. Go. gave

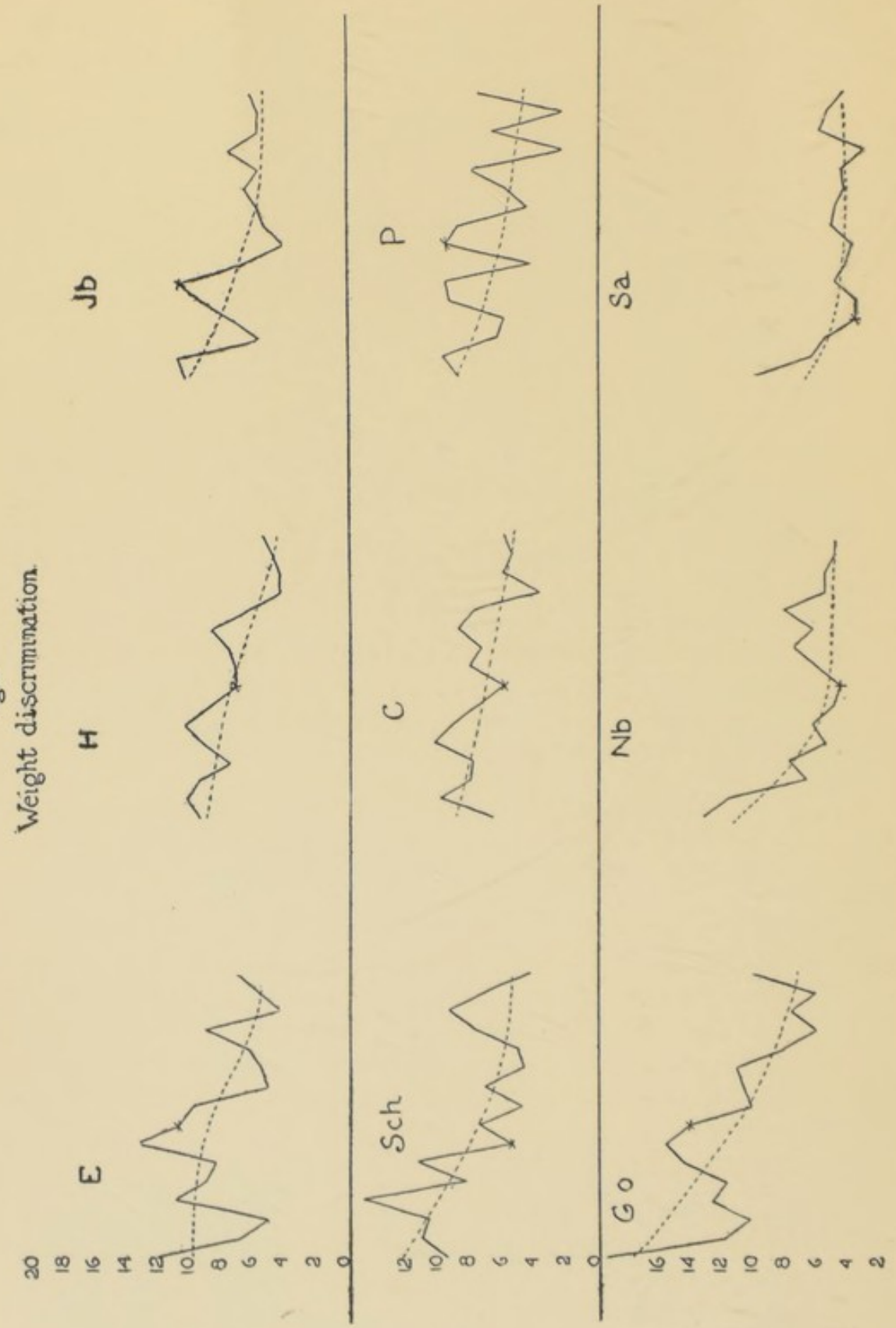


Fig 2.
Weight discrimination.
E
H
Jb
Sch
C
P
Go
Nb
Sa
20
18
16
14
12
10
8
6
4
2
0

the impression of being very careless and indifferent, she took about half the time that Sa. did. Nb., who took about three sevenths the time Sa. did and was the quickest, has a curve more like Sa.'s than has any one else. Taking the average of the first two trials and of the last two (gross score), the gross gain, percentile gain, the time taken on the average, and ranking the nine subjects by each of these scores, we get:

TABLE XLVII

	Rank at Start. (1 = Least Error)	Rank at Finish. (1 = Least Error)	Rank for Av. of Total Series (1 = Least Error)	Gross Gain. (1 = Most Gain)	Percentile Gain. (1 = Most Gain)	Time Taken. (1 = Least Time)
E.	3.5	8	8	7	9	7
H.	5	1.5	6	3	2	4.5
Jb.	7	7	3.5	4.5	6	2.5
Sch.	6	6	7	4.5	5	8
C.	2	5	5	9	8	6
P.	3.5	4	2	6	4	4.5
Go.	9	9	9	1	3	2.5
Nb.	8	1.5	3.5	2	1	1
Sa.	1	3	1	8	7	9

from which the correlations by the method of rank differences are as follows:

Position at start and at finish	$R = +.27$
Position at start and average in the whole series	$+.45$
Position at start and gross gain	$-.98$
Position at start and percentile gain	$-.65$
Average in whole series and time taken	$-.04$

This means that, with these subjects at least, their performance at the first two days' trial was relatively more like their average performance than it was like their performance during the last two days. Those who were poorer at the start made a greater relative gain and a much greater gross gain than those who were better at the start. Within the range of accuracy attained there was practically no relationship to the speed of judgments.

For the control cases also the correlation of accuracy and speed in this was — .07, very near the figure for the practising group, and meaning again practically no relationship.

To notice the improvement if any during the daily test the average errors of the first twelve and the last twelve judgments of each subject were compared. The twelve were of course carefully distributed over the whole range of weights. The errors are as follows:

	First 12	Last 12
E.	6.7	6.8
H.	5.1	5.5
Jb.	5.4	5.9
Sch.	5.5	5.1
C.	5.1	5.3
P.	5.1	5.4
Go.	8.2	6.5
Nb.	5.4	5.7
Sa.	4.1	5.0
Average	5.6	5.7

There is no "warming up" effect discoverable from the first half to the second half of the test daily. On the whole there is scarcely any difference, though for some subjects there is a decided increase in error, which in Sa.'s case may be due to fatigue, since she was the slowest.

The scoring of the *a's* test was not so easy, because of unequal numbers of *a's* in the daily tasks of two pages each. Instead of retaining the gross time taken to cover two pages, it seemed fairer to make the following reduction: find the time that would have been required (proportionately) to cover a space including 250 *a's* with the same accuracy as was actually shown for the whole two pages, *i. e.*, with the same proportion of errors and omissions. This reduction is accomplished by use of the formula,

$$\frac{\text{time taken}}{\text{number marked}} \times 250.$$

Thus, the score for a subject who in 420 seconds marked 286 *a's* is

$$\frac{420}{286} \times 250 = 367.$$

This score is, essentially, the time for covering a given space, and therefore grows smaller with increase in efficiency.

In the following table are given the daily scores for each individual, and also the total number of *a's* in the day's task. The curves as plotted from these scores are shown in Fig. 3. The greatest difference is from the first to the second day's trial.

There are several curves fairly similar, Jb.'s and P.'s, for instance, also H.'s and Go.'s, perhaps E.'s and Nb.'s. When smoothed out, there are seven very similar, namely those of all except H. and Go. The two most unlike are Go.'s and C.'s, the former irregular, showing a poor record at the start and a rapid improvement, the latter very smooth, with a good record at the start and gradual but steady improvement. In percentile improvement the two were nearly equal.

