ARCHIVES OF PSYCHOLOGY

EDITED BY
R. S. WOODWORTH

VOLUME VI
1917-1921

NEW YORK
1922

CONTENTS

RECITATION AS A FACTOR IN MEMORIZING

ach
G

RECITATION AS A FACTOR IN MEMORIZING

BY

ARTHUR I. GATES, Ph.D.

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. Woodworth

No. 40, SEPTEMBER, 1917

COLUMBIA UNIVERSITY CONTRIBUTIONS TO PHILOSOPHY AND PSYCHOLOGY

Volume XXVI, No. 1

NEW YORK

THE SCIENCE PRESS

Agents: G. E. Stechert & Co.: London (2 Star Yard, Carey Street, W. C.); Leipzig (Koenigstr. 37); Paris (16, rue de Condé)

UNIVERSITY PRINTING OFFICE
COLUMBIA UNIVERSITY

TABLE OF CONTENTS

PREFACE

The present investigation was begun in the Psychological Laboratory of the University of California in the spring of 1916. The experiments in which children served as subjects were conducted in a public school in Oakland, California, while those upon adults were, for the most part, carried out in the Psychological Laboratory of Columbia University.

The writer has been fortunate in having enjoyed, during the course of the work, endless encouragement, suggestions, and assistance from a large number of people. To Mr. N. Ricciardi, Principal of the school visited, I am indebted for the privilege of conducting the experiments upon his charges as well as for the ready help in arranging details for the work. To the many teachers whose class-rooms I invaded, I am indebted for the kindliest toleration and for a great deal of valuable assistance. My debt of gratitude to Professors G. M. Stratton and Warner Brown of the University of California and to Professors J. McKeen Cattell, R. S. Woodworth, E. L. Thorndike, H. L. Hollingworth, and Dr. A. T. Poffenberger of Columbia University, is very great. To my friend Charles E. Martin, I am indebted for valuable suggestions and criticisms in the preparation of the manuscript.

I

INTRODUCTORY STATEMENT OF THE PROBLEM

The process of learning as carried on by most adults depending upon their native resources or practical experience, is frequently interrupted by attempts at recitation or voluntary recall of what has been learned. We tend to introduce an attempt at recitation at the earliest possible moment, usually long before a perfect reproduction is possible. In that case, as a rule, we refer promptly to the material being studied in order to complete the perusal. For example, many years ago Francis Bacon observed, "If you read anything over twenty times you will not learn it by heart so easily as if you were to read it only ten, trying to repeat it between whiles, and when memory failed looking at the book." [1] The spontaneous methods of learning of many people resort so naturally to these attempted reproductions that we can hardly refuse to believe that they are helpful. Yet most of us would admit that the dominating idea behind such a procedure is the fear of studying the lesson more than is absolutely necessary, and it is by no means clear that introducing the recitation too early in the learning process may not result in loss of time. This gives rise to several practical questions, such as:—Is an attempted recitation of as much value in learning as another perusal or reading, and is a recitation at one stage of the learning as valuable as at another?

It is at once obvious that the solution of such questions is of tremendous import for the work of the school. It is imperative that recall or recitation, as a factor in learning, should be analysed and its quantitative importance determined. Although several studies of the problem have been made within the last decade, facts that will permit indisputable application to the work of the school-room are still wanting. The amount of experimentation required to solve the problem adequately is much greater than would at first thought appear, since different results might be expected according to the age and training of the subject, the kind of material employed, the length of the lesson or the purpose of the learner, *i. e.*, whether the material is to be 'learned by heart' or only partly learned. The general condition of the problem is indicated by a recent statement of Meumann, who, after summarizing the work in the field, con-

[1] *Novum Organum*, 1620, translated by James Spedding, edition of 1863, p. 229.

cluded that[2] "it is indispensable that such experiments be repeated and confirmed before the results are applied to pedagogy."

That great variations in the methods of learning exist even among adults with college training will be surprisingly evident to anyone who will select at random twenty such individuals and carefully observe the means employed in learning a stanza of poetry or a series of nonsense syllables. Many cases of alleged 'poor memories' may prove to be due to poor methods of study. One would even more confidently expect to find among children a greater number of ineffective methods of studying. In fact Miss M. J. Baldwin made a study of this matter[3] and found that such was the case. She undertook, by means of questionnaires and observations of pupils in Grammar and High School, to determine the methods of study as well as the methods employed in testing their attainment. The methods of study, as one might expect, were various. Some employed one kind or another of attempted recall, such as trying to say or write the main ideas, but more than one-fourth simply read the lesson through time after time. In some classes from fifty to sixty per cent. of the pupils came to the recitation without having once attempted to test their mastery of the lesson in any definite way.

Manifestly, when so many of our years are to be spent in studying, it is imperative that some information concerning such broad functions as reading and recitation as factors in learning should be obtained and applied. While volumes have been written on methods of study and on the economy of learning, so far as children are concerned, no objective data are available demonstrating the relative value of these two functions which are fundamental in any attempt to learn. Earlier investigators have found in the case of many adults that the optimum combination of recitation and reading may lead to the mastery of a given lesson in one-half the time required to learn it by reading alone. If such findings should hold for children, and if it is generally true, as Miss Baldwin found, that twenty-five per cent. or more of the pupils in the schools rely entirely upon reading in their learning, the loss of time and energy is appalling.

The present study presents the results of an effort to answer a practical problem of the school-room—namely, What are the relative values of learning by reading as compared to learning by recitation in the case of school children working under school conditions and with the ordinary school-room methods of attack? It will be seen later that all the previous work on this subject has employed adults as

[2] *Vorlesungen zur Einführung in die experimentelle Pädagogik*, vol. III, 1914, p. 130.
[3] 'Studies in Development and Learning', *Archives of Psychology*, 1909, No. 12, pp. 65–70.

bered *A*, is again given one or more stimuli and required to pronounce upon the relation if identity which the latter holds (or hold) to the former. Thus the method of reproduction is to be distinguished from that of recognition by the fact that in the former only one artificially given stimulus is required in each experiment, while in the latter at least two are necessary.''

Bean (2) says that the curve of forgetting can be properly measured by reproduction or recognition and that the rate of forgetting differs with the method.

Among the investigators to conduct experiments with the primary interest to determine the difference between recall and recognition is McDougall (13). He presented the material in two forms, first he exposed ten monosyllabic words simultaneously for a period of ten seconds to each of his two subjects. In the second case they were presented verbally. Afterwards the observers were asked to write down all that they could recall within a minute. Then from a set of twenty containing the original ten, they were requested to identify as many as possible. The difference in the per cent. recalled and recognized is about twenty. The results were as follows:

Subject	Recall	Recognition
	VISUAL	
A	54.3%	69.7%
B	59.5%	84.3%
Average	56.9%	77.0%
	AUDITORY	
A	50.0%	74.0%
B	57.0%	74.3%
Average	53.5%	74.15%

The data are too meager to draw any general conclusions, except perhaps that more can be recognized than recalled.

Heine (10) gave the recognition test by presenting the objects one at a time and asked the subject to state if each aroused a "distinct impression" of being known or if it was difficult to say, or if the impression was absent. The time of each was taken and the correct recognitions were quicker than the wrong or doubtful ones. Heine also found that retroactive inhibition had no effect on recognition, but did affect recall.

Strong (16) has used the recognition method extensively. One of the most noticeable results of experimental work is that recognition is easier than recall and has a higher score. Strong found that about 17 out of 20 words, after being presented once, could be recognized immediately afterwards; this is a much higher score than one would expect for recall. Advertisements were shown by Strong one each second in sets varying in length from 3 to 150 and recognition tests were made immediately afterwards. The per cent. or relative amount recognized decreased with the increase in quantity shown, but the decrease was much slower than in the case of recall memory. For 150 the average per cent. recognized by forty subjects was 35 or about 52 words. This is a much higher score than one would expect by the recall method. Strong obtained the curve of forgetting for recognition. The general shape of the curve is similar to the one Ebbinghaus found for recall. The loss of memory is great at first and then proceeds more slowly.

When studying incidental memory Myers used sometimes recall, sometimes recognition, and sometimes both methods of testing according to which the material was best adapted. In connection with the testing of incidental memory for coins he writes: "The test was begun by merely having the circles 'drawn,' but at the suggestion of E. K. Strong, Jr., the recognition test was added and finally used alone. It is certainly preferable because of the time saved both in applying the test and in dealing with the results, though there is the disturbing element due to the large number of circles before the eye." Later when Myers (14) made a comparative study of recognition and recall, he drew the following conclusions:

"1. The recognition efficiency is about two and one half times that of recall, and this ratio varies slightly with different amounts of stimuli and with different intervals of time. However, the total number of words correctly scored in recognition was about four times the number recalled. The difference in ratios is due to the penalty assigned because of chance.

"2. Great individual difference obtains for both recognition and recall, but for recognition it was proportionately higher than for recall, and the difference of time intervals and length of stimuli-list affect the variability of recall more than that of recognition.

"3. The correlation between recall and recognition is surprisingly low: many who recall only one or two words or even none have a remarkably high record for recognition.

. . . .

"5. For long intervals of time more of the answers for recognition are correct than those for recall and this superiority of recognition-answers increases with increase of time interval.

. . . .

"7. The comparative order of frequency for each word, in recognition and recall is about the same for the first words of the stimuli-list, but there is a wide variation for those least frequently recalled.

subjects. But as Meumann has said,[4] "We do not know whether recitation is of the same value for children, nor whether the combination of readings and recitations for optimum results is the same as for adults." In most cases, moreover, the earlier researches were conducted under rather rigorously controlled conditions. The subjects were not permitted to study in their habitual manner; sometimes the material was presented tachistoscopically with a fixed tempo of presentation, sometimes articulation was prohibited or other restrictions enforced. In the present work, so far as practicable, conditions were made as nearly normal as possible. The material selected is comparable to that with which the pupils were accustomed to deal in their daily work. The children studied in much the same manner that they would employ in learning a vocabulary, a spelling lesson, or a history or geography lesson, with the knowledge that at the end of the study period they would be given a written examination. Details of material and methods, however, will be reserved for a later page.

In addition to the experiments upon school children, adult subjects were also tested with similar materials and methods. The data thus obtained will make possible a more adequate comparison of the present findings with those of other investigations and will be of assistance in better interpretation of the results by virtue of the more reliable introspective observations which would be expected from the more experienced learners.

From this study it is hoped that some information will be secured on the following points:

1. The relative value of learning by reading as compared to learning by recitation.
2. The differences in the functions involved in the two methods of learning.
3. The optimum time at which to introduce recitation into the learning process.
4. The relation of the two methods of learning as dependent upon the age or school status of the learner.
5. The relation of the two methods as dependent upon the kind of material employed.
6. Incidental information concerning the learning methods of children and adults.

In the next section a brief summary of the work previously done on the problem will be presented.

[4] *Op. cit.*, vol. III, p. 130.

II

SUMMARY OF PREVIOUS STUDIES ON THE PROBLEM

In 1908 M. Dimitre Katzaroff made a direct attempt[1] to obtain evidence upon the relative value of reading as compared to recitation as factors in the learning process. Series of eight or ten pairs of nonsense syllables were presented serially on a drum which revolved so as to expose each pair for two seconds. A four-second pause was made between perusals. Adult subjects were employed exclusively, each learning three or four series at each sitting, a rest of five minutes being given between series. After a certain number of readings attempts were made at recall by exposing the first word of each pair and calling for the second word. Each syllable was left in view for twenty seconds unless its associate was recalled earlier, and in cases of failure to recall the learner was prompted orally. Various combinations of study and recitation were tried as may be seen from Table I.

The test of memory was made by the 'Treffer' method, the original first members, however, being exposed in a new order after intervals of twenty-four, forty-eight, or seventy-two hours. The time required for each response was measured by a Muensterberg chronoscope. Table I gives a summary of the results.

The general result of these experiments is that recitation, after a certain number of original readings, is more valuable than additional readings. In most cases, the advantage of recitation is very great, measured by the amount correctly recalled, and usually the reaction time is less. Tables D and E also show quite conclusively that recitations grouped are more effective than recitations interspersed with readings.