Fig. 3.
Marking a's

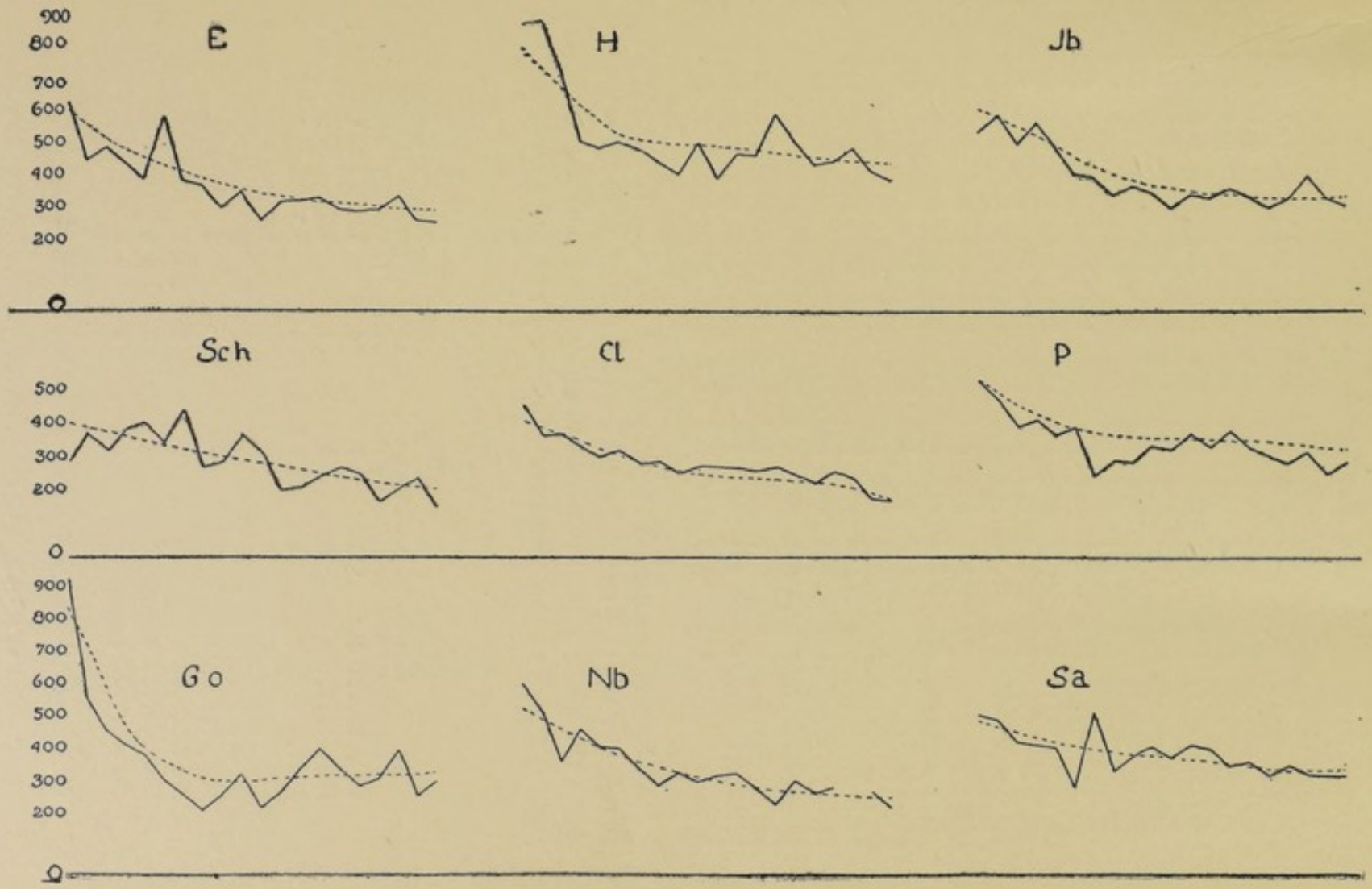


TABLE XLVIII

SCORE IN A'S TEST

Day	a's Possible	E.	H.	Jb.	Sch.	C.	P.	Go.	Nb.	Sa.	Av.
1	299	642	882	550	298	463	542	933	611	508	603
2	314	461	892	596	374	367	480	558	517	498	527
3	299	506	737	514	330	371	398	457	372	428	457
4	285	447	508	581	392	341	418	425	476	426	446
5	345	402	502	497	413	314	366	390	418	414	413
6	374	595	520	418	352	326	391	312	416	298	403
7	355	395	496	408	442	287	257	273	348	519	381
8	368	388	458	347	282	296	302	219	299	362	328
9	365	307	421	381	293	258	293	273	342	386	328
10	318	365	517	360	373	285	341	334	310	417	367
11	409	270	409	314	323	285	329	232	324	382	318
12	333	331	484	347	209	280	383	281	335	420	341
13	327	334	481	340	210	266	343	349	285	412	335
14	268	345	611	365	251	278	390	413	242	356	361
15	315	306	520	340	277	252	343	350	315	365	341
16	360	299	453	315	262	229	312	303	275	327	308
17	334	304	465	325	170	267	297	326	294	360	312
18	313	349	493	413	209	248	334	402	?	336	348
19	410	279	431	347	246	179	262	271	279	327	291
20	409	274	400	319	146	173	304	307	236	333	277
Average	338	380	534	354	293	288	354	370	352	394	

As before, comparing the average of the first two days with the average of the last two days, taking also the average for the whole 20 days, the gross gain, the percentile gain, and ranking the nine subjects for each of these and also for speed and for accuracy, we get:

TABLE XLIX

	Position at Start 1=Shortest Time	At Finish 1=Shortest Time	Average Position 1=Shortest Time	Gross Gain 1=Most	Per Cent. Gain 1=Most	Speed 1=Least Time	Accuracy 1=Fewest Errors
E.	5	4	7	4	5	1	7
H.	9	9	9	1	4	9	4
Jb.	7	8	4.5	5	7.5	7	9
Sch.	1	2	2	9	7.5	3	6
C.	2	1	1	6	2	2	5
P.	4	5	4.5	7	6	6	1
Go.	8	6	6	2	1	4	8
Nb.	6	3	3	3	3	5	3
Sa.	3	7	8	8	9	8	2

The correlations by the method of rank differences are:

Position at start and at finish..... $R = .72$
Position at start and average position..... .58
Position at start and gross gain ..... —.90
Position at start and percentile gain..... —.38
Speed and accuracy..... —.37

Here the subjects kept their relative positions through the test fairly well. Those who were poorest at the start made a greater relative gain than those who were better, and had almost a guarantee that they would make a greater gross gain. The quicker ones are rather less accurate than the slower ones.

In the *sorting test* one subject so misunderstood directions that all her records had to be discarded; another so confused her first nine entries that they too could not be used; a third showed a carelessness in entering whole minutes rather than seconds. Only seven and a half complete records were therefore available of which one is not so reliable as are the others.

The following table gives the daily scores, from which the curves as plotted are shown in Fig. 4. The scores on the three occasions on which the "control time for movements only" was recorded are indicated on each curve by a cross. The missing curves are suggested by a dotted line.

TABLE L

SECONDS REQUIRED TO SORT 60 COUNTERS

Day	E.	Jb.	Sch.	C.	P.	Go.	Nb.	Sa.	Average
1	240	90		150	120	240	150	145	162
2	120	90		60	60	180	90	85	98
3	60	100		75	89	180	60	90	93
4	60	85		55	90	240	60	80	96
5	60	170		56	90	240	90	72	111
6	90	90		65	95	240	90	68	105
7	90	80		55	65	120	90	114	88
8	90	70		50	70	120	72	89	80
9	60	110		50	50	120	72	75	77
10	120	105	120	50	55	150	54	77	91
11	60	80	60	56	60	120	66	90	74
12	60	60	100	53	68	120	66	95	78
13	90	60	85	50	60	60	60	69	67
14	60	110	60	40	50	60	85	100	71
15	45	50	90	45	60	60	60	69	80
16	60	50	115	45	55	60	50	60	62
17	60	108	60	45	60	60	60	62	64
18	60	60	90	45	61	60	60	68	63
19	60	60	60	50	60	60	65	66	60
20	60	60	90	43	50	60	50	65	60
Average	80	84	85	57	68	127	72.5	82	

The curves most alike are those of P. and Nb., though when smoothed out those of Sa. and C. are also similar. Those most unlike are Go.—irregular and rapidly improving—and C., very regular with almost all the improvement at the beginning. Since Go.'s scores were so poorly kept, a better instance of dissimilarity might be C.'s

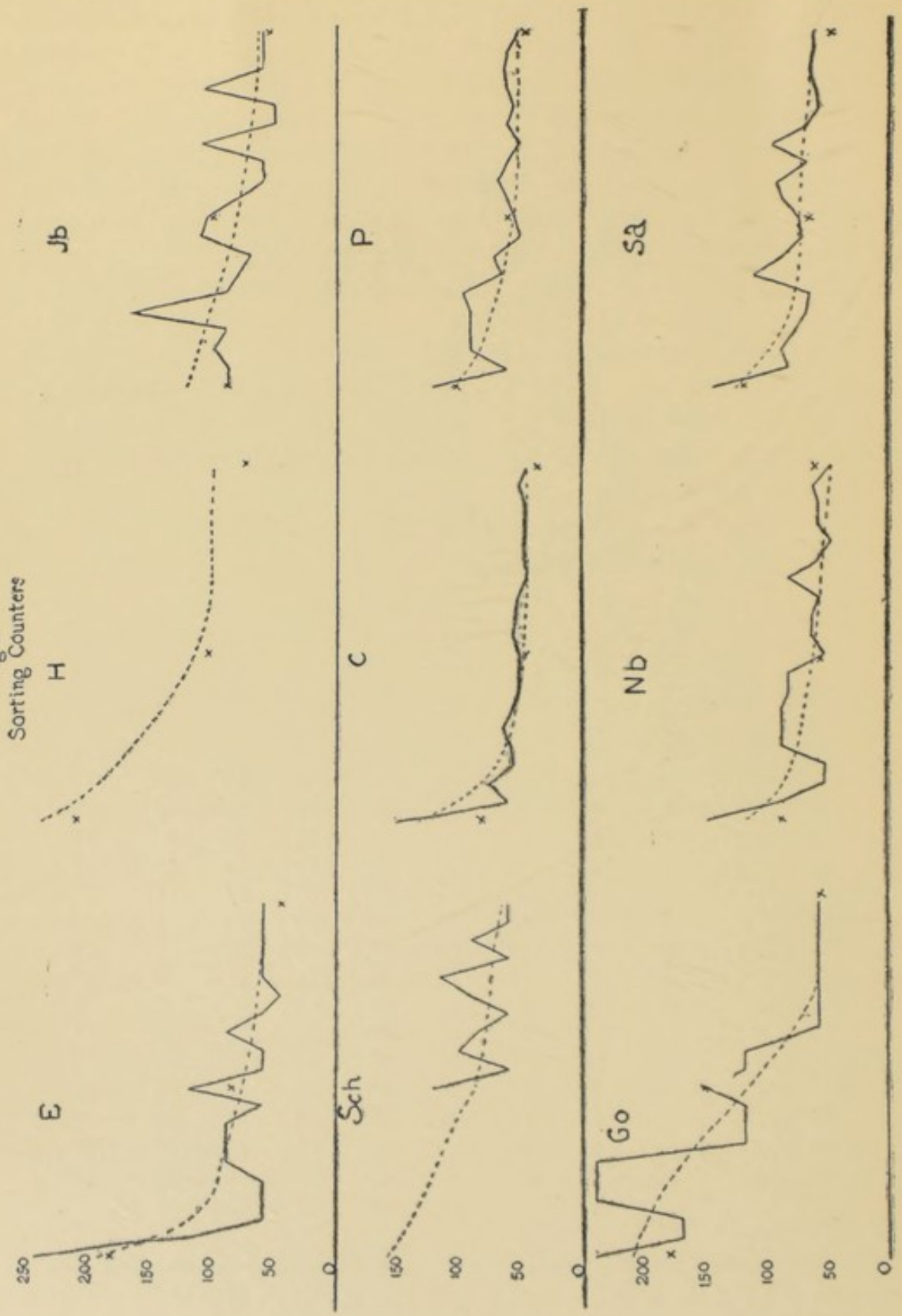


Fig.4
Sorting Counters
E
H
Jb
Sch
C
P
Go
Nb
Sa
250
200
150
100
50
0

curve and Jb.'s, the latter showing great irregularity from day to day and the reverse of improvement near the beginning.

Below are the rankings of the subjects according to position at the start (average of two days), position at the finish, average position, gross gain and percentile gain.

TABLE LI

	Position at Start 1=Least Time	At Finish 1=Least Time	Average 1=Least Time	Gross Gain 1=Most Gain	Per Cent. Gain 1=Most Gain
E.	6	4.5	4	2	2
Jb.	1.5	4.5	6	7	7
Sch.	—	8	(7)	?	?
C.	3	1	1	4	3
P.	1.5	2	2	6	6
Go.	7	4.5	8	1	1
Nb.	5	4.5	3	3	4
Sa.	4	7	5	5	5

From these the correlations by the method of rank differences are:

Position at start and at finish	$R = .56$
Position at start and average position	.58
Position at start and gross gain	−.92
Position at start and percentile gain	−.86

Here there was more change in the relative position through the test than in the marking *a's*. It should be noted, however, that the "positions" were very close together at the end since nearly all got down to about 60 seconds or slightly less in handling the 60 counters.

Again therefore, since all finish nearly alike, those who were poorest at the beginning made the greatest relative and gross gain.

In the *mental multiplication* tests only digits from 3 to 8 were used in the multiplicand, and from 2 to 7 in the multiplier. In arranging examples care was taken to have no two consecutive figures alike in both multiplicand and multiplier—to minimize unnecessary confusion. The subjects all dreaded this test at the outset, but after two days' work with it they gained confidence in their ability. No suggestion was given any of them as to using or discarding visual or auditory imagery, nor as to devices for lessening the number of figures to be remembered. But they were asked to note any change in attitude or method that helped or hindered them. The following notes are interesting.

E. after the second day decided that a pause between examples was not worth while. For a time she visualized a series of dots as a help in placing partial products.

H. found it better to do her adding as she went along rather than to keep one partial product in mind while getting another.

Jb. discarded visualizing as it was a hindrance. She tried saying the partial products aloud for awhile, finally took to adding two partial products before finding the third.

P. also hit upon this method as early as the third day and kept to it thereafter.

C. occasionally adopted a device, such as, with a multiplier like 625, dividing by 4 instead of multiplying by 25. This was seldom possible however. Occasionally she noted that the answer seemed to come automatically, in one process without consciously thinking through the steps. "It opened out before me." C. was specializing in the mathematics department, so was probably better prepared with devices and automatic calculations than any of the others.

In scoring this test, errors were penalized by adding on .2 of the time taken for 1 error, .3 for 2 or 3 errors, .4 for 4 or 5 errors, and .5 for 6 errors in the final answer. As it happens, subjects who are usually accurate seem doubly penalized by this, since with them the consciousness or suspicion of error lengthens their time in any case, whereas with the habitually inaccurate an error more or less made no appreciable difference in the time taken.

Records for each of the 60 examples were kept to see if any particular one was much more difficult or easy than the rest; but both good and bad scores were made with almost every example, and none could be singled out as specially difficult or easy.

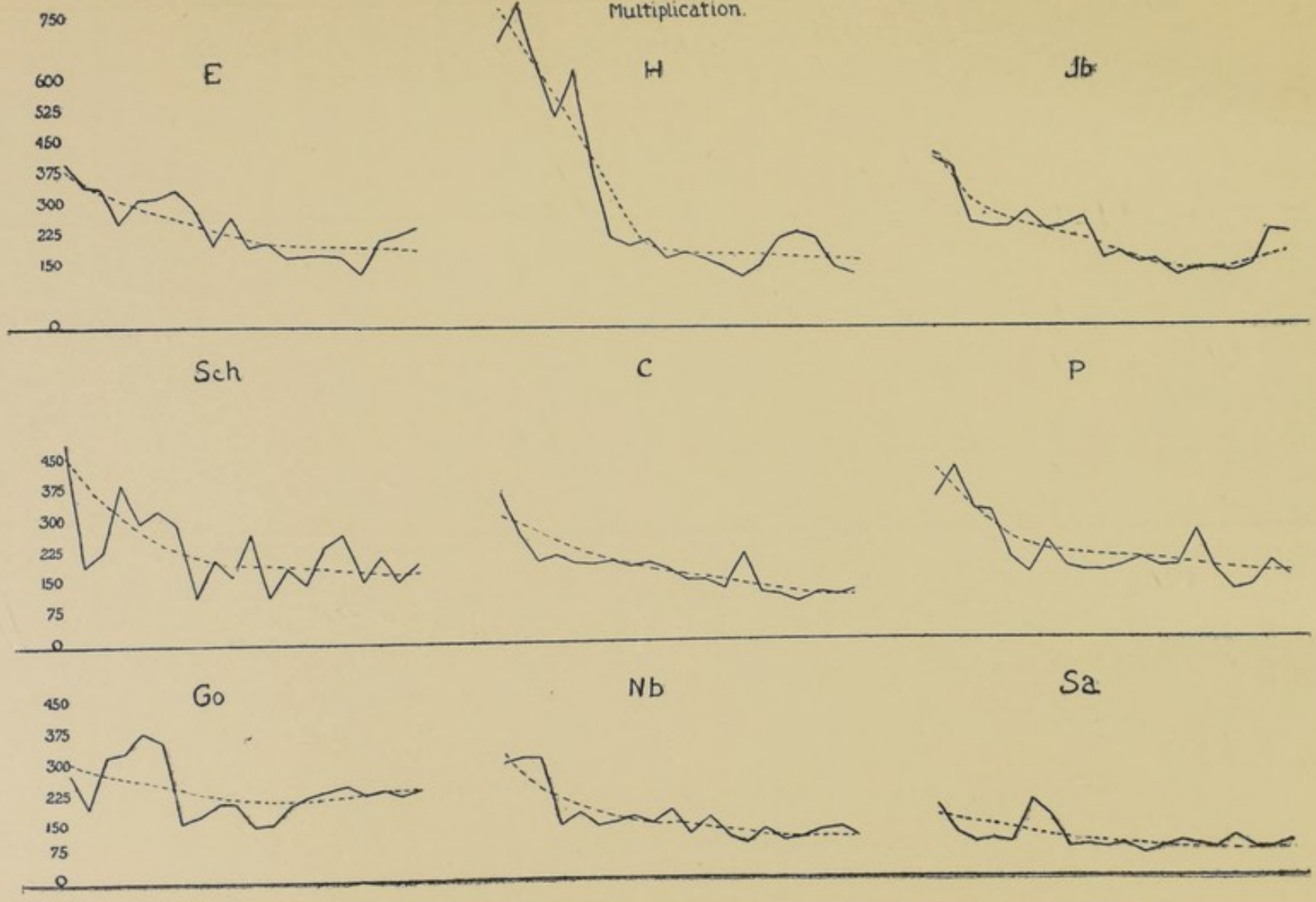
In the table that follows the daily average score for each subject is given, that is, the average score on three examples for 20 days. The curves as plotted from them are in Fig. 5.

As each point on the curve represents an average rather than a single trial, the curves may be considered partly smoothed already. The two most regular and most alike are those of C. and Sa.; the most irregular is that of Sch.; the most unlike any other is that of H., though after the sixth day when her scores are within the range of those of the other subjects, her curve is more regular, and not unlike E.'s or Jb.'s.

The curves representing separately the factors of speed and accuracy are shown by a continuous and a dashed line respectively in Fig. 6.

From this it will be seen that there is very little if any improvement in accuracy, but a good deal in speed. Also, of the most accurate subjects, H. is the slowest, Nb., C., and Sa. are the quickest. Also that there is more individual difference revealed in speed than in accuracy, judging by the amount and regularity of improvement in each.