The greater effectiveness of recitation is explained by Katzaroff as being due in the main: *First,* to a greater control over the conditions of learning. *Second,* to a greater activity of the learner during recitation, "in the readings, the subject is passive, calm, indifferent; in recitations he is active, he has to seek, he rejoices when he has found and is irritated at the syllables which evade his call."[2] *Third,* to a greater confidence with reference to the material learned which is brought about by recitation.

[1] 'Le Rôle de la récitation comme facteur de la mémorisation', *Archives de psychologie,* 1908, **7** pp.224–259.

[2] *Op. cit.,* p. 257.

For several reasons, it would be unsafe to consider Katzaroff's findings as typical. In the first place the results were obtained from too few and highly trained adults. Individual peculiarities may play too prominent a rôle. Moreover, practice effects were not sufficiently taken into account, and finally the mode of presentation was not the same in the two methods. During the perusals by reading the total presentation was visual, but during recitation oral

TABLE I

Showing a summary of results obtained by Katzaroff, op. cit.

Table	Number of subjects	Number of sittings for each	[3]Combination L reading R recitation	Test after number hours	Per cent. correct	Reaction time in seconds
A	1	4	L 10 R L5	48	43	5.6
			L 10 R R5		50	4.0
B	3	4	L8 L7	72	6	8.0
			L8 R7		20	6.7
C	2	4	L8 L7	72	9	7.8
			L8 R L6		15	8.6
D	1	3	L4 L6	24	17	5.0
			L4 R6		46	2.9
			L4 RL RL RL		25	4.5
E	1	3	L4 L6	24	4	2.9
			L4 R3 L3		62	3.1
			L4 RL RL RL		54	2.9

presentation was added to the visual, since the prompts were made by means of the experimenter's voice.

A more recent experiment by Thorndike[4] has given very different results. Twenty-eight adult students learned four vocabularies of twenty pairs each, the second by attentive reading and rereading, the first by reading the first members and trying to recall the second members of the pairs. The third vocabulary was learned in the same manner as the second, and the fourth in the same manner as the first. The results are given in Table II.

There is no apparent superiority in the method involving recall; in fact, the method of reading and rereading seems to give slightly better results. Professor Thorndike explains that "This, however, was partly due to the overlearning of the first vocabulary, there

[3] For the sake of brevity, let L equal reading and R equal an attempted recitation. Thus L10 R5 means ten original readings followed by five attempted recitations.

[4] 'Repetitions versus Recall in Memorizing Vocabularies', *Journal of Educational Psychology*, 1914, 5, pp. 596–597.

being a tendency to take profitable risks in the vocabularies after the first." And moreover, "The experiment was too crude and too slight to give numerical results worth presenting in detail."

A more extensive study has recently been reported by Alexander Kühn.[5] Three kinds of material were employed: vertical rows of

TABLE II

Recall		Reading		Reading		Recall	
Time (seconds)	Number correct	Time (seconds)	Number correct	Time (seconds)	Number correct	Time (seconds)	Number correct
18.6	13.9	16.1	16.9	15.8	15.7	14.7	14.6

twelve nonsense syllables, vertical rows of twelve one-syllable substantives, and short verses. Two methods of studying were employed, one in which the learner read and reread until he was confident of his mastery of the material and another in which the subject was permitted to employ recitation as soon as he desired.

TABLE III

Material	Method of learning[6]	Tempo Controlled				Tempo Free			
		Number of repetitions		Time		Number of repetitions		Time	
		Average	Median	Average	Median	Average	Median	Average	Median
Verses	R	7.5	7.3	2′09″	2′06″	6.5	6.4	1′51″	1′49″
190 tests	L	8.8	8.8	2′26″	2′25″	8.2	8.2	2′19″	2′15″
6 subjects	L—R	1.3	1.5	17″	19″	1.7	1.8	28″	26″
Words	R	15.3	14.6	2′35″	2′31″	11.2	10.9	2′04″	2′01″
132 tests	L	20.8	20.4	3′24″	3′21″	18.8	18.7	2′59″	2′58″
9 subjects	L—R	5.5	5.8	49″	50″	7.6	7.8	55″	57″
Syllables	R	22.7	22.7	3′56″	3′56″	18.1	17.7	3′33″	3′38″
132 tests	L	36.6	36.7	6′06″	6′09″	31.8	31.7	5′22″	5′17″
9 subjects	L—R	13.9	14.0	2′10″	2′13″	13.7	14.0	1′49″	1′49″

In some experiments the tempo of presentation was controlled by means of a metronome, the subjects being required to repeat a word or a nonsense syllable each 0.65 of a second, while in the case of poetry a syllable was repeated each 0.4 second. In other tests, the subjects studied in their preferred tempo. In all cases, simultane-

[5] 'Über Einprägung durch Lesen und durch Rezitieren', *Zeitschrift für Psychologie*, 1914, **68**, pp. 396–481.

[6] R—recitation, L—reading. Sum of tests for all subjects.

ous visual presentaton was employed, the subjects prompting themselves and correcting their own errors. For the learning of syllables and words, trochaic rhythm was specified. The subjects were university graduates and instructors. Each subject learned a dozen or more lessons by each of the two methods. Table III gives a summary of the results.

Table IV gives the relation of reading minus recitation to recitation, computed from the data of Table III.

TABLE IV

The Relation of Reading minus Recitation to Recitation

Material	Tempo controlled				Tempo free			
	Repetitions		Time		Repetitions		Time	
	Average	Median	Average	Median	Average	Median	Average	Median
Verses	0.17	0.20	0.13	0.15	0.26	0.28	0.25	0.24
Words	0.36	0.40	0.31	0.33	0.68	0.71	0.44	0.47
Syllables	0.61	0.62	0.55	0.56	0.76	0.79	0.51	0.52

The general result is that for all materials recitation is a more effective method of learning than reading only. Rather striking individual differences, however, were found, ranging all the way from certain subjects who required more than twice as long to learn a series of nonsense syllables when no recitation was permitted, to others (in all three subjects out of thirteen) for whom reading was an equally or even more effective method of learning than recitation. The latter are representative of a peculiar type of mechanical learners to whom we shall refer again. On the whole, however, the advantage of recitation is clear.

It appears that the advantage of recitation differs considerably according to the kind of material being studied; the more senseless and less connected the material, the greater the advantage of recitation over reading. Thus, Table IV shows the superiority of recitation to be rather small in the learning of verses, about twice as great for learning series of words, and larger still for learning nonsense syllables. The advantage of recitation also differs according to the method of studying that is employed, being in all cases greater when the learner is left to employ his own tempo than when the tempo is controlled by the experimenter. It was found also that in the reading method the subjects were unable to judge so well when the material was mastered and often 'made too early an attempt to recite'. In general, it was found that the controlled tempo hampered the learning to a greater or less extent.

After a lapse of twenty-four, forty-eight, and ninety-six hours, the material was relearned, the same methods being employed that were used in the original learning. Table V shows the average results for six subjects.

TABLE V

After twenty-four hours. Total number of tests—84

Method	Tempo controlled				Tempo free			
	Repetitions		Time		Repetitions		Time	
	Average	Median	Average	Median	Average	Median	Average	Median
R	4.1	4.0	1′06″	1′05″	3.8	3.8	1′00″	0′58″
L	5.0	5.0	1′21″	1′20″	4.6	4.6	1′12″	1′12″
L−R	0.9	1.0	15″	15″	0.8	0.8	12″	14″
After forty-eight hours. Total number of tests—68								
R	3.9	3.8	1′03″	1′02″	3.7	3.7	0′57″	0′55″
L	4.5	4.4	1′14″	1′13″	4.5	4.3	1′12″	1′12″
L−R	0.6	0.6	11″	11″	0.8	0.6	15″	17″
After ninety-six hours. Total number of tests—38								
R	3.8	3.9	1′03″	1′02″	3.3	3.3	0′53″	0′52″
L	4.8	4.9	1′21″	1′20″	4.4	4.6	1′12″	1′15″
L−R	1.0	1.0	18″	18″	1.1	1.3	19″	23″

The results indicate the superiority of recitation as a factor in relearning but do not show that the material learned by the recitation method is better retained than the material memorized by reading alone. In the case of a few individuals, tests were made after various intervals by the 'Treffer' method. From the data given by Kühn (p. 463) the following averages for ten subjects have been computed:

TABLE VI

Per cent. of material recalled after an interval of from five to ten minutes. Test by 'Treffer' method

Method of learning	Lists of words		Lists of syllables	
	Tempo controlled	Tempo free	Tempo controlled	Tempo free
R	66.8	72.3	36.4	32.6
L	59.3	56.7	25.4	22.7
L−R	7.5	15.6	11.0	9.9

It is apparent that the material learned by recitation is better retained. In the case of words, material learned by means of a fixed tempo is not so well retained as material learned by a free tempo, but this result is not clear for nonsense material, for which the opposite, if anything, is true.

Three subjects were each given twelve tests by the 'Treffer' method first after ten minutes and again with other material after twenty-four hours. The results are given below.

TABLE VII

Per cent. of material recalled after ten-minute or twenty-four-hour lapses
Studying done with free tempo

After ten minutes		After twenty-four hours	
Reading	33.5	Reading	6.2
Recitation	44.1	Recitation	11.9

From this data, Kühn concludes "the advantage of learning with recitation for retention is much greater after a pause of a day than after a pause of a few minutes." [7]

Kühn found a great deal of interesting information concerning individual peculiarities in learning. Many individuals had a constant tendency to begin to recite too soon, or too late, for the best results; some were unable to limit themselves to 'pure' reading, more or less recitation unintentionally creeping in; some subjects were found who seemed to derive no benefit whatever from continued readings beyond a maximum of four or five; and others obtained better results under the reading method than when recitation was a factor. Kühn found the latter subjects to employ a peculiar form of 'automatic' or 'mechanical' method of learning, in which the usual method of building up associations between items and binding them into some form of compact 'schema' or structure was not employed. Simple visual imprinting was the most effective procedure.

Kühn came to the general conclusion, "that recitation is more effective because it leads to a more fundamental, many-sided working over of the material" (p. 443). In recitation the items are more attentively observed, the list is more carefully analysed, striking words are picked out, and a better 'schema' of reconstruction is employed. In the case of those individuals who rely upon the various sorts of associative aids in learning, recitation is very helpful, but the few who learn mechanically can do as well or better by merely reading.

[7] *Op. cit.*, p. 466.

So far no information has been cited with regard to the stage at which it is best to introduce the first recitation. This question was taken up and answered, in a measure, by Witasek.[8] Rows of ten pairs of nonsense syllables were exposed successively at the rate of one per second by means of a Wirth memory apparatus. The subject studied aloud both in reading and recitation, and all corrections and promptings were made orally by the experimenter.

TABLE VIII (from Witasek, p. 267)

Showing the superiority of a group of recitations over a group of readings, absolutely and relatively [9]

Number of preliminary readings	Superiority in number of repetitions		Superiority in number of seconds		Superiority in number of prompts	
	Absolutely	Relatively	Absolutely	Relatively	Absolutely	Relatively
Of five recitations over five readings						
6	2.3	90 per cent.	87	90 per cent.	4	80 per cent.
11	2.0	80 per cent.	73	80 per cent.	6	100 per cent.
16	1.4	70 per cent.	54	70 per cent.	3	60 per cent.
Of ten recitations over ten readings						
6	1.5	70 per cent.	62	70 per cent.	5	80 per cent.
11	2.0	60 per cent.	61	60 per cent.	3	60 per cent.
Of fifteen recitations over fifteen readings						
6	1.8	60 per cent.	62	60 per cent.	4	60 per cent.

In the case of recitation, however, the first syllables of the pairs were spoken by the experimenter. An interval of six to seven seconds was given between repetitions. Three series of syllables were learned each day, a three-minute rest being allowed between lessons. The method of computation used by Witasek is, in many cases, somewhat confusing. He frequently makes use of the term 'imprinting value' ('Einprägungswert') which means the value of

[8] 'Über Lesen und Rezitieren in ihrer Beziehungen zum Gedächtnis', *Zeitschrift für Psychologie*, 1907, **44**, pp. 161–185, 246–278.