Fig.5.
Multiplication.



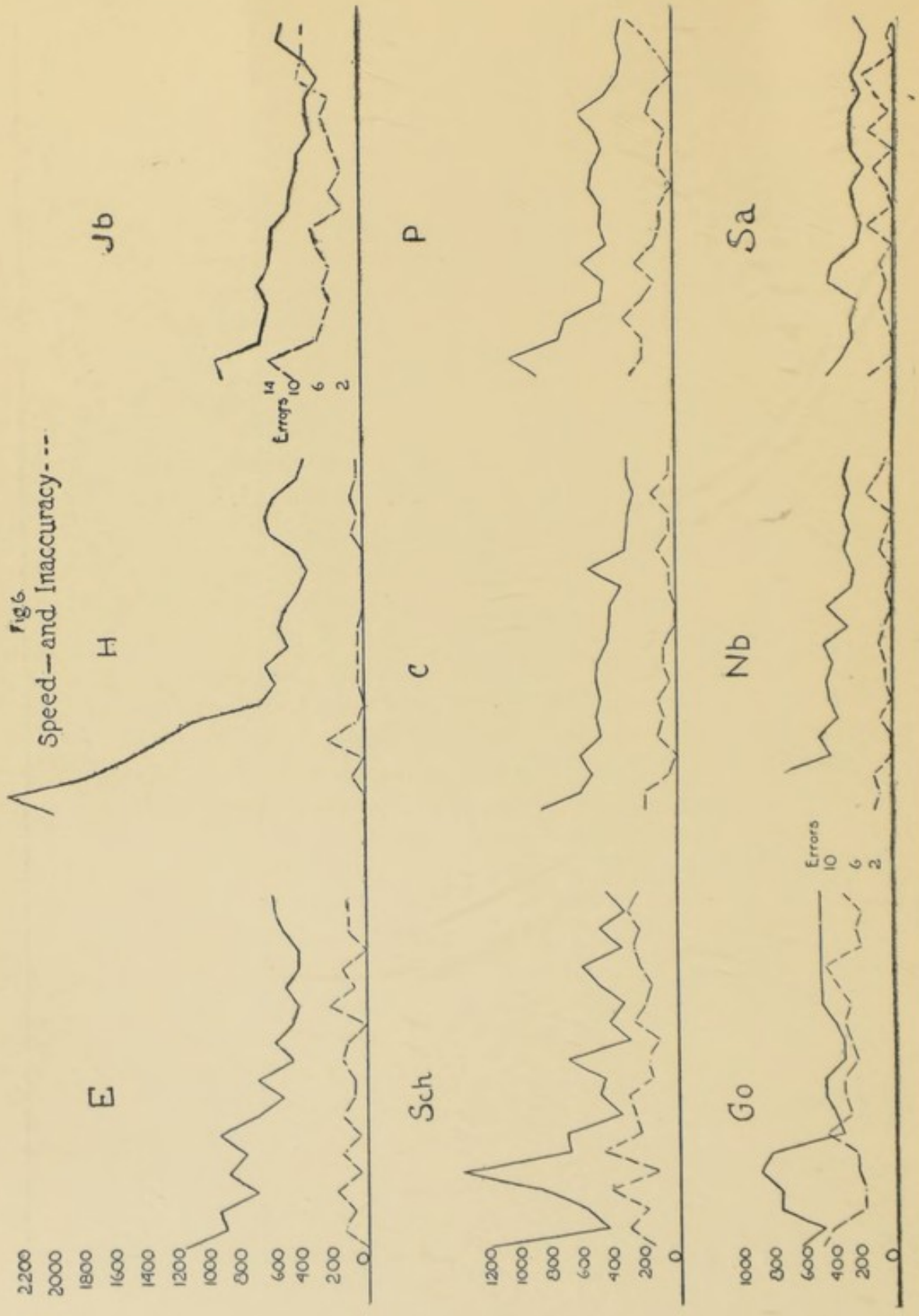


Fig 6
Speed—and Inaccuracy- - -
E
H
Jb
C
P
Sch
Nb
Sa
Go
Errors
14
10
6
2
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0

TABLE LII

SCORES IN MULTIPLICATION TEST

Day	E.	H.	Jb.	Sch.	C.	P.	Go.	Nb.	Sa.	Average
1	400	685	408	492	360	344	284	303	189	389
2	348	790	382	196	262	419	218	310	119	338
3	344	642	251	234	198	315	332	308	97	302
4	255	510	242	395	207	307	338	145	108	282
5	316	628	240	303	189	193	380	175	99	280
6	320	389	278	328	187	160	356	142	203	262
7	340	220	235	289	191	240	164	150	165	221
8	294	209	240	117	180	174	182	166	85	187
9	204	227	260	214	184	165	214	146	88	189
10	276	176	168	166	168	160	210	182	81	176
11	196	184	180	275	145	174	156	113	93	169
12	212	170	154	116	145	192	156	164	69	152
13	175	145	155	193	121	172	208	110	84	151
14	176	120	121	147	216	176	234	95	98	153
15	175	145	135	239	109	261	240	130	94	170
16	166	219	141	278	96	172	249	111	83	168
17	140	229	116	153	76	113	228	110	108	142
18	216	211	147	217	96	126	234	127	82	161
19	225	145	233	148	94	188	228	132	81	164
20	242	131	225	195	96	133	237	110	97	163
Average	251	307	216	235	166	212	242	163	106	

Below are the rankings of the nine given as for the other three tests considered so far. Jb. and E. are perhaps penalized here as their last few records were worse than say the fourteenth and fifteenth. Otherwise the correlations would all be closer. It must be remembered too that the steps in the *speed* ranking are much more unequal than in some of the other tests.

TABLE LIII

	Position at Start (1 = Least Time)	At Finish (1 = Least Time)	Av. Position (1 = Least Time)	Speed (1 = Least Time)	Accuracy (1 = Fewest Errors)	Gross Gain (1 = Most)	Per cent. Gain (1 = Most)
E.	6	7.5	8	8	5	7	7
H.	9	4	9	9	1	1	1
Jb.	8	9	5	4	8	6	6
Sch.	5	6	6	7	7	5	5
C.	4	2	3	2	4	3	2
P.	7	5	4	5	6	2	4
Go.	2	7.5	7	6	9	8	8
Nb.	3	3	2	3	2	4	3
Sa.	1	1	1	1	3	9	9

The correlations are:

Position at start and at finish ..... $R = .44$
Position at start and average position..... .58
Position at start and gross gain..... —.63
Position at start and per cent. gain..... —.52
Speed and accuracy ..... .10

The same general conclusions would be drawn as for the other tests, except that there is a slight positive relationship between speed and accuracy. Possibly the quasi-automatism in the familiar arithmetic processes noticed by C. may account for this.

In the *maze test* the scoring was done—as with other subjects—by adding .1 to the time taken for 1 or 2 touches, .2 for 3 or 4 touches, .3 for 5 or 6 touches and so on. The daily scores resulting are given below and the curves plotted from them in Fig. 7.

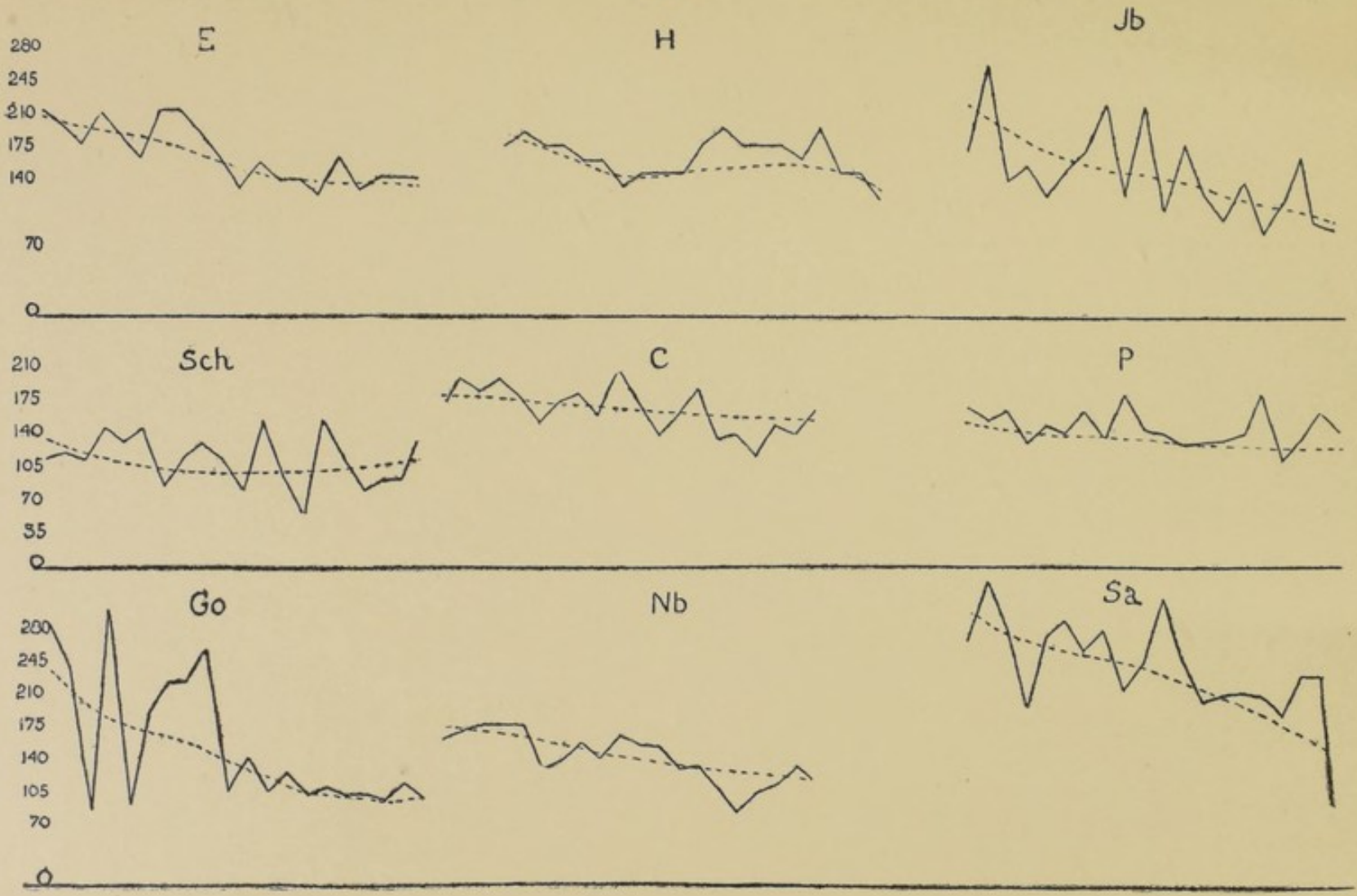
TABLE LIV

E.	H.	Jb.	Sch.	C.	P.	Go.	Nb.	Sa.	Average
216	180	174	118	165	170	288	165	172	194
204	195	264	121	198	154	240	174	333	209
180	180	143	117	187	165	90	180	280	169
216	180	159	150	198	130	306	180	195	190
192	165	121	135	181	148	96	180	273	166
168	165	154	153	160	143	192	132	290	173
216	135	176	89	181	165	228	148	259	177
216	150	224	117	190	135	228	161	281	189
195	150	120	132	160	182	264	144	215	173
168	150	221	117	209	152	108	168	244	171
132	181	108	84	176	140	144	158	316	159
168	198	187	154	140	128	108	156	247	165
144	181	120	100	160	130	132	132	203	145
144	180	100	55	190	132	108	135	210	139
126	180	142	156	130	139	114	110	212	145
168	165	88	121	140	182	108	88	210	141
132	198	120	89	120	115	108	108	190	131
144	150	168	96	149	135	102	117	231	143
144	148	99	96	143	165	120	135	231	142
144	120	90	144	165	144	102	120	82	123
Av. 171	162	148	117	162	147	159	144	238	

It must be remembered that these are only single trials; also, from experience with other subjects, notably the long-term group and R. and Wy., that a conscious attention to speed is accompanied by decreased accuracy. No track was kept by these nine subjects as to whether they attended more to speed or to accuracy. The oral directions emphasized the latter, but the general conditions of the test—timing themselves and having to enter the time—would probably emphasize the former. From these facts then very irregular curves would be expected, which is exactly what is shown.

Go.'s apparent regularity in the second half is due partly to her careless entries of whole minutes, partly to her consistently high number of touches. H.'s comparative smoothness is due to her almost perfect record for accuracy. When these curves are smoothed out C. and P. are most alike, Sch. and Sa. most unlike.

Fig 7
Maze


E
H
Jb
280
245
210
175
140
70
0
Sch
C
P
210
175
140
105
70
35
0
Go
Nb
Sa
200
245
210
175
140
105
70
0

The rankings are given below as for the other tests, and also the correlations worked out from them.

TABLE LV

	At Start (1=Least Time)	At Finish (1=Least Time)	Average Position (1=Least Time)	Speed (1=Least Time)	Accuracy (1=Fewest Touches)	Gross Gain (1=Most)	Per cent. Gain (1=Most)
E.	5	6	8	4	8	5	5
H.	4	5	6.5	8	1	6	6
Jb.	6	1	4	5	5	3	2
Sch.	7	3	1	2	7	4	4
C.	3	7.5	6.5	7	3	8	8
P.	1	7.5	3	6	4	9	9
Go.	8	2	5	1	9	1	1
Nb.	2	4	2	3	6	7	7
Sa.	9	9	9	9	2	2	3

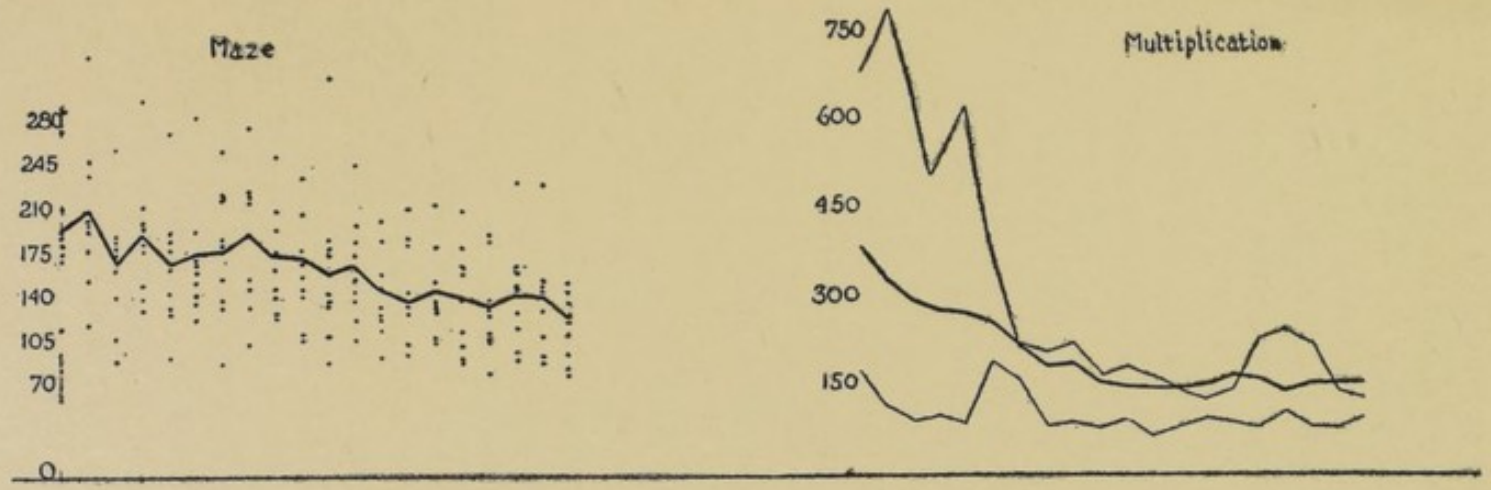
The correlations are:

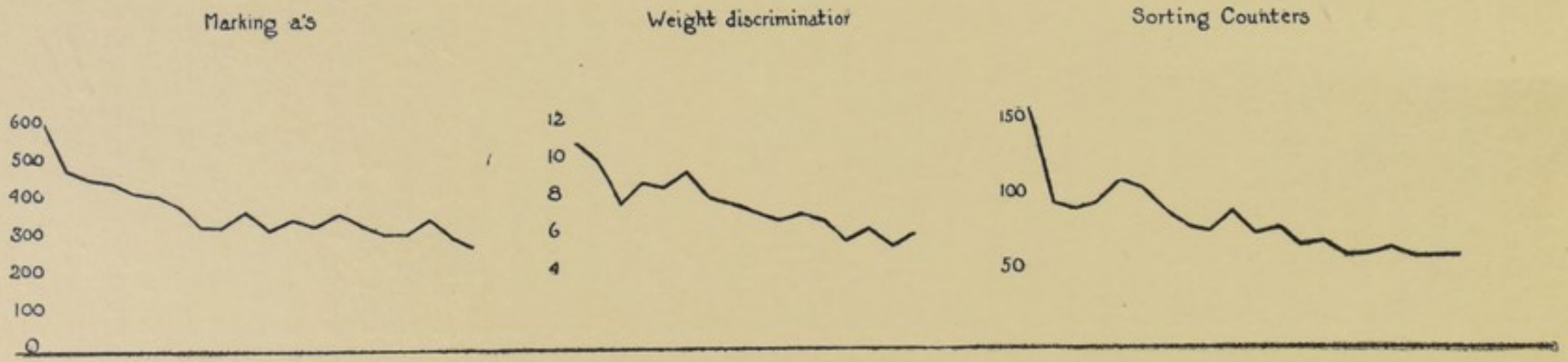
Position at start and at finish	$R = -.21$
Position at start and average position	.33
Position at start and gross gain	−.95
Position at start and per cent. gain	−.90
Speed and accuracy	−.93

In this test the subjects do not keep their relative positions through the series; and, as might be expected, speed and accuracy are almost completely inversely correlated.