[9] To make the meaning of the table clearer the procedure may be described more in detail. To begin with, the lists were read six, eleven, or sixteen times as indicated. One hour later, the lists were in one case read, and in the other case, recited, five, ten, or fifteen times as indicated, and immediately after the learning of the lists was completed by further recitations. From these data, the saving in the total time required to learn was computed for the groups of readings and for the groups of recitations. The table above presents, absolutely and relatively, the *differences* between the savings brought about by reading and by recitation—the differences always being in favor of the groups of recitations as shown.

a repetition in reducing the time or repetitions needed to complete the learning. Table VIII shows the superiority of a group of recitations over an equal group of readings, after a given number of preliminary readings, in reducing the time required after an interval of an hour, to complete the learning so that the entire list could be recited without prompts in ten seconds or less. The table is based upon the work of seven university graduates and faculty members, tested three or four times, a total of twenty-four tests.

The table shows in summary form a finding which is demonstrated in more detail by Witasek, *e. g.*, that the imprinting value of successive readings declines very rapidly after the first few. That

TABLE IX (from Witasek, pp. 184–185)

Number of original readings	Number of additional readings	Number of recitations	Total repetitions	Speed of recitations after one hour	Number of prompts
6	0	0	6	78″	7.9
6	5	0	11	75″	7.2
6	0	5	11	63″	6.3
6	0	10	16	69″	5.8
6	10	0	16	74″	7.5
6	5	5	16	66″	6.0
6	0	15	21	66″	5.5
6	15	0	21	73″	6.7
6	5	10	21	65″	5.9
6	10	5	21	66″	5.7
6	10	10	26	69″	5.7
6	5	15	26	65″	6.2

is to say, readings are pronouncedly subject to a law of diminishing returns. In this respect, recitations proved to be a better form of repetition. It is apparent, from the table, that recitations introduced into the learning at almost any point are of more value than continued readings. However, the superiority of recitation seems to be somewhat greater when introduced after six than when introduced after eleven or sixteen readings. This would seem to indicate that recitations, too, are subject decidedly to the law of diminishing returns. But Witasek explains that in these particular tests with a large number of repetitions "the readings unconsciously become very similar to recitations." The learner, finding the readings to become more and more fruitless, is unable to restrain a natural inclination to partially recite.

So far experiments have merely confirmed the current opinion that recitations, if not introduced too early in the learning process, are of more value than continued readings. It remains to enquire into the combination that will yield the richest returns in proportion to the outlay of time and energy. Table IX shows the relative effectiveness of several combinations as measured by the speed of the first recitation after an interval of an hour, together with the number of prompts. The procedure in this recitation was as follows: The first member of each pair was exposed, the subject responding with the second member, whereupon the first member of the next pair was exposed and so on. If the subject responded incorrectly, he was corrected by the experimenter, and if the subject could not

TABLE X (from Witasek, p. 184f)

Number of original readings	Number of additional readings	Number of recitations	Total repetitions	Speed of third recitation after one hour	Number of prompts
6	0	0	6	37″	1.8
6	5	0	11	34″	1.7
6	0	5	11	22″	0.8
6	0	10	16	20″	0.8
6	10	0	16	32″	1.6
6	5	5	16	19″	0.5
6	0	15	21	15″	0.5
6	15	0	21	27″	0.5
6	5	10	21	18″	0.7
6	10	5	21	19″	0.7
6	10	10	26	20″	0.9
6	5	15	26	16″	0.4

respond at all within ten seconds, the experimenter gave the response orally and exposed the next syllable in the series.

The results shown in this table are not very clean cut and in some respects are rather surprising. Within the various groups showing an equal number of repetitions, it is quite clear that a combination of recitations with readings leads to a more successful recitation after an hour than when reading only was employed in the study period. The advantage does not appear to be very great, however. What is quite surprising is that a small number of repetitions of any sort (six or eleven) leads to nearly as effective a recitation after an hour as a larger number (sixteen, twenty-one, or twenty-six). From this it would appear that repetitions beyond eleven are pretty largely wasted, and accordingly recitation, contrary to Witasek's earlier contention, must be subject also to a law of greatly diminishing returns.

The above table (Table X) which was computed from Witasek's original data, shows that the results for the third attempted recitation after the interval of an hour are quite different from those based on the first recitation after the interval. The second, fourth, or fifth recitation would have shown a similar difference.

In the case of these later repetitions, the advantage of recitation as a factor in the original learning is quite pronounced. It is apparent also, that although the law of diminishing returns is still seen to operate, its influence is very much less marked than appeared in the results for the first recitation after the interval.

Table XI exhibits the results in terms of the total time required to learn the series in two sittings separated by an hour.

TABLE XI (from Witasek, p. 274)

Work of the first sitting				Work of the second sitting			
Combination	Time of reading	Time of reciting	Pauses between repetitions	Time in recitations	Pauses between recitations	Sum with pauses	Sum without pauses
L6 R0	60	0	35	262	56	413	322
L11 R0	110	0	70	236	49	465	346
L16 R0	160	0	105	228	42	535	388
L21 R0	210	0	140	202	42	594	412
L6 R5	60	96	70	143	28	397	299
L6 R10	60	166	105	163	35	529	389
L6 R15	60	206	140	133	28	567	399
L11 R5	110	80	105	145	28	468	335
L11 R10	110	119	140	124	21	514	353
L11 R15	110	142	175	109	14	550	361
L16 R5	160	69	140	142	21	532	371
L16 R10	160	121	175	153	28	637	434

Table XII shows the data of Table XI rearranged, the combinations being arrayed in the order of their effectiveness with the percentages of time devoted to reading and to recitation.

The findings indicate that a small amount of work at the first sitting pays better than a large amount; that is to say, the series can be learned more quickly in the end if only a small proportion of the total time is devoted to the first study while the larger portion is saved for the review an hour later. There is considerable evidence that better results are obtained if the original study period is partly devoted to recitation; for example, 6 *Ls* plus 5 *Rs* gives better results than 11 *Ls;* 11 *Ls* plus 5 *Rs* is much better than 16 *Ls* and so on. However, the most potent factor is the distribution of the recitations. The best results are obtainable when the original period includes about twenty-five per cent. of the total learning

TABLE XII (based on Table XI)

Combination	First sitting		Second sitting		
	Per cent. of time for reading	Per cent. of time for recitation	Per cent. of time for recitation	Sum with pauses	Sum without pauses
L6 R5	24	33	43	397	299
L6 R0	23	0	77	413	322
L11 R0	39	0	61	465	346
L11 R5	39	24	37	468	335
L11 R10	35	37	28	514	353
L6 R10	18	45	37	529	389
L16 R5	50	20	30	532	371
L16 R0	50	0	50	535	388
L11 R15	33	45	22	550	361
L6 R15	17	55	28	567	399
L21 R0	59	0	41	594	412
L16 R10	42	30	28	637	434

time, allotting about half of this time each to reading and to recitation. Beyond this amount, recitations introduced into the review are much more effective than recitations in the first period of study.

On the whole, so far as the matter of the relative values of reading and recitation as factors in learning are concerned, Witasek's method of attack is subject to several defects. In the first place, too many variable factors are introduced. The influence of various divisions of the lesson, without regard to the methods of study employed, makes interpretation difficult. Moreover, as Witasek himself points out, the conditions of the experiment were such that the readings, especially after a number of perusals, became, unintentionally, very much like recitations. Witasek's procedure may also be charged with most of the defects found in Katzaroff's experiments; in reading, the subject prompted himself by looking at the forgotten syllable, but during recitation, promptings were made orally by the experimenter; the subjects were few and all were trained adults, whose habitual mode of studying may have been seriously interfered with by the particular conditions of the experiment; and finally practice effects were by no means fully eliminated.

From his study, Witasek drew the general conclusion that recitation, as compared to reading, is a much more effective method of study. The difference in favor of recitation was attributed in the main to a 'higher degree of attention' which was made possible by virtue of the opportunity afforded the subject to gauge his progress in the learning and apply himself to the portions that offered difficulty. The higher grade of attentiveness is closely correlated with an apparently 'greater activity' shown during recitation. In reading the subject is likely to relax into a state of passive receptivity, in recitation, the attitude is one of alert, searching (*'sich besinnen'*) activity.

In an experiment by Miss Abbott,[10] the problem has been attacked from a somewhat different point of view. Miss Abbott endeavored to determine the learning types of a limited number of individuals and to utilize this information in the interpretation of the numerical results. As material, lists of thirty nonsense syllables and sixty English words were used. An apparatus was provided such that the words or syllables could be exposed singly for any time desired. A fixed time (sixteen minutes) was allowed for the study period, this time being divided up into various combinations of reading and recall.

The groups of words and syllables were presented in various ways as shown in Table XIII.

TABLE XIII

Series	First eight minutes spent in	Exposure time per item	Interval between items	Interval between first and second learning period	Second eight minutes spent in
a	visual imprinting	1″	0	1′	visual imprinting
b	"	1″	0	15′	"
c	"	1″	0	45′	"
d	"	1″	0	1′	Recall
e	"	1″	0	15′	"
f	"	1″	0	45′	"
g	"	1″	0		
h	visual imprinting and recall	1″	3″	1′	visual imprinting and recall
i	visual imprinting and recall	1″	3″	15′	visual imprinting and recall
j	visual imprinting and recall	1″	3″	45′	visual imprinting and recall
k	visual imprinting and recall	1″	3″		

All series were allotted a sixteen-minute study period except series g and k which received but eight minutes. In series a, b, c, and g no opportunity is given for recall, the whole time being spent in 'Einprägung'; in series d, e, and f the first eight minutes is

[10] 'On the Analysis of the Factor of Recall in the Learning Process', *Psychological Review Monograph*, 1909, **11**, pp. 159–177.

spent in imprinting followed by eight minutes of recall; while in series h, i, and j, three-fourths of the time is devoted to recall, which is interspersed with the 'Einprägung' occupying the three-second intervals indicated in the table.

In all cases the subjects worked under certain restrictions. During the presentation of the material in series a, b, c, d, e, f, and g, the subject was not to form any associative links between the items and while one item was before him, he was not to think of another. During the recall period in the d, e, f series, while the subject was permitted to image the items and form such associations

TABLE XIV (from Abbott, p. 173)

Percentages of words and syllables correctly recalled after four hours

Series	a	b	c	d	e	f	g	h	i	j	k	Subject
Words	23	23	24	23	33	25	23	41	59	42	8	V
Syllables	42	38	40	35	40	17	22	68	92	70	47	
Words	8	8	13	29	12	8	5	97	98	97	62	W
Syllables	20	15	12	20	30	27	8	97	97	98	62	
Words	15	28	23	28	15	18	13	31	34	30	18	X
Syllables	13	13	22	22	17	18	5	30	27	48	10	
Words	67	78	83	48	42	43	28	53	55	17	23	Y
Syllables	67	53	70	67	63	50	57	63	67	53	43	
Words	54			61				56				Z
Syllables	60			50				55				

as he wished, he was not allowed to pronounce them or to write them down.

Five students of psychology acted as subjects in the tests. The image type of each was determined as follows: Subject *V* employed inner speech and concrete visual imagery; Subject *W* was of a motor-auditory verbal type; *X*, motor-auditory with some visual imagery; *Y* was strongly visual, never pronounces a word, just lets it 'soak in'; and *Z* was of mixed type, employing different kinds of imagery at different times.

The test of memory consisted in requiring the subjects to write, four hours after the test, all the words or syllables they could remember. Table XIV shows the results in the form of percentages of the total lists that were correctly reproduced.

Table XV was derived from Table XIV by subtracting the results obtained in series g, from the results obtained in series a, b, c, d, e, f, respectively. This table, consequently, presents the gain brought about by the second eight minutes of reading or recall as compared to the results obtained by the first eight minutes imprinting alone.

TABLE XV (from Abbott, p. 173)

Showing the advantage of sixteen minutes study over eight minutes

Series	a	b	c	d	e	f	Subject
Words	0	0	1	0	10	2	V
Syllables	20	16	18	13	18	—5	
Words	3	3	8	24	7	3	W
Syllables	12	7	4	12	22	19	
Words	2	15	10	15	2	5	X
Syllables	8	8	17	17	12	13	
Words	39	50	55	20	14	15	Y
Syllables	10	—4	13	10	6	—7	

Table XVI gives the results for words and syllables combined together with the averages for a, b, c; d, e, f; and h, i, j, respectively, based on the data from Table XIII.

TABLE XVI (from Abbott, p. 174)

Showing the combined results for words and syllables

Series	a	b	c	Average	d	e	f	Average	h	i	j	Average	g	k	Subject
	29	28	29	*29*	27	35	22	*29*	50	70	51	*56*	23	21	V
	12	11	12	*11*	26	18	14	*19*	97	98	97	*97*	6	62	W
	14	23	23	*20*	26	16	18	*20*	31	32	36	*35*	10	16	X
	67	70	79	*72*	54	49	46	*49*	57	59	29	*48*	38	30	Y

From Table XV it appears that with the exception of three cases the additional eight minutes of reading or recall results in a greater amount of material recalled, and for all subjects except Y, the value of the additional study is more pronounced in learning nonsense syllables than in learning words.

The most significant comparisons appear in Table XV. Subject *W*, of auditory-motor type, does much better in series h, i, j, than in d, e, f, which in turn gives better results than a, b, c. That is, the methods restricting learning to visual imprinting alone are the poorest of all; the method giving eight minutes of imprinting followed by eight minutes recall is much better; while the method giving three-fourths of the time to interspersed recall gives results about nine times as good as the first. Subjects *V* and *X* agree in showing sixteen minutes of visual imprinting to be as effective as eight minutes of imprinting followed by eight minutes of recall, but each shows to better advantage when three-fourths of the time is spent in interspersed recall (series h, i, j). These subjects ordinarily employed auditory-motor imagery or inner speech and were undoubtedly greatly hampered by some of the restrictions placed upon them in the d, e, f series. Subject *Y*, who possessed strong visual imagery, learning by simply allowing the items to 'soak in', does very well in the method of visual imprinting and very poorly in either method employing recall.

From this study, Miss Abbott draws the following conclusions:

1. That the factor of recall is always an aid in the learning process.

2. That when recall comes after the *Einprägung* of the material, immediate recall is of more value than delayed recall and its value decreases as the delay increases in length.

3. That the recall is of greater value when it is interspersed with the *Einprägung*.

4. That localization is one of the factors which go to make recall an aid to memory, but that the relative importance of this factor is determined by individual type.

5. That the relative value of recall and *Einprägung* depends on individual type.

To the present writer, it seems that the third conclusion, *e. g.*, "That recall is of greater value when it is interspersed with the Einprägung," is not entirely borne out, at least not in such form as to be applicable to every-day, non-restricted methods of learning. In the first place, the methods employing the interspersed recall devote twenty-five per cent. more time to it than do the methods in which the recalls are grouped. Again, the severe restrictions placed on the first eight minutes of learning by reading in the series a, b, c, d, e, and f are avoided in the series in which recall is interspersed with reading. It will be recalled that Katzaroff in experiments in which restrictions were less severe, and employing a larger number of subjects, found that recitations grouped gave better results than recitations interspersed with readings. (See p. 5.)

A study by Clemens Knors,[11] although not primarily concerned with the present problem, contributes some information concerning three different methods of memorizing paired material. *Method A* is similar to the 'reconstruction method' introduced by Miss

[11] 'Experimentelle Untersuchungen über den Lernprozess', *Archive f. d. g. Psychologie*, 1910, **17**, pp. 297–362.

Gamble.[12] The series was first read through once, the subject then attempted to recite both members of the pairs; the series was then read again, followed by another attempt at reconstruction and so on until learned. *Method B* was the same except that, in recitation, the first members of the pairs were exposed, the learner attempting to recite the second members only. In *Method C* the subject read and reread the series until he felt that they were mastered; whereupon he was tested by exposing the first members of the pair as in *Method B*. It will be noted that *Method A* permits the recitation of both members of the pairs, *Method B* of but the first member, while *Method C* permits no recitation during the learning.

In all methods the subject read or recited aloud, the number of perusals and the number of promptings being recorded. The scores are given in the form of the total amount of material that was read by the subject plus the amount supplied him in the form of promptings or corrections by the experimenter. The following sample will show how the score was computed. Suppose a series of fourteen syllables is learned by eight readings plus seven attempted recitations, during which fifty-one syllables were supplied by the experimenter. Then the total score would be eight (the number of readings) plus fifty-one (the total number of prompts) divided by fourteen (the number of syllables in the list). That is, score = $8 + 51/14 = 11.64$.

From the original data given by Knors, the results shown in Table XVII have been computed. Unfortunately Knors did not print all of the raw data that he collected, so that some of the tables are incomplete. The subjects are three adults (A, B, C) and four children (a, b, c, d) eleven to thirteen years of age. The table presents the average score of three or four tests for each individual.

Although the results are somewhat irregular, a few points can be made out. Sections *H* and *I* indicate that, for adults, *Method A*, which requires the recitation of both members of the pairs, is superior to *Method B*, in which but the second member is recited. For Subjects *A* and *C* the differences are very great. The same subjects, however, show but a slight superiority of *Method A* over *Method C* in which reading alone was involved.

Although the findings for the children are very irregular, some differences between the methods seem clearly to appear. When the series of nonsense syllables to be learned is long (Section L), *Methods A* and *B* are both superior to *C*, which permits reading only; but when the series is short (Section *M*) the differences are very small. The differences between *Methods A* and *B* in either

[12] 'A Study in Memorizing Various Material by the Reconstruction Method', *Psychological Review Monograph*, 1909, **10**, No. 4.

case are so small as to be negligible. In the case of senseful words, the advantage of *Method A* over *C* is very great, the two standing, for different individuals, in various ratios ranging from seven to five up to four to one. It appears in general, then, that children, as compared to adults, profit much more through the employment of recitation in learning.

However, but little reliability can be placed upon the scanty findings of these experiments. The number of subjects is too small and the quantitative results are too meager. The time of the various readings and recitations not being kept, there is some doubt whether any of the methods would show a distinct advantage with respect

TABLE XVII (from Knors)

Subjects	H Fourteen nonsense syllables. Methods			I Eighteen nonsense syllables			J Ten sense words			K Eighteen sense words		
	A	B	C	A	B	C	A	B	C	A	B	C
A	6.9	12.3		7.9	14.8		4.5		4.8	5.2		7.2
B	10.4	14.5		11.1	13.2		3.9		3.8	4.7		6.8
C	9.5	13.2		7.9	12.2		4.1		4.1	3.8		5.6
Subjects	**L Fourteen nonsense syllables. Methods**			**M Eight nonsense syllables**			**N Ten sense words**			**O Fourteen sense words**		
	A	B	C	A	B	C	A	B	C	A	B	C
a	17.9	12.2	23.7	14.1	8.9	12.9	5.7		7.3	6.5		12.5
b	11.1	10.8	29.3	9.1	11.3	12.0	4.3		10.1	5.9		19.5
c	11.1	13.3	30.7	7.8	8.7	13.1	5.0		19.2	6.1		22.4
d	10.5	13.6	19.6	10.6	9.9	11.2	6.0		11.5	7.8		17.8

to the total amount of time required to learn. Unfortunately Knors did not print all of his raw data, but from what does appear, it is clear that the variability of the performances, especially those of the children, is very great. Subject *A*, for example, in three tests of learning series of nonsense syllables by *Method A*, shows an average score of 9.0 with a P. E. of 2.2. On the whole, it would not be safe to consider Knors's results as more than suggestive.

SUMMARY OF RESULTS

Without doubt, this brief enumeration of the results of such a medley of experiments has left but a vague impression with regard to present status of the problem under consideration. Perhaps an effort to summarize the findings will assist somewhat to a better understanding. Such an effort, however, is fraught with difficulty. When one considers the individual differences possible among the subjects, the variations in materials, in method of presentation, and in the methods of scoring and the like, it can be readily understood that direct comparison of many studies is quite out of the question. Perhaps it will be worth while, first, to review the methods of attack employed in the several studies and, by throwing the differences into relief, pave the way for a concise summary of the outstanding results that will then be presented and for an understanding of the relation of the present study to those which have gone before.

1. *Differences as regards materials.* Most of the studies have employed the method of paired associates; as material, nonsense syllables in pairs, senseful words in pairs, digits paired with nonsense syllables, and foreign words paired with the vernacular have been used. Single series of senseless or senseful words of various lengths have also been used, and, in some cases, connected sense material such as prose or poetry. It is possible that the results might differ considerably according to the kind of material used; in fact, Kühn and Knors found that this was decidedly the case.

2. *Subjects.* With the exception of the few experiments with four boys, conducted by Knors, well educated adults have been employed. In nearly every case, moreover, the number of subjects has been entirely too small to eliminate differences which might be due to the influence of previously acquired habits of study as well as the more innate differences such as those considered by Miss Abbott.

3. *Methods of presenting the material.* In most cases, the material has been presented visually, but as was noted above (p. 5) sometimes the method of presentation changes within the lesson. Both Witasek and Katzaroff presented their material visually when the subject was reading, but during the recitation the material was presented orally. In some cases, the material is printed large, in some, small; in some it is held in the hand; in others, it is at a distance or thrown on a screen; sometimes the material is presented simultaneously, sometimes serially. The tempo of presentation is an important matter also. In nearly all cases, the tempo was controlled and varies greatly from experimenter to experimenter. Kühn presented syllables at the rate of one each 0.4 second, Abbott one per second, and Katzaroff one every two seconds. No one knows how closely these rates corresponded to the habitual tempo of the learner, and, what is more, as was shown by Kühn and as we shall see again later, the natural tempo of recitation is considerably

slower, on the average, than that of reading. In short, the results are affected somewhat by the rate of presentation, influence of which is likely to be different upon reading than upon recitation.

4. *Methods of reciting or recalling.* Aside from employing a fixed rate of presentation of the material, the recitation or recall often worked under other restrictions. For example, Miss Abbott in some tests restricted the learning to mere inner visualization of the data. In other cases, Knors for example, the subject was required to read and recite aloud. As a result of these various controls, the methods of learning became highly artificial; seldom was a subject permitted to study in the manner that he would spontaneously adopt, and too often the restrictions were not the same for recitation as for reading.

5. *Testing the learning and computing the results.* Sometimes the lesson consisted of a certain number of repetitions, in which cases the learning was never complete. The success attained might be measured by the rate of the next recitation following immediately or after an hour (Witasek), or simply by the amount of material that could then be reproduced immediately or after an interval, or by the time required to complete the learning then or later. Sometimes, a certain amount of time was given for study, and the amount that could be reproduced immediately or after an interval (Abbott) was taken as a measure of the learning. Sometimes, the assignment was learned at a sitting (Kühn), the score being based on the time or repetitions required to learn. Other things, such as the number of prompts required (Witasek), or the recitation time (Katzaroff), have been introduced as a measure of success. Add to these differences the highly ingenious yet anything but clean-cut methods of computation, such as those introduced by Witasek and Knors, and it is clear that to adequately compare the results of these studies one with another, is next to impossible.

Certain other sources of error, such as neglect of practice effects, fatigue, diurnal variations in efficiency add to the uncertainty. So it is only with all these differences and sources of error in mind that an attempt will be made to give a brief summary of the general status of the problem.

First. A predominance of evidence points to a greater effectiveness of recitation, compared to reading, as a factor in learning in the case of adults, at least.

Second. This rule holds true only after the learning has advanced somewhat by virtue of preliminary readings, but the exact point at which it is best to introduce recitation into the learning, or the optimum distribution of readings and recitations within the lesson, has not been satisfactorily demonstrated.

Third. The more reliable experiments, such as those of Kühn, indicate that the advantage of recitation over reading is greater in learning senseless, non-connected material than in learning senseful, connected material.

Fourth. The matter of the relative value of recitations grouped, as compared to recitations interspersed with the readings, is still an open question.

Fifth. No satisfactory evidence is at hand indicating that the general results found for adults will hold in the case of children of grammar or high school age and training.