Now to examine the data for answers to the questions raised: first, is a mean curve for a test representative of the test or do individual curves differ too much from it and each other to make it reliable? After all, since any average tells little unless accompanied by a statement of the variability, and since a curve of practise is nothing but a series of such non-significant averages, one would not expect a mean curve to be representative of anything beyond the fact of change. Still, the changes in rate of improvement as shown by the mean curve may be different with different functions, or there may be one typical curve of practise to which all functions approximate. In Fig. 8 are shown five mean curves, one for each test. That for the maze is accompanied by a scattering of dots to show the distribution of the nine around each average point; that for mental multiplication is accompanied by the two most distinctly different curves, those of H. and Sa. to show the range. Without these representations of variability there is nothing to distinguish one curve from the others. All alike show greater improvement near the beginning and only slight irregularity after about the seventh day.

Fig. 8.
Mean Curve For Each Test.





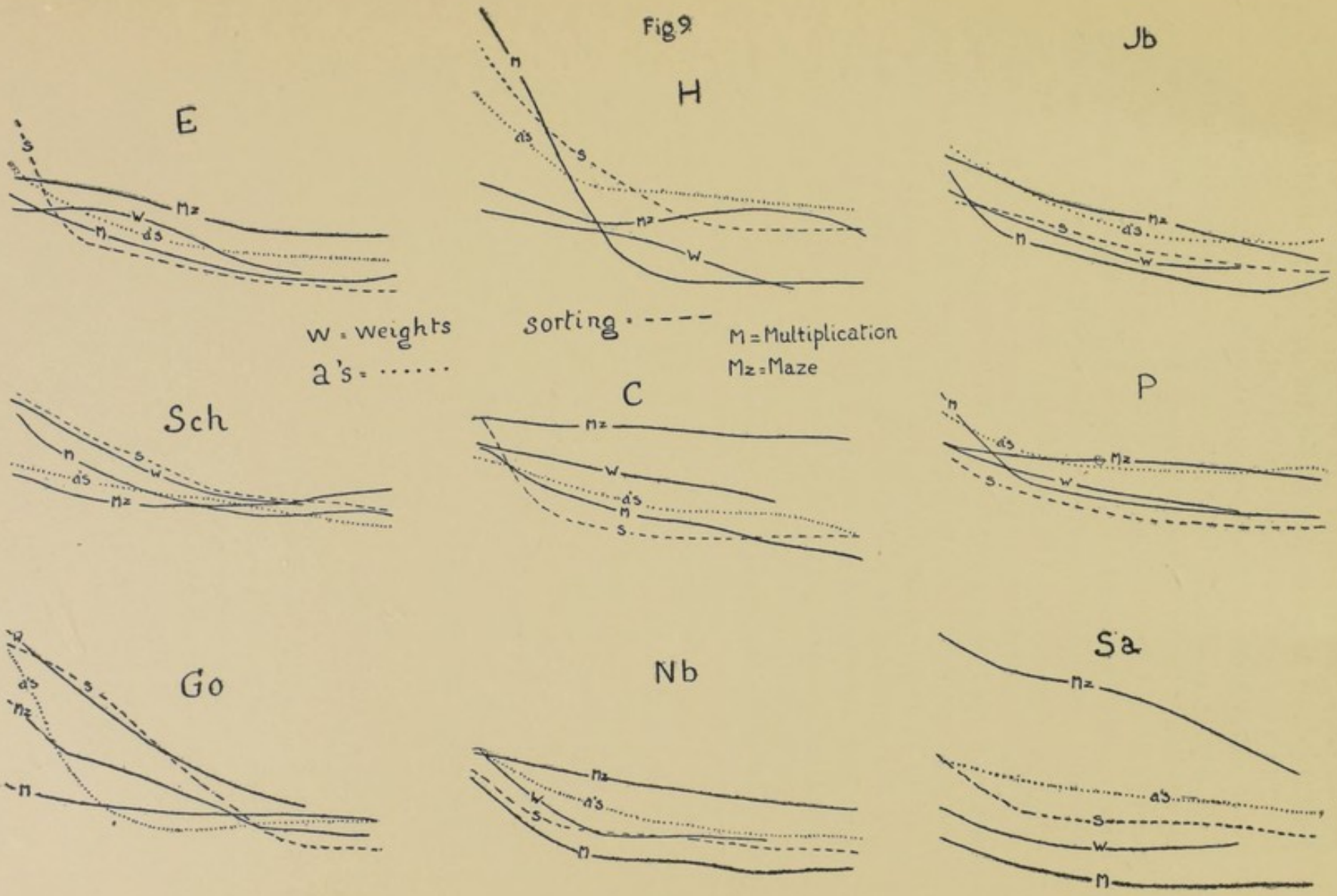
All the functions do seem to approximate one typical law for changes in the rate of improvement.

The second question, are the changes in the rate of improvement different with different individuals or is there one typical curve of practise to which all individuals approximate is answered so far as these data go by Fig. 9. In this are shown the nine sets of smoothed curves, one for each individual. Those for C. and P. are different from those of H. and Go., the former being level and smooth, the latter with a sharp slant near the beginning. Sa. also belongs to the former group, only her relative position in the various tests is very different. Jb. and Nb. show a moderate slant in practically all; E. and Sch. have a mixture of types. This may mean that practise does disclose easily recognizable individual differences, that some people improve rapidly at first, others at about the same rate all the time. Or it may mean only that giving a few trials shows at the beginning a great range of abilities and that the range is lessened with practise. Those who are poor in ability have the greatest leeway to make up and so improve rapidly, while every one improves rather slowly once a certain degree of ability is reached. Thus if comparison is made after the sharp initial slant is over, individual curves will resemble each other in form very closely. In general it seems most probable that if all individuals could start with absolutely zero practise and their changes in rate of improvement up to the limit of improvement be measured, that their curves would resemble each other very closely. The apparent differences as found are so largely caused by the very different levels at which they start, as well as to chance variations in their daily performances.

Individual differences do however occur in the consistency of performance shown by the relative freedom from irregularity in the slope of the curve. If the irregularities of C. and Sa. on the one hand and of E., Go., and Sch., on the other, were computed, the general tendency of the three last to more irregular progress than that shown by the former two would be found much greater than would be expected by chance. This difference is, however, simply one form of the general differences in variability of performance, not anything peculiar to the learning process by itself.

However, since all C.'s curves are not alike, nor all Go.'s, it may be that there is some truth in the third condition suggested, namely, that a curve reveals special not general ability in an individual. That is, that in some kinds of work an individual who is good in any case when compared with others will make steady though slight improvement, while one who is relatively poor will either improve rapidly at first and irregularly for a considerable period, as Sch. in

Fig 9


E
H
Jb
Sch
C
P
Go
Nb
Sa
W = weights
a's = ...
sorting = ----
M = Multiplication
Mz = Maze

mental multiplication, or he will improve very little if at all, perhaps regularly as E. in the maze but more likely irregularly, as E. in weights and sorting, and Sa. in the maze.

In other kinds of work the individual's initial ability may be relatively very different but his tendency toward great irregularity in practise or the reverse may persist. Even in a test such as judgments of lifted weights where all nine curves are more or less irregular, those from C. and Sa. and perhaps Nb. who were notably regular in the other tests are less irregular than those from E., Go., and Jb., who were irregular in other tests as well.

Finally, if irregularity is disregarded and all curves smoothed out, only those facts conforming to the "law of the practise curve" are represented, namely, that a person improves in any work most rapidly at first and makes little and slow improvement after reaching a certain degree of ability. From this point of view, since smoothed mean curves resemble each other no matter whence their derivation, practise must tend to make people more alike.

IV

CONCLUSIONS

REVIEWING this experimental study as a whole, it may be said to offer evidence in reply to certain criticisms of the method of mental tests.

1. In the first place the kind of tests given are said to be of little significance, that knowing how many A's an individual can cancel in a given time, or how many objects he can sort or how many opposites he can name tells us very little about him. This is probably true to a certain extent, since the simpler the performance the more alike individuals will probably be. Complex processes from real life may often be more significant but are necessarily less precise, less convenient, less well recorded and scored, and may therefore be limited to the descriptive stage of investigation. Making more precise measurements need not exclude descriptive work, however, for, in individual tests at least, details of temperament, speed in responding, intelligence in understanding and following directions can be noted, while in addition there will be the objective record to serve as basis of comparison. Then too, with careful experimentation, the tests proven most typical or significant can be selected and administered in the best way. For instance, the easy opposite test given by the time-limit method seems to be a truer measure of the speed of association than the first-idea test by the amount-limit method. The straight maze if improved with respect to length and continuity of movement would probably be more significant and precise as a measure of speed and accuracy of movement than is the hitting 100 dots.

2. In the second place, the criticism that a single trial is unreliable is true but need not be exaggerated since other facts such as state of fatigue, time of day, temporary embarrassment, inclination for work and familiarity with the environment and the kind of material used also enter in to make trials unreliable. To overcome this in part, at least two trials should be made of any test, preferably in addition to a few minutes fore-exercise in similar work. Fewer tests each administered oftener would give a truer estimate of an individual and a better basis for comparison and correlations. It might be advisable to allow sufficient time for each test to get the

average divergence of the obtained result for an individual from the true result down to some standard of reliability agreed upon by various investigators.

3. In the third place, the criticism that giving only a few trials measures not the mental process supposedly tested but merely adaptability to strange conditions such as apparatus, instructions, working for speed, and the particular requirements of the test is seldom of weight. Early improvement due to this alone is rare, and even so could be checked by proportionate fore-exercise and the choice of a proper test.