Sixth. A considerable, but not thoroughly convincing amount of evidence indicates that the efficacy of the two methods of study depends entirely upon the learning or imagery type of the individual.

Seventh. The two broad functions, learning by reading and learning by recitation, have not as yet been adequately analyzed into their constituent functions.

III

THE SUBJECTS, MATERIALS, AND METHODS OF PROCEDURE

As was mentioned earlier, the present study was devised to answer a practical question of the school-room—namely, What are the relative values of learning by reading as compared to learning by recitation in the case of school children working under school conditions and with the ordinary school-room methods of attack? So far as practicable, everything was done to secure normal conditions for the work. The details concerning subjects, materials, methods of study, and computation of results will now be considered.

Experiments were conducted with adult subjects as well, the data from which will be used for comparative purposes and for purposes of determining more exactly the functions operative in the two methods of study. For the sake of convenience, the experiments upon adults will be described in a later section where the results are presented.

The Subjects

The subjects used were pupils of a grammar school of Oakland, California. The members of the first, fourth, sixth, and eighth grades acted as subjects for the experiments in which the nonsense syllables were used, and the third, fourth, fifth, sixth, and eighth grades for the tests with sense material. Each class consisted of from forty to forty-five pupils.

The school in which the experiments were conducted is situated in a residential suburb of Oakland and draws its pupils from the homes of business men and artisans of moderate means. In general the school stands in the first class.

As will be explained later in detail, the pupils were grouped by grades rather than by age for the tests. The following table summarizes the distribution of the members of the several grades according to age.

Materials Used

The materials were of two sorts, senseless, non-connected material and connected, sense material in the form of biographies. The nonsense syllables were constructed in a manner similar to that of

Age in years	6	7	8	9	10	11	12	13	14	15	16	17
Grade 1	24	13	4									
Grade 3		1	11	21	8	1						
Grade 4			5	16	11	4	1					
Grade 5				4	17	12	4	1	1			
Grade 6					2	7	13	9	2	1	1	1
Grade 8							1	8	13	12	7	

Müller and Schumann.[1] The sense material was constructed by the writer from material found in J. McKeen Cattell's *American Men of Science* and *Who's Who in America*. Samples are appended. While this material is senseful and connected, the organization of different parts of the whole is not so complete and systematic as would be generally found in poetry or prose, in which the ideas are more closely related and the material more closely unified by rhythms, accents, and natural pauses. This biographical form of material was used because it was desirable to approximate the kind of material that the pupils were accustomed to study in their regular history, geography, or grammar lessons.

The nonsense syllables were mimeographed in vertical columns on cards and were handed out one to each student. The sense material was mimeographed on sheets which were likewise distributed to the pupils.

Preliminary tests were conducted in order to determine the amount and difficulty of the material to include in the lesson as well as to give the subjects some preliminary practice in the tests before the actual experimentation began. The kind and amount of material was arranged so that the lesson was somewhat too large for the best students to master in the time allotted.

In the case of nonsense syllables, the series contained for the eighth grade sixteen syllables; for the sixth, fifteen; and for the fourth, fourteen. The pupils of the first grade were unable to read or write these syllables, so the teacher kindly constructed series of twelve syllables of a kind they were accustomed to manipulating, such as *ad*, *en*, *ig*, *op*, *ot*, etc. These syllables were written with a black crayon by the teacher on large strips of heavy paper.

The sense material was also arranged to suit the capacities of the different classes. For the eighth grade the biographies of five men served as a lesson; for the sixth and fifth grades, the same biographies for but four men were used. For the fourth grade easier biographies of four boys were used; while for the third grade, the biographies of three boys sufficed. Samples of the material are appended.

[1] Described by Meumann in *The Psychology of Learning*, pp. 365–368.

The following is a sample of the material used by the fifth, sixth, and eighth grades:

JAMES CHURCH, born in Michigan, February 15, 1869. Studied in Munich, and later studied Forestry and Agriculture. Director of Mt. Rose Weather Observatory in 1906. Studied evaporation of snow, water content, and frost.

JOHN CLARK, born in Indiana, June 4, 1867. Studied Surgery and became a doctor in Philadelphia. Taught at Johns Hopkins. Has visited Italy and Russia. Has a brother in Vancouver.

MORTON CLOVER, born in Ohio, April 25, 1875. Studied Chemistry at Michigan. Worked in Manila for eight years. Wrote articles on the content of dogwood, of sugar, and acids. Now lives in Detroit.

CLARENCE CORY, born in Indiana, September 4, 1872. Studied in Purdue and Cornell Universities. Now lives in Berkeley. Is Professor of Engineering and Dean of Mechanics. Since 1901 has been Consulting Engineer of San Francisco. Is a member of the British Institute.

GEORGE CURTIS, born in Massachusetts, July 10, 1872. Studied at Harvard on Geography. Won Gold Medals at Paris in 1900. Member of Boston Scientific Society. Went on the Dixie Expedition in 1902.

The following is a sample of the material used by the third and fourth grades:

HARRY, is 14 years old. His father is a farmer. Around the farm are red stones, black-berry bushes, red clay, green clover, and small trees. Harry is in the eighth grade, and is tall and slender. He likes dancing and singing.

JAMES, was born in June, 1905. He is going to be a carpenter. He can make a chair, a stool, a box, a gate, and a window. His mother has white hair and wears a black dress. His father is fifty-five years old.

HAROLD, was born in New York. He came to California when six years old. He is now fifteen years old and has a gun, a bicycle, a kite, a pair of skates, and a baseball suit. He is going to be a lawyer and live in Seattle.

FRED, was born in March, 1898. He lives on 31st and Parker Streets. He goes to business college. He is tall, has black hair and blue eyes, wears a gray suit and brown necktie. His home is made of brick and granite.

Since, as will be seen later, the same sort of tests were repeated several times, it was necessary to construct different texts equal in number to the tests given.[2] An attempt was made, of course, to make the various texts of equal difficulty, but as is usually the case, they probably vary considerably. That such differences in difficulty as may exist will not invalidate the results to any considerable extent, will be made clear later.

Method of Conducting the Tests

Several very conspicuous sources of error are to be contended with in experimental work of the present sort. That such errors have found their way into the work of previous investigators on this topic, has already been pointed out. The more important sources of error are as follows:

[2] In the case of nonsense material five tests and five texts were used; in the case of sense material six tests and an equal number of texts.

(*a*) *Practice effects.* In a series of five or six practice periods of from five to ten minutes each, it would be expected that practice effects would be considerable. Some of the earlier studies have not taken this sufficiently into account.

(*b*) *Unequal difficulty of texts.* Since one individual must repeat a similar test with many different texts, any inequality in their difficulty will affect the results. Even series of nonsense syllables may differ greatly in difficulty for different individuals.

(*c*) *Individual differences.* In the case of most of the earlier investigations the subjects were so few that individual peculiarities may have played a large role.

(*d*) *Diurnal variations in efficiency and fatigue.* It is imperative that comparative experiments should be conducted at the same hour of the day with subjects as nearly as possible in the same state of physical fitness, unless some adequate estimate of these influences be introduced as a check. In this respect nearly all of the earlier investigators have been negligent.

In order to eliminate, as far as possible, the effects of such sources of error, the method described below was employed in the work.[3]

A class, consisting of forty or more pupils on the average, was divided into a number of sections or squads,[4] the number of squads, for reasons which will be evident, being made equal to the number of methods of study that were tested. Each squad thus consisted of seven or eight pupils, the personnel remaining unchanged throughout. Different texts, of as nearly equal difficulty as possible, were of necessity used. A particular squad was tested but once on a single day, and to complete the series for each squad required five or six days. The accompanying table shows in detail the manner in which the tests were conducted. The procedure was as follows: At nine a. m. of the first day, squad one was given its first test under method one,[5] using text one. Immediately after, squad two studied the same text, according to method two; then squad three worked under method three and so on. On the next day, squad three was taken out at the first hour and studied text two according to method two; at the next hour squad four worked under method three with the same text and so on. Thus the squads progressed, during the five days, through all the trials, texts, methods, and hours. The outcome, as shown under the column indicated 'Total' is that from the point of view of the methods employed, which is the only factor with which we are concerned; all other influences are balanced or neutralized.

[3] The first grade was handled as a whole and not by squads as were the others. To be taken into new surroundings under the charge of a stranger proved to be too disturbing for these little children.

[4] Five for the learning of nonsense material, six for the learning of sense material.

[5] 'Method' refers to the manner in which the material was studied, see p. 30.

Differences in practice effects are neutralized because the sum total of practice for any one method is the same as for all others. Individual differences are neutralized because each subject has studied under each method, and no one more than once. The errors arising from differences in the difficulty of the texts are avoided, because each method has to its credit one group working with each of the six texts. The influences of diurnal variations in efficiency or

	Day 1	Day 2	Day 3	Day 4	Day 5	Total [6]
Method 1	Squad 1 Trial 1 Hour A Text 1	Squad 2 Trial 2 Hour E Text 2	Squad 3 Trial 3 Hour D Text 3	Squad 4 Trial 4 Hour C Text 4	Squad 5 Trial 5 Hour B Text 5	All squads All trials All hours All texts
Method 2	Squad 2 Trial 1 Hour B Text 1	Squad 3 Trial 2 Hour A Text 2	Squad 4 Trial 3 Hour E Text 3	Squad 5 Trial 4 Hour D Text 4	Squad 1 Trial 5 Hour C Text 5	All squads All trials All hours All texts
Method 3	Squad 3 Trial 1 Hour C Text 1	Squad 4 Trial 2 Hour B Text 2	Squad 5 Trial 3 Hour A Text 3	Squad 1 Trial 4 Hour E Text 4	Squad 2 Trial 5 Hour D Text 5	All squads All trials All hours All texts
Method 4	Squad 4 Trial 1 Hour D Text 1	Squad 5 Trial 2 Hour C Text 2	Squad 1 Trial 3 Hour B Text 3	Squad 2 Trial 4 Hour A Text 4	Squad 3 Trial 5 Hour E Text 5	All squads All trials All hours All texts
Method 5	Squad 5 Trial 1 Hour E Text 1	Squad 1 Trial 2 Hour D Text 2	Squad 2 Trial 3 Hour C Text 3	Squad 3 Trial 4 Hour B Text 4	Squad 4 Trial 5 Hour A Text 5	All squads All trials All hours All texts

fatigue are neutralized, since each method has been tried by one squad working at each of the different hours.

Almost ideal arrangements were made for conducting the tests.[7] In a well lighted and well ventilated room about twenty-two by fourteen feet in size, a library table large enough to seat about a dozen people was provided. The situation of the room was such that practically all noise and distractions of whatever kind were avoided. Care was taken to keep the physical conditions of the room as constant and comfortable as possible. Fresh air was kept in circulation, an abundance of light was admitted, and the temperature was kept constantly between fifty-seven and sixty degrees Fahrenheit.

[6] In the case of sense material, six methods, squads, texts, etc., were used instead of five.

[7] For this I am greatly indebted to the school principal, Mr. N. Ricciardi.

Since there were but seven or eight pupils undergoing a test at a time, the experimenter who stood at the head of the table could easily keep an eye on the work of each individual. Any attempt on the part of a pupil to copy from another, to loaf, or use improper methods of any sort, could be instantly detected. Such policing was quite unnecessary and such violations of rules as did occur were in most cases unintentional. However, such factors which might result in the unreliability of the data were urgently sought, and in cases where such an unreliability was known or suspected, the entire data of that child were thrown out. In addition to the observations of the writer, the opinion of the teacher, especially with reference to doubtful cases, was sought and freely obtained. Each teacher listed the pupils in her room according to the following request, "Please list your estimates of the intelligence of the pupils in your room, in order of rank, putting the most intelligent as Number one, ——. Use your own methods of estimating and your own conception of what intelligence is. Please do not, however, make it a mere record of class standing according to grades received, and mere maturity should not be considered." The teachers also fulfilled a request to give the names "of such pupils that you think on account of feeble intelligence or inattentiveness, lack of persistence, indolence or inclination toward dishonesty in work, etc., would be unreliable subjects for experimental purposes." The teachers were consulted also in particular cases when the occasion arose.

As a result of these precautionary measures the work of a few pupils was discarded. The following were the chief factors which seemed to justify discarding a pupil's data:

First. Absence from one of the tests. In case a pupil missed one or more of the tests, his entire work was discarded. This was necessary because in succeeding tests he would be one or more stages behind in practice. The absentees on return were allowed to continue the work without being told that the data would not be used, as a precaution against creating any ill feeling among the pupils.

Second. Copying from others or using unfair methods of any sort. Intentional or unintentional disregard of rules was very rare.

Third. Lack of interest or loafing. Occasionally a pupil from lack of interest or less worthy motives, felt inclined to be balky or to 'quit' for a moment in the midst of a test. The data of such were discarded.

Fourth. Mental defectives. A few pupils were found to be markedly below the average in the test work. Consultation with the teacher confirmed the suspicion of sub-normality and the data of such were discarded, although they went through the work with the other pupils.

Fifth. Physical defectives. Bad cases of eye defects, weakness from previous illness, and school-yard accidents occasionally interfered with maximal performance to such an extent that the data were excluded.

On the whole, however, such cases were very rare and the spirit among the pupils was of the finest. A keen spirit of competition arose with reference both to an individual's own previous record and to the records of other individuals, such that in nearly every case the results were the products of the pupils' very best endeavors. The number of pupils who completed all of the tests in a satisfactory manner ranged from thirty-seven to forty-one in the various grades.

Methods of Studying

A single squad having been seated at the table in the separate room, a copy of the material was passed out face downward before each pupil, and the following instructions were given: "On each of these cards is a list of nonsense words [show a sample]. They are called nonsense words because in English they have no meaning. Now the object of the test today is to see how many of these words you can learn in a certain short time.

"We will proceed like this. I will give you two signals to start. At 'Ready' you take the card at the corner like this and at 'Go' you turn the card over and begin to study.

"Now you are going to study for a while in one way and then later you are going to study in a very different way. To begin with you are to study by reading this list of words over and over from beginning to end [illustrate]. Remember you are to read only. You should never look away from the paper; never close your eyes to see if you can say the words; in fact never say a single word unless you are actually looking at it, actually reading it. Remember you are to read through from the first to the last every time.

"After you have read the words through and through in this way for a while, I am going to give you a signal 'Recite'. When I say 'Recite' you are to hold your paper in front of you so that when you are looking straight ahead, you look over the top of it and you can see it by glancing downward a little like this. Now you are to try to say to yourselves as many of the syllables as you can without looking at the card. When you cannot remember the next word look down at your card and then go on saying as many of them as possible without looking. Glance at the card again whenever you cannot remember. Go through the list from the first word to the last in this way and continue until the word 'Time' is given. Remember you are not to look at the words unless you absolutely have to.

"When the learning period is over I am going to ask you to write as many of these words as you can."

It should be remembered that every class had received previous practice in the learning. The first grade had been given two trial tests of five minutes each, and every other grade one or two trials of eight minutes each, the data from which were not used.

Following is a table showing the absolute and relative amounts of time devoted to reading and to recitation in each method.

NONSENSE MATERIAL

Grade one

Method	Time of reading	Time of recitation	Per cent. reading	Per cent. recitation
1	5′	0′	100	0
2	4′	1′	80	20
3	3′	2′	60	40
4	2′	3′	40	60
5	1′	4′	20	80

Grades four, six, and eight

1	9′	0	100	0
2	7′12″	1′48″	80	20
3	5′24″	3′36″	60	40
4	3′36″	5′24″	40	60
5	1′48″	7′12″	20	80

SENSE MATERIAL

Grade three

Method	Time of reading	Time of recitation	Per cent. reading	Per cent. recitation
1	7′30″	0	100	0
2	6′	1′30″	80	20
3	4′30″	3′	60	40
4	3′	4′30″	40	60
5	1′30″	6′	20	80
6	45″	6′45″	10	90

Grades four, five, six, and eight

1	9′	0	100	0
2	7′12″	1′48″	80	20
3	5′24″	3′36″	60	40
4	3′36″	5′24″	40	60
5	1′48″	7′12″	20	80
6	54″	8 06″	10	90

The study period was made somewhat shorter for the first and third grades, because it was found that steady application for longer periods was quite fatiguing.

At the end of each study period the pupils promptly placed the text papers face downward and began at once to write the material upon sheets that were provided. They were instructed to give the material in the original order as far as possible. In the case of nonsense syllables, the recall was pure reproduction, but when the sense material was used, the names of the individuals whose biographies were studied were written on the board in proper order. This was the only aid that was given. Ample time was allotted in which to write the material remembered.

Three or four hours later, tests for retention were given. The test consisted in simply asking the pupils to write, as before, all the material they could remember. No aids were given except that the names, in the case of sense material, were written on the board as in the immediate test.

Notes were kept of all manifestations of the children's work such as movements of the lips, whisperings, rhythmical movements of the head, or hands or feet, tappings of the fingers, directions of the gaze, etc., in fact, of all appearances which might be of later service in interpreting the results. The judgments of the pupils were frequently called for upon such matters as the methods which they liked or disliked, why the nonsense syllables were hard to learn and the like. These will be dealt with later.

The method thus far described applies only to the work with the school children. Different methods were employed upon the adult subjects and they can most conveniently be described on a later page where the results are presented.

TREATMENT OF THE DATA

Method of scoring the nonsense syllables. The nonsense syllables were scored by giving three points for a syllable correct in form and position; two points for a syllable correct in position with one letter incorrect; two points for a syllable correct in form but not in correct position; one point for a syllable with two letters correct but in wrong position. For example:

Correct list	Reproduced list	
pib	pib	= 3
bah	dah	= 2
rem	bug	= 1
lor	rem	= 2
cug	lag	= 0

Although more exact methods of scoring nonsense syllables are available, it was thought that the additional precision that might be obtained by their use would scarcely justify the additional labor involved.

All of the nonsense syllables were scored by a person who had no acquaintance with the nature of the experiment. In order to test the personal equation as manifest in the scoring, a set of forty lists were graded by two individuals, neither being aware that the lists were to be, or had been, corrected by another. The variations were found to be very small and due to variable errors so that the averages were about the same. The average score for forty papers was for one grader 22.81, for the other 22.88. From these figures the P. E. was computed by means of the formula $\text{P. E.} = .84435 \frac{\text{A. D. dis.}}{\sqrt{n}}$

The P. E. thus determined is 0.021. The personal factor involved in the scoring of results is thus too small to be of significance.

The material of the first grade pupils (two letter syllables) was graded by simply counting the number of syllables that were correct in spelling. These children had had but little experience in writing on paper and as a consequence their syllables were mixed up so badly that it was impossible in many cases to be sure what order was intended. Consequently, correctness in form, only, was considered.

Method of scoring the sense material. The sense material was scored by dividing the original texts into details, ideas, or facts that were mentioned, to serve as a guide. One credit was given for the correct reproduction of each of these 'details' when they fell under the proper name. When a detail, such as a birthplace, was correctly reproduced but applied to the wrong person, one-half a unit was given. In some cases the credits of one-half or three-fourths were given to details or facts partly correct, depending upon the judgment of the reader.

Part of the sense material was scored by one individual and part by another, neither of whom was acquainted with the experiments in general. To test the reliability of the judgments, forty papers were scored independently by each. Variations of small magnitude but greater than for the nonsense material were found, but these were due to variable errors that compensated each other in the long run, producing on an average of forty scores very slight differences. The P. E., computed as above, is 0.015. This P. E. is so small in comparison with the P. E.'s of the averages that it has not been taken into consideration in the final computations of the results.

METHODS OF COMPUTING THE RESULTS

The results show the average scores based upon the methods of grading just mentioned. To be more accurate, the tables show a grand average of the averages of the several squads for each method of study. The work of several pupils, for various reasons that have been cited, was rejected, with the result that the final number of individuals in some squads is greater than that in others. Since, from the point of view of any particular method, the practice effects of each squad differed from every other, to permit the results of a squad to enter the final average with full weight, would distort the figures in a degree amounting to the average difference in efficiency due to the greater amount of practice of the one over the other. This overweighting was avoided by averaging each squad separately and then making an average of these figures.

For the same reason the P. E.'s could not be computed in the regular manner but must be based upon the results of the individual squads. Assuming that the averages of the several squads would be equal except for differences due to practice, fatigue, and diurnal variations, the deviations of the figures within each squad from the average of that squad were computed. A sum of the deviations for all individuals from the average of their squad was thus obtained and divided by the total number of individuals in the class, thus giving the Average Deviation. The P. E.'s were then computed according to the formula: [8]

$$\text{P. E. tr. av.} - \text{obt. av.} = 0.8453 \frac{\text{A. D. dis.}}{\sqrt{n}}$$

[8] See Thorndike, E. L., *Mental and Social Measurements*, New York, 1912, pp. 186 ff.

IV

QUANTITATIVE RESULTS

It was pointed out earlier that the amount of material given as a lesson was slightly greater than the best students could learn in the time allotted. Learning was never complete, although in the case of many individuals it was nearly so. With nonsense syllables as material, the average scores for the best methods are for different classes from fifty to seventy-three per cent. of the highest possible score. For the sense material, the best average scores are in the neighborhood of forty per cent. of the highest possible scores. This fact should be kept in mind during the consideration of the results which follow. For convenience of expression, we shall speak of 'methods' in which there was a 'combination of twenty per cent. reading with eighty per cent. recitation', etc., but it must be remembered that such expressions have a strictly local meaning, for several reasons. In the first place, such 'combinations' lead only to partial learning of the data. Perhaps the same combination would lead to very different results if applied to the time required to completely learn the lesson. A second consideration is that a 'combination' has reference only to the particular kind and the particular amount of material here used. The optimum combination would doubtless be different according to the difficulty and length of the lesson. These matters will be given more consideration on a later page.

RESULTS FOR THE LEARNING OF NONSENSE SYLLABLES BY CHILDREN

Table XVII shows the results of the immediate test for nonsense syllables in the form of average with P. E.'s computed in the manner described in the previous chapter. Table XVIII shows the same data transformed to relative scores in which the average of each class for all five methods is considered 100, serving as a basis for the other scores. The P. E.'s were changed to correspond. Figure 1 shows graphically the data of Table XVIII, the average being denoted by the heavy line, which is enclosed within two light lines representing on either side the area including the P. E.[1]

For the fourth, sixth, and eighth grades the results are clear. The results for Grade one were a disappointment and should be considered

[1] After the manner originally suggested by Professor J. McKeen Cattell.

apart from the others. The averages for this grade seem to indicate that Method Five, in which the amount of recitation is greatest, produces the poorest results while the methods involving more reading show to better advantage. The P. E.'s, however, show the averages not to be highly reliable and their significance is slight. One reason for this may lie in the fact that a less refined method was used in the case of this grade (see p. 27). In all probability, moreover, the inexperience of these beginning pupils accounts for the results in a large measure. They were simply unable to adjust themselves to

TABLE XVII

Showing the average score for each grade for the various methods of study

Method		1	2	3	4	5
Combination in mins. and secs.		L9' R0	L7'12" R1'48"	L5'24" R3'36"	L3'36" R5'24"	L1'48" R7'12"
Combination in per cent.		L100 R0	L80 R20	L60 R40	L40 R60	L20 R80
Grade eight	Average score	16.92	23.86	25.79	27.28	35.51
	P. E.	0.61	0.69	0.65	0.66	0.86
Grade six	Average score	13.21	20.18	22.64	25.15	30.52
	P. E.	0.61	0.84	0.60	0.91	1.07
Grade four	Average score	9.45	12.00	16.10	16.95	20.93
	P. E.	0.57	0.46	0.56	0.75	0.79
Combination in mins. and secs.		L5' R0	L4'R1'	L3'R2'	L2'R3'	L1'R4'
Combination in per cent.		L100 R0	L80 R20	L60 R40	L40 R60	L20 R80
Grade one	Average score	6.2	6.1	6.2	5.6	4.7
	P. E.	0.22	0.27	0.20	0.20	0.21