4. In the fourth place, the criticism that tests measure the degree or amount of previous similar experience rather than actual capacity is true not only of such tests but of any form of mental measurement. It should operate only against expecting too much from the tests, not against their use, but rather, in fact, in favor of repeating them at stated intervals. The only alternative—testing subjects with no similar previous experience or else those whose training had brought them to the physiological limit—would be impracticable, and out of the question. In general, tests of a novel, little-trained function such as *grouped objects* or the $a - t$ test show greater susceptibility to practise than those of a frequently used, much trained function such as addition.

5. In the fifth place, in estimating the nature and degree of improvement in a function with repeated trials the nature of the units used to express such improvement must be taken into consideration, and misleading statements based upon one form of measurement only must be guarded against. Moreover, when comparisons of changes are to be made, whether between different processes in an individual or a group, or between different individuals in one process, it becomes still more important to use more than one way of treating measurements.

6. In the sixth place, the criticism that practise may influence individuals each by a law of his own and processes each by a law of its own does not seem to hold so far as the general law of improvement goes. On the whole, higher mental functions are sooner susceptible to practise than are sensory functions, the more so again if they are novel. Individuals with low standing can and do improve the most, judging objectively, though even so they may not, in conveniently measurable periods of time, overtake those whose standing was high at the beginning. Characteristic variability or consistency of performance may be disclosed whatever the process and whatever the change in improvement.

APPENDIX

KEY FOR CORRECTION OF OPPOSITES

Right, scored 2. (Second choice, scored 1.) *Wrong scored 0*

Above	Below, beneath, *under, down*
Absent	Present (here)
Adroit	Awkward, clumsy (unskilful, unskilled)
After	Before (ahead)
Apart	Together (with, near)
Asleep	Awake
Backwards	Forwards (frontwards)
Barbarous	Civilized (humane), *tame, cultivated*
Best	Worst
Big	Little (small)
Bless	Curse
Broad	Narrow (thin)
Broken	Whole (mended, unbroken, intact)
Brother	Sister
Buy	Sell
Cheap	Dear, expensive
Clumsy	Adroit, deft, skilful, neat (adept, agile, graceful), *clever*
Come	Go
Country	City, town
Create	Destroy, annihilate, tear down (abolish, spoil)
Day	Night
Dead	Alive, living
Deceitful	Sincere, straightforward (truthful, honest, frank, candid, honorable), *open, true, ingenious, upright*
Degrade	Elevate (exalt, uplift, raise, ennoble, promote, advance, restore, honor)
Diligent	Lazy, indolent
Elation	Depression, dejection (despondency, low-spiritedness)
Enrage	Pacify (subdue, appease, calm), *quiet*
Exciting	Depressing, quieting, soothing (calm, restful)
Expand	Contract, condense (decrease, narrow), *enclose*
to Float	Sink (anchor)
Forcible	Weak (gentle), *gently*
Frequently	Seldom, rarely (not often, occasionally)
Generous	Stingy, parsimonious (miserly, greedy, mean), *avaricious*
Gentle	Rough (rude, harsh)
Genuine	False, spurious (counterfeit, sham, insincere, artificial, unreal, imitation, fictitious), *fake, bogus, adulterated, spurious*
Grand	Simple, trivial (poor, petty, modest, ordinary, humble, mean, ignoble, plain, commonplace, insignificant), *tawdry, mediocre, lowly*
Here	There
Hinder	Help, aid, further (promote, advance, assist, hasten, quicken)

Hold	Let go, release, drop (lose, give up, loosen), *give, loose*
If	Unless (although, certainly)
Ignorant	Wise (informed, learned, knowing, educated, intelligent)
to Lack	Have, possess, abound (have in abundance, gain), *attain*
Land	Water (sea)
Less	More
Level	Uneven, slanting, sloping, inclined (rugged, hilly, mountainous, irregular, undulating), *jagged, rough, bumpy, broken*
Loquacious	Taciturn, silent (quiet, reticent, reserved)
Mine	Yours (his, theirs), *your*
Motion	Rest (still, standstill, stillness, quiet)
Obscure	Clear, lucid (plain, evident, light, bright), *open, significant*
Over	*Under* (below, beneath)
Part	Whole, meet (totality, entirely)
Past	Future (present)
Permanent	Temporary (transitory, transient, fleeting), *ephemeral, evanescent, unstable, changing*
Permit	Forbid, deny, prohibit (prevent, refuse), *hinder*
Precise	Inexact (careless, slovenly, disorderly, lax, indefinite, vague, inaccurate), *irregular, loose*
Proud	Humble, *cosmopolitan, democratic*
Repulsion	Attraction, *liking, encouragement, acceptance*
to Respect	Despise (look down on, disregard, insult), *abhor, scorn, loathe, dislike*
to Reveal	Conceal, keep secret (hide, obscure, cover up, keep back)
Rough	Smooth, gentle (calm, tender), *easy*
Rude	Polite, civil, courteous (cultured, sophisticated, obliging, gentle), *refined, fine, polished*
Separate	Together, combined, meet, join, connect (collective, united, continuous)
Serious	Frivolous, gay (merry, laughing, joking, jocular, mirthful, lively), *jocose, funny, silly, cheerful*
Simple	Complex (hard, wise, clever, complicated, difficult, intricate, profound, elaborate)
Son	Daughter (father)
Spend	Save (keep, hoard), *hold*
Stormy	Calm (clear, quiet, fine, peaceful, smooth, tranquil), *fair, mild*
Straight	Crooked (curved)
Stupid	Sensible, bright, clever (smart), *wise, alert*
Take	Give (leave, let alone)
Tall	Short
Unless	If (in spite of, though), *because*
Vertical	Horizontal (slanting), *crooked, perpendicular*
Weary	Fresh (refreshed, rested, brisk, lively), *energetic*
Wicked	Righteous, good (holy)
Wild	Tame, cultivated (civilized)
Win	Lose

PARAGRAPHS USED IN THE EBBINGHAUS COMBINATION TEST

I–XX were specially prepared for the long-term group. The remaining paragraphs, prepared by other investigators, were used with the short-term group.

I.

The argument amounts .. this, that like consequents must .... like antecedents. But it is impossible for the antecedents to be ..... alike, in that the thoughts and feelings ..... give rise to my movements are immediately given, while ..... which give rise to ..... people's movements are ... given. The question presents ....., whether this essential ..... in the mode of existence .. the antecedents does not wreck the analogy.

II.

From the facts thus ... presented, it would be natural to infer ..... mind and body are, in respect of action, on a footing .. equality. The interactionist, at this point, might be tempted to set up the ..... that every fact showing the influence of .... upon mind can be matched with a .... showing ... . of .... upon ...., and that by as much as the former demonstrates the mind's dependence, the ..... demonstrates its power.

III.

In every actual case of perception, the entire fact is not ..... the presence of a physical ..... to consciousness, but at the same ...., and as a condition of that presence, the existence of a train of ..... and effects connecting the object .... the percipient's ..... If I ..... a table, this involves the presence in the ..... world, along with the table, of light-rays passing from the ..... to the eye, and ..... passing from the eye to the brain.

IV.

Parliament had hitherto ..... very little attention on our Eastern possessions. Since the death of George II., a rapid ..... of weak administrations each of ..... was in turn flattered and betrayed by the Court, had held the ..... of power. Intrigues in the palace, riots in the capital, and insurrectionary ..... in the American colonies had left the advisers of the Crown little time to study Indian politics. When they did interfere their interference was ..... and irresolute. Lord Chatham had ..... a bold attack on the Company, but his plans were rendered ..... by the strange malady which about that ..... began to overcloud his splendid genius. At length it was generally felt that Parliament could no longer ..... the affairs of India.

V.

Very similar to this was the state of India sixty years .... Of the existing governments not a single one could lay ..... to legitimacy. There was scarcely a province in which the real sovereignty and the ..... sovereignty were not disjoined. Titles and forms were still ..... which implied that the heir of Jamerlane was absolute ..... when in reality he was a captive. The Nabobs were, in some ..... independent princes; in others, they had, ... their master, become ..... phantoms and the Company was supreme. Among the Mahrattas the heir still .... the title of Rajah; but he was a prisoner, and his prime minister had ..... the chief of the state.

VI.

In a rude state of society men are children with a greater variety of ideas. It is in such a state of society that we may ..... to find the poetical temperament in its ..... perfection. In an enlightened ... there will be much

intelligence, much ....., much philosophy, abundance of just classification and subtle ....., abundance of wit and eloquence, abundance of verses and even of .... ones; but little ..... Men will talk about the old poets and comment on them, and to a certain extent ..... them, but they will scarcely be able to ..... the effect which poetry produced upon their ruder ....., the ecstasy, the plenitude of belief.

VII.

One of his gifts was a voice habitually deep and sonorous yet capable of ..... very low and gentle at the ..... moment. About his ordinary bearing ..... was a certain fling, a fearless expectation of success, a confidence in his own ..... and integrity much fortified by contempt for .... obstacles or seductions of ..... he had had .. experience. Mr. B. perhaps liked him the ..... for the difference between ...., and certainly for being a stranger. One can begin so many things with a ... person!

VIII.