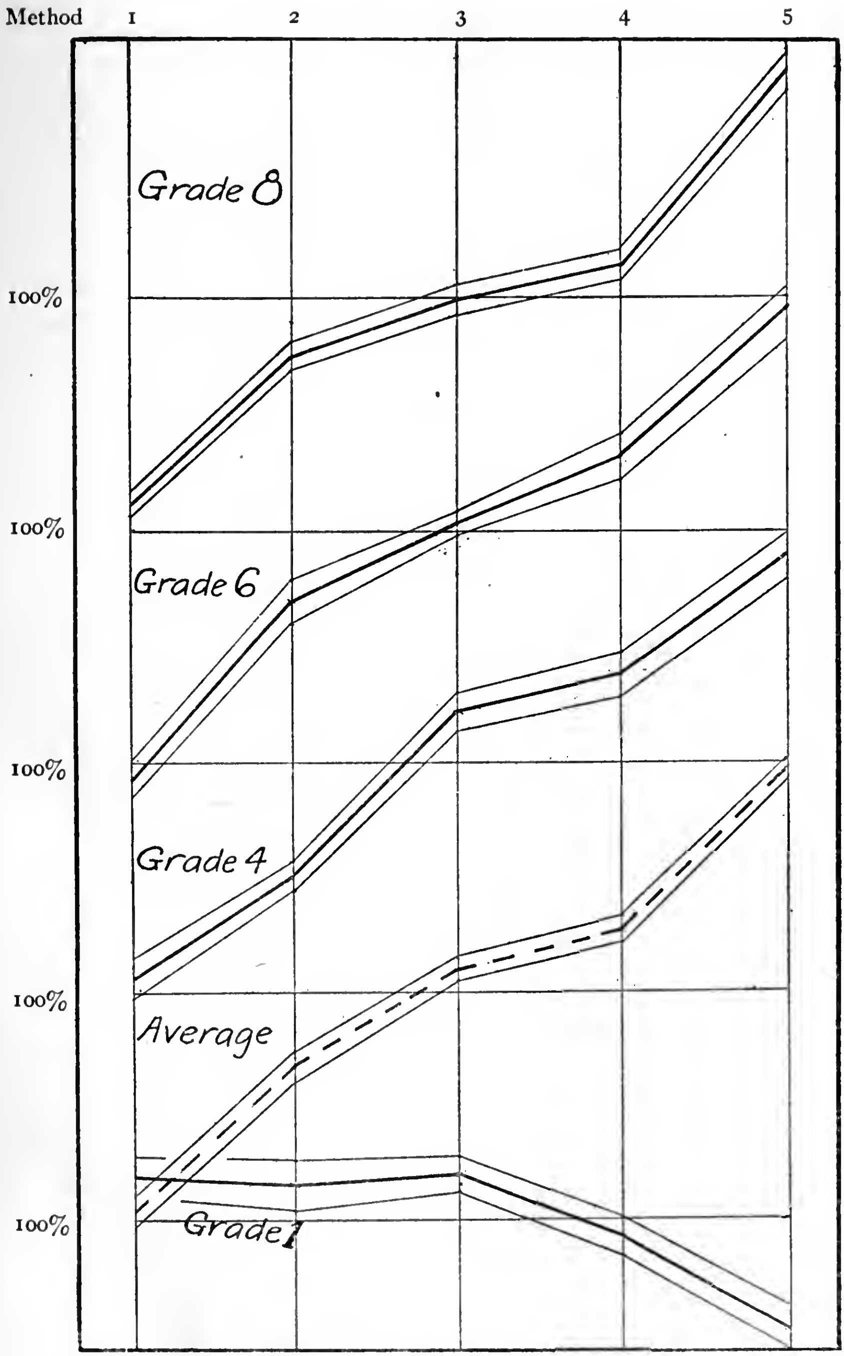
TABLE XVIII

Showing the data of Table XVII on a relative basis

Method		1	2	3	4	5
Grade eight	Relative score	65.40	92.23	99.69	105.45	137.26
	P. E.	2.37	2.69	2.53	2.57	3.35
Grade six	Relative score	59.13	88.35	101.34	112.57	136.61
	P. E.	2.74	3.78	2.70	4.09	4.81
Grade four	Relative score	63.42	80.53	108.05	113.75	134.42
	P. E.	3.42	2.76	3.36	4.50	4.74
Grade one	Relative score	107.64	105.90	107.64	97.22	81.59
	P. E.	3.80	4.67	3.46	3.46	3.63

FIGURE 1

Based on the data of Tables XVIII and XIX



the experiment. Many employed practically the same methods of study throughout, as could be seen from observation in some cases. Others made an effort to follow the prescribed directions which often resulted in poor records, especially in the methods in which recitation began very early. A great deal of time was lost in fruitless endeavor to recall syllables that were not as yet forthcoming. These young children were not skilled enough in testing their knowledge and prompting themselves where needed, which probably accounts for the apparent inferiority of the results obtained in Methods Four and Five, in which the reading periods were short.

For the fifth, sixth and eighth grades the results are convincing. The averages show a very great superiority of Method Five, in which the most recitation is introduced, over Method One which employs reading only. Twice as much is learned by the former as by the latter method in the same time. The small P. E.'s indicate a high reliability of these extreme differences, as well as the fact that the same thing is shown by all three classes. The graphs (Figure 1) show that there is an increase in the amount learned as the relative amount of recitation becomes larger, a fact shown by all three classes. The amount of this increase is not constant, however, being marked by a particularly great difference between Method One which permits no recitation and Method Two which gave twenty per cent. of the time to recitation. The effect of a minute and a half of recitation is very marked. The increase in effectiveness is fairly constant from Method Two to Three to Four, but the step from Four to Five is somewhat greater than any one of these. The most probable explanation for this exceptional score in the case of Method Five is that it was usually productive of a little more enthusiasm than other methods. The children anticipated this as the 'record breaking' method.

The reliability of the differences between the methods has been computed in a different way, as shown in Table XIX and displayed graphically by the broken line curve in Figure 1. This table shows the averages of the three grades (eighth, sixth, fifth) for each method with the P. E.'s of the averages computed according to the formula:[2]

$$\text{P. E. tr. av.} - \text{obt. av.} = .6745 \frac{\sigma \text{ dis.}}{\sqrt{n}}$$

The P. E.'s should be magnified to some extent in this table for the reason that the number of cases is very small (three) and that influence of any factor tending to create differences between the groups considered, for example, the effects of maturity or length of

[2] See Thorndike, E. L., *Mental and Social Measurements*, p. 188 ff.

school training, would make the P. E.'s larger. However, the P. E.'s are still very small.

TABLE XIX

Showing the average percentile scores with P. E.'s for Grades four, five, and eight

Method	1	2	3	4	5
Average score	62.65	87.04	103.59	110.59	136.09
P. E.	1.01	1.88	1.40	1.45	0.45

The following table shows more plainly the differences between the various methods and the P. E.'s of those differences, the computation being based on the preceding table. The formula employed for obtaining the P. E. of the differences is:[3]

$$\text{P. E. diff.} = \sqrt{(\text{P. E. av.})^2 + (\text{P. E. av.})^2}$$

TABLE XX

Showing the differences of the various methods in percentages with the P. E.'s of the differences

Differences of methods	Differences of methods	Differences of methods
2—1 = 24.39 ± P. E. 2.11	3—2 = 15.99 ± P. E. 2.34	4—3 = 7.56 ± P. E. 2.01
3—1 = 40.38 ± P. E. 1.72	4—2 = 23.55 ± P. E. 2.37	5—3 = 33.06 ± P. E. 1.47
4—1 = 47.94 ± P. E. 1.76	5—2 = 49.05 ± P. E. 1.93	5—4 = 25.50 ± P. E. 1.52
5—1 = 73.44 ± P. E. 1.10		

The differences are all conspicuous and reliable.

Differences in results among classes

A glance at Figure 1 will show that the findings for Grades four, six, and eight are very similar. In all grades Method Five is about twice as good as Method One. About the only difference is that the fourth grade does not do well, relatively, with very short periods of recitation. The difference in percentages[4] of Method Two (1′ 48″ Recitation) over Method One (all reading) is for the eighth grade 26.83 per cent., for the sixth grade 29.22 per cent., for the fourth grade 17.11 per cent. Grade four shows the slightest superiority of Method Two over Method One, but its value is rendered somewhat doubtful since the sixth grade shows a slightly greater superiority than does Grade eight. Computing the superiority of the average score from all Methods over the score of Method

[3] See Thorndike E. L., *op. cit.* [4] Based on Table XVIII.

Two, the results are: Grade eight, a superiority of 7.73 per cent.; Grade six, 11.12 per cent., and Grade four, 19.47 per cent. From these figures it appears that the older children are able to do better, relatively, with the short recitation periods.

Summary

To summarize, then, it may be stated: (*First*) that for the learning of nonsense material by children, recitation after a few initial readings is of much greater value than continued readings. (*Second*) That after preliminary readings for 1′ 48″, the more quickly the attempts at recitation are introduced, the better results will be obtained. (*Third*) No conspicuous differences appear in the results for the different classes with the exception of Grade one. (*Fourth*) which for reasons mentioned on p. 27 must be treated as a distinct case.

RESULTS FOR SENSE MATERIAL

Table XXI shows the results of the immediate test for sense material in the form of average scores. Table XXII shows the same data on a relative basis in which the average score for all methods

TABLE XXI

Showing the average score for each grade for the various methods of study [5]

Method		1	2	3	4	5	6
Combination in mins. and secs.		L9 Ro	L7′12″ R1′48″	L5′24″ R3′36″	L3′36″ R5′24″	L1′48″ R7′12″	L54″ R8′06″
Combination in per cent.		L100 Ro	L80 R20	L60 R40	L40 R60	L20 R80	L10 R90
Grade eight	Average score	20.77	22.39	24.84	24.95	25.28	23.75
	P. E.	0.72	0.87	0.70	0.69	0.50	0.82
Grade six	Average score	15.13	16.55	18.01	17.70	17.77	16.63
	P. E.	0.75	0.59	0.69	0.68	0.82	0.68
Grade five	Average score	11.79	13 95	15 21	15.96	15.33	15.74
	P. E.	0.40	0.43	0.48	0.56	0.50	0.55
Grade four	Average score	14.61	16.91	16.36	18.81	17.62	17.20
	P. E.	0.77	0.78	0.86	0.77	0.70	0.71
Combination in mins. and secs.		L7′30″ Ro	L6′ R1′30″	L4′30″ R3′	L3′ R4′30″	L1′30″ R6′	L45″ R6′45″
Combination in per cent.		L100 Ro	L80 R20	L60 R40	L40 R60	L20 R80	L10 R90
Grade three	Average score	8.66	10.34	11.18	14.12	13.10	12.09
	P. E.	0.39	0.49	0.49	0.46	0.56	0.54

[5] The highest possible score, approximately, is for Grade eight, 60; for Grades six, five and four. 48; and for Grade three, 36.

for that grade equals 100. The P. E.'s are computed as described above (p. 34). Figure 2 shows graphically the data of Tables XXII and XXIII.

TABLE XXII

Showing the data of Table XXI on a relative basis

Method		1	2	3	4	5	6
Grade eight	Relative score	87.78	94.62	104.98	105.45	106.80	100.03
	P. E.	3.01	3.64	2.93	2.89	2.09	3.43
Grade six	Relative score	89.21	97.58	106.19	104.36	104.77	98.06
	P. E.	4.42	3.48	4.09	4.01	4.83	4.01
Grade five	Relative score	80.42	95.15	103.75	108.86	104.57	107.36
	P. E.	2.72	2.93	3.27	3.81	3.41	3.75
Grade four	Relative score	86.34	99.94	96.69	111.17	104.13	101.65
	P. E.	4.54	4.60	5.07	4.54	4.13	4.18
Grade three	Relative score	74.78	89.29	96.54	121.93	113.12	104.40
	P. E.	3.35	4.21	4.21	3.95	4.81	4.64

A glance will show that the results here obtained differ from those received with nonsense material. In general the advantage of reading with recitation as compared to reading alone is less great. Moreover it appears that introducing the recitation too early proves to be of no value; in fact, for the lower grades it may prove to be a positive hindrance. This point will be taken up later. All grades agree in showing reading alone to be a poor method of study, while a combination of forty per cent. reading with sixty per cent. recitation seems to give best results.