He had never put any question concerning the nature of his illness, nor had he betrayed any ..... as to how far it might be likely to cut ..... his labors or his life. On this point, as on all others he ..... from pity; and if the suspicion of being pitied for anything surmised or known in ..... of himself was embittering, the idea of calling ..... a show of compassion by frankly ..... an alarm was intolerable. Every proud mind knows something of this ..... and perhaps it is only to be ..... by a sense of fellowship deep ..... to make all efforts at isolation seem mean and petty ..... of exalting.

IX.

Her belief that Rosamond could manage her papa was well founded. Mr. Vincy had as ..... of his own way as if he had been a prime minister: the force of ..... was easily too .... for him as it is for most pleasure-loving, florid ...; and Rosamond was ..... forcible by means of that mild persistence which enables a soft living substance to make its ... in spite of opposing rock. Papa was no rock. He had no fixity but that of alternating impulses sometimes ..... habit, and .... was altogether unfavorable to his taking a decisive line of ..... in relation to his ..... engagement.

X.

Soldier wake, the ... is peeping
Honor ne'er was ... in sleeping,
Never .... the sunbeams still
Lay unreflected on the ....:
'Tis when they are glinted ....
From axe and armor, spear and jack,
That they promise ..... story
Many a page of deathless .....
Shields that are the foeman's terror
Ever ... the morning's mirror.

Soldier, ..., thy harvest, fame;
Thy study, conquest; war, thy ....

XI.

And is she happy? Does she see unmoved
The .... in which she ..... have lived and loved
Slip without ..... bliss slowly away,
One after one, ... like to-day?
Joy has ... found her yet, nor ever will,
Is it this ..... which makes her mien so still
Her features .. fatigued, her eyes, tho' sweet,
So sunk, so rarely ..... save to meet
Her children's? She moves slow; her voice alone
Hath yet an infantine and silver tone,
But .... that comes languidly: in truth
She ..... one dying in a mask of youth.

XII.

Move eastward, happy earth, and leave
Yon orange ..... waning slow;
From fringes of the ..... eve
O, happy planet, ... go;
Till over thy dark shoulder glow
... silver sister ....., and rise
To glass herself in dewy eyes
That ..... me from the glen below.

Ah, bear me with ...., lightly borne,
Dip forward under ..... light
And move me to my marriage ....
And round ..... to happy night.

XIII.

Professor Crocker presented his trained animals yesterday afternoon and ..... and was greeted .. large houses on both ..... The production is unique and ..... an interesting lesson in ..... education, some .. the tricks ..... by the four-footed actors being really ..... His troup consists of 25 animals, and .... has a role to .....

XIV.

Weather that was pleasant only at times, and at ..... times threatening or rainy made ..... unpleasant conditions ... yesterday's observance of Dominion day, and .... a damper on many festivities. The morning dawned bright and ..... and scores of ..... parties left the city on excursions. Towards noon it became cloudy and there were some ..... Again it cleared up, only to be ..... later by heavy thunder, lightning and rain, though the ..... in the city was light ..... to what it was in the surburbs.

XV.

The longshoremen of the Cunard ..... pier who struck yesterday ..... the steamship Umbria arrived to ..... the company to pay them sixty ..... instead of fifty-five cents an hour for Sunday ....., returned to work to-day. Their demand was not ..... The chairman of the ..... said to-day that he was at a loss to ..... the reason for the action of the ..... men. He said the union did not ..... the strike.

XVI.

The magnetic dip needle is made in the form of a lozenge, ... to the horizontal needle, but it is poised or ..... by ..... of a shaft running through the center of the lozenge at right ..... to it, and is held in ..... by agate bearings as ..... in figure 20. In some types the cradle ..... the horizontal shaft is poised on a steel needle. The needle is thus ..... to take up a position ..... and south and to incline on its .....

XVII.

It is natural to believe in great men. Nature seems to ..... for the excellent. The world is upheld by the veracity of .... men; they make the earth wholesome. They who lived .... them found life glad and nutritious. Life is sweet and tolerable only in our belief in .... society; and actually, or ideally, we manage to .... with our superiors. We call our children and our lands by their ..... Their names are ..... into the verbs of language, their works and effigies are in our ....., and every circumstance of the ... recalls an anecdote of them.

XVIII.

If he had been an English nobleman on a pleasure tour, or a newspaper courier, he could not have ..... more quickly. The post boys wondered at the fees he ..... amongst them. How happy and green the country ..... as the chaise whirled ..... from milestone to milestone, through neat country towns where landlords ..... out to welcome him with ..... and bows; by pretty roadside inns where the signs .... on the elms, and horses and men were drinking under the checkered ..... of the trees; rustic hamlets ..... round ancient grey churches, and through the friendly English landscape. To a traveller returning ..... it looks so kind.

XIX.

Nay, ye should not weep, my children!
Leave it to the faint and weak;
Sobs are ... a woman's weapon
Tears befit a maiden's .....
Weep not, ... of MacDonald!
..... not thou, his orphan heir.
Not in shame, but ..... honor
Lies thy slaughtered ..... there.
Weep not, but when years are over
And thine arm is ..... and sure,
. . . . .
Let thy heart be ..... as iron
And thy wrath as fierce .. fire,
Till the hour when ..... cometh
For the race that slew thy sire!

XX.

An electrical storm of ..... severity passed over this district last night, which burned barns, killed cows in the field, put telephones and ..... lines out of commission, knocked ..... trees, and did a great deal of ..... generally. The flag staff was struck and splintered and the slates were .... off the

roof. A barn was burned with a large ..... of hay, and a driving shed was ..... destroyed. Crops in all ..... were almost pounded into the .....

XXI.

We confess to something of sympathy ... the correspondent ... hinted yesterday that ... children are ... over and killed by automobiles, the .... is not always that .. the automobilist, ... sometimes rests in some measure on those who do not ..... their children to avoid unnecessary ..... It is a plain ...., of course, that public highways are ... the use of the whole population, ... that the automobilist is ..... every obligation .. keep the limitations of his rights and privileges .. mind as he goes along, but the road is his .. well as other peoples.

XXII.

If we are ..... well, thoroughly sound, we .... not be depressed. The perfectly healthy animal ... no worries. The remedy has already .... indicated. Regretfully it is .. simple .... very few people take the trouble to ..... it. .... it is clearly and widely recognized that ..... is stupid, that its .... is simple where ..... is no organic trouble, worry will ..... Worry is simply a .... of what, ... the sake of a nice large word, is called "neurasthenia," nerve-depletion. ..... plenty of recreation, plenty of fresh air, and the ..... man will not worry.

XXIII.

Park Hill on the Hudson offers you a solution .. the home problem to-day. No home seeker .. investor ... afford to ignore its claims. Escape the wear and tear .. the city's noise ... rush .. this open air paradise, just .. the city's edge, .. all respects an ideal home location ... yourself and family. ..... are cottages containing every improvement waiting ... you to step .. and make yourself comfortable. It not .... commands the most beautiful view around New York ... is protected for all time ..... intrusion. Choice lots now ..... on very easy terms.

XXIV.

A law .. defence of property rights in the broadest sense .. observed ..... almost abolish international conflicts. Gentlemen .. not fight with fists .. money differences ... do they refer them .. courts of honor. Civil courts are for that ..... and are as useful for nations as for men. The sanction of international law must .. merely moral, for a long time .. least. But in ..... that there should be ... moral sanction there must .. a moral code. The principles of .... a code are deducible .... treaties to which nations have set their hands ... seals.

XXV.

I asked the slovenly, ... cheerful female ... answered the bell ... the landlady, wondering the while .... I should say when I was asked ... references. The merriment had not been called forth .. anything amusing .. my appearance, .. my vanity had feared, ... by a story which a man sitting .. ... head of the table was just finishing. The only vacant chair .. the room was beside him, and, rather awkwardly, ... I felt that they were ..... my measure, I made my ... toward it. As I ... down he greeted .. with a polite bow.

XXVI.

The occult in everyday affairs is the ..... of this new book .. Robert Chalmers. ..... one of the thrilling stories of ..... the volume is composed .. the tale of some awful mysterious happening, some supernatural ..... beyond the ..... of material reasoning of mortal man .. explain, which comes .... the life of some ordinary, everyday man. The opening ..... tells of a dinner ..... to a man deeply versed in occultism .. his American friends. To these he gives many hints ... suggestions of momentous things which he ... plainly see ..... for them .. the future.

XXVII.

We believe we can prove .. you that this investment is .. secure ... the dividends so sure, that it justifies you .. withdrawing money .... the Savings Banks, ... it is earning 3½% and putting it .. our business where it will earn 7%. We are a New England enterprise, managed .. New England men, and we have behind .. a record .. fourteen years of unbroken success. ..... you have much or little you can not ..... to let slip this opportunity of doubling the ..... from your savings. Prompt action in this matter will ..... you well.

XXVIII.

On the ....., it didn't cost me a dollar. In fact, though at ..... I have found myself ..... of considerable sums of ready money, I have never .... a man of property .. the strict sense of the word. I abandoned my ....., the law, .. I did not .... its practice so lucrative .. I had hoped. For some years thereafter I traveled largely .. the Mississippi River. It ... the decline in steamboating ... the adoption .. less leisurely methods of travel .... cut into my income and forced .. to come North and ..... in trade.