The following table (XXIII) shows the average results for all classes combined, with the P. E.; the methods of computation being the same as those previously described.

TABLE XXIII

Showing the average percentile (relative) score for all grades combined

Method	1	2	3	4	5	6
Relative score	83.71	95.32	101.63	110.35	106.67	102.30
P. E.	1.64	0.99	1.26	1.90	1.01	0.74

In the average results, Method Four seems to be distinctly superior to Method Three and possibly superior to Methods Five and Six. In order to more accurately determine the reliability of the differences between the methods, Table XXIV was computed after the method earlier described (p. 38) using the data from Table XXIII.

TABLE XXIV

Showing the differences between the relative scores for the several methods with the P. E. of the differences

Differences of methods	Differences of methods	Differences of methods
2—1 = 12.61 ± P. E. 1.91	3—2 = 6.31 ± P. E. 1.60	4—3 = 8.72 ± P. E. 2.27
3—1 = 17.92 ± P. E. 2.06	4—2 = 15.03 ± P. E. 2.14	5—3 = 5.04 ± P. E. 1.61
4—1 = 26.64 ± P. E. 2.50	5—2 = 11.35 ± P. E. 1.41	6—3 = 0.67 ± P. E. 1.46
5—1 = 22.96 ± P. E. 1.92	6—2 = 6.98 ± P. E. 1.23	5—4 = 3.68 ± P. E. 2.15
6—1 = 18.59 ± P. E. 1.79		6—4 = 8.05 ± P. E. 2.03
		6—5 = 4.37 ± P. E. 1.25

This table shows, that for the average results, every method is clearly superior to Method One (all reading), the smallest difference, that between Methods Two and One, being more than six times the P. E.[6] It is also certain that every method except Method One is superior to Method Two, the smallest superiority being four times the P. E. The difference between Method Three and Four is also quite reliable, being four times the P. E. The superiority of Method Five over Method Three is more than three times the P. E.; while there is no real difference between Methods Six and Three. There is no evidence that Method Four is superior to Method Five, but Four is superior to Six, and Five is also superior to Six by small but reliable differences.

We are safe in concluding then, that in general, Method One, which includes no recitation, is the poorest method, while Method Four or Five is the best. Method Two is considerably superior to Method One and Method Three is better than Two. That is to say, the best results are obtained when the recitation is introduced after one and one-half to three and one-half minutes of preliminary reading. Beginning earlier or later than this leads to poorer results.

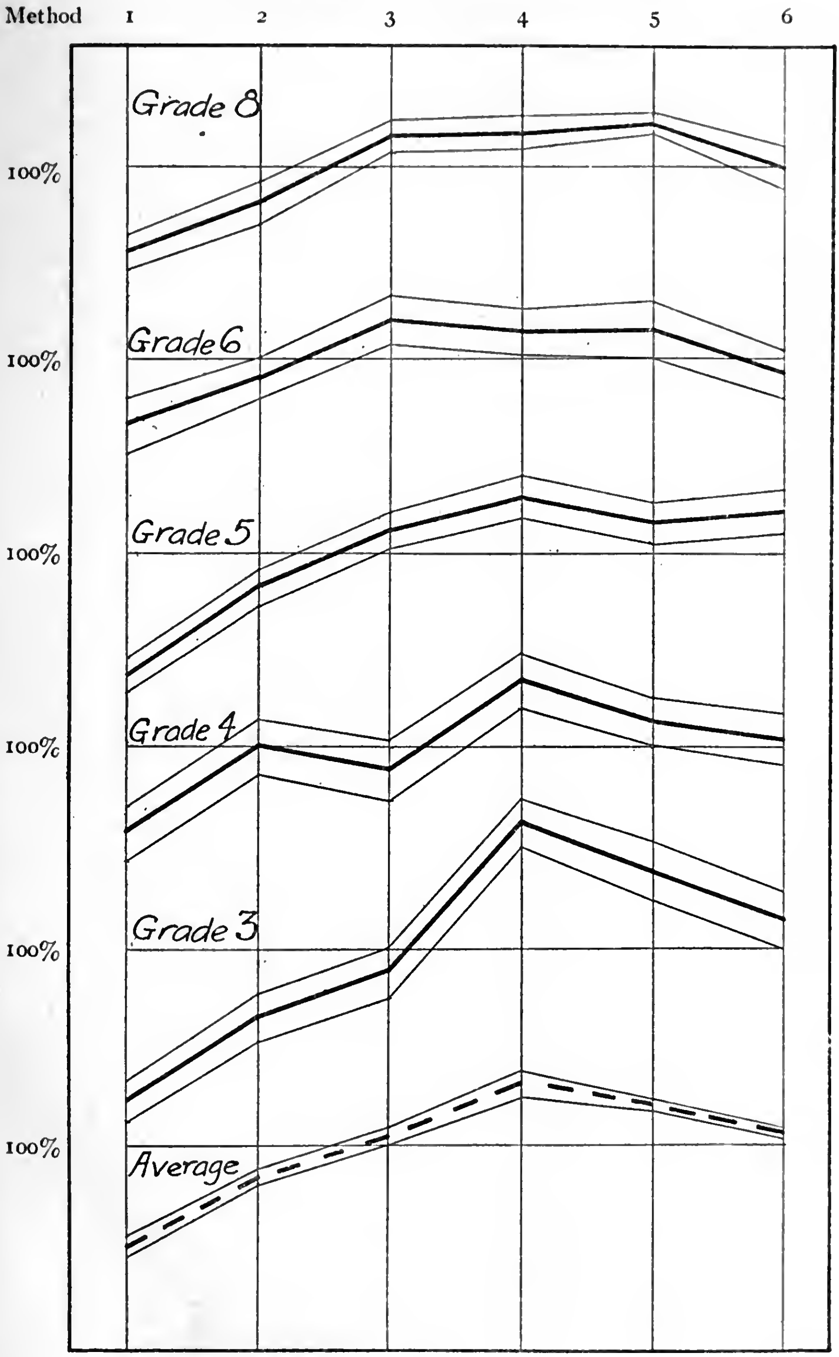
Differences in results among the classes

Such are the findings in general, but it was noted earlier that the classes differ in certain respects. These differences appear quite clearly in Figure 2. In the first place, the difference between the

[6] It should be repeated that the P. E. should be very large for the reason that all the apparent differences in the curves for the various classes (see Figure 2) result in attenuation of the P. E.

FIGURE 2

Based on the data of Tables XXII and XXIII



best and the poorest method is greater for the lower grades. *The superiority of the best over the poorest method is:*

For grade eight	19.02 per cent.	Average for 8 + 6 = 18.00
For grade six	16.98 per cent.	
For grade five	28.44 per cent.	
For grade four	24.83 per cent.	Average for 3 + 4 = 35.99
For grade three	47.15 per cent.	

The differences do not increase uniformly with the grades, but if we average the differences for the eighth and sixth grades, also for the third and fourth, the latter figure is exactly twice the former, while the fifth grade lies midway. The older children are doubtless not so badly handicapped by the lack of an opportunity to recite; or, stated in another way, the younger children are more dependent upon the factors involved in recitation in their learning.

Another difference also appears. Optimum results may be obtained by introducing recitation earlier in the period in the case of the upper grades. For Grades five, six, and eight the differences between Methods Three, Four, Five, and Six are nil or unreliably small, but in the case of Grade four the difference between Methods Six and Four is 9.52 per cent. ± P. E. 6.08; the same difference for grade three being 17.53 per cent. ± P. E. 6.0. The introduction of the recitation period too early has a deleterious effect upon the learning of the two lower grades. The probable explanation of this difference between the grades is to be found in the better adaptation of the older or more experienced learner to the conditions of the test. When recitation is introduced too early, the younger pupils waste time and energy in fruitless endeavor to recall the material. Positive errors of recall are probably numerous also and thus retard the learning. The older pupils, on the other hand, realizing that so early an attempt at recitation would be unprofitable, continue for some time to read, or divide the repetitions between reading and recitation, reciting those few sections which can be recalled, but referring promptly to their paper when the material is not forthcoming. Another explanation is possible, but less probable, *e. g.*, that the results are due to a real difference among the classes in ability to make rapid headway in the first few minutes of study. That this is not highly probable is indicated by the fact that under optimum conditions all classes learned approximately the same proportion of their respective lessons in the given time.

A final difference is that the upper grades, in comparison with the lower, do better when recitation is not introduced until fairly late, *i. e.*, when the proportion of reading is greater. Table XXII discloses the fact that for Grades six and eight Method Three (sixty per cent. reading) is as good as any other, but for Grades three and four, Method Three is considerably inferior to Methods Four, Five, or Six. For Grade three the superiority of Method Four over Method Three is 25.39 per cent. ± P. E. 5.65; for rade Gfour, 14.48 per cent. ± P. E. 6.70. Grade five lies between the extremes, showing a small (5.11 per cent.) but scarcely reliable superiority of Method Four over Method Three. This difference cannot easily be accounted for, precisely, with the evidence at hand. It is probably due to the greater experience of the more advanced students in learning material—history, geography, and other lessons—in which reading plays a very important part. By virtue of this experience, the older children were more skillful in employing the most fruitful methods of attack in reading which virtually amounted to less pure reading, *i. e.*, reading which was in some degree recitation. The younger children stuck more strictly to pure reading. These matters, however, must be waived to a later consideration. It is only necessary here to suggest that such class differences, whatever the explanation for them may be, are of marked pedagogical importance.

Summary of Results for Sense Material

1. In general, best results are obtained by introducing recitation after devoting about forty per cent. of the time to reading. Introducing recitation too early or too late leads to poorer results.

2. In general, the optimum combination of reading and recitation, under the conditions of the present tests, shows a superiority over reading alone by about thirty per cent.

3. The lower grades differ from the upper grades in three respects.

   a. The advantage of the best combination of reading and recitation over the method of learning by reading alone is twice as great for the lower grades, the average for grades three and four being 35.99 per cent. as compared to 18.00 per cent. the average for grades six and eight.

   b. Introducing recitation earlier than the stage indicated in (1) above, had a disadvantageous effect upon the learning of the lower grades, but little or no ill effect upon the work of the upper grades.

   c. The upper grades, in comparison with the lower, learn more effectively under the methods involving a relatively large amount of reading.

RESULTS AS REGARDS RETENTION OF NONSENSE MATERIAL

Tests for retention of nonsense syllables were given from three to four hours after the learning period, the exact intervals varying for different classes but being always the same for all squads of any one class. The pupils were simply asked to write down in proper order all the syllables they could remember. It is impossible to determine the unreliability of the retention results due to review, intentional or otherwise, on the part of the pupils during the interval between the learning period and the tests. That a few pupils did review the material during the interim was obvious from the fact that they obtained a higher score in the retention tests than in the immediate test. Such results were, of course, discarded. It was impossible to detect other cases in which the reviewing was less extensive. With the exception of a few suspicious cases, the results showed little or no indication of such procedure. An effort in the way of appeal from teachers and the experimenter was made to discourage such practices, and, on the whole, there are good reasons for believing the results, aside from the exceptions mentioned, are quite reliable enough for broad interpretation. It would be unwise, however, to give the data much weight for the interpretation of fine differences, such as the differences between closely related classes.

Table XXV shows the results in the form of averages with P. E.'s computed in the manner previously described. Table XXVI likewise shows the results on a relative basis.

TABLE XXV

Showing the average scores obtained in the retention tests

Method		1	2	3	4	5
Grade eight	Average score	7.02	12.55	13.66	17.55	22.89
Interval four hours	P. E.	0.42	0.78	0.57	0.67	0.88
Grade six	Average score	5.23	7.12	9.91	12.58	20.38
Interval three hours	P. E.	0.45	0.54	0.72	0.82	1.19
Grade four	Average score	3.49	5.89	8.35	10.58	14.25
Interval three hours	P. E.	0.38	0.56	0.53	0.47	0.53
Grade one	Average score	4.03	3.17	3.57	3.37	3.11
Interval three hours	P. E.	0.18	0.24	0.21	0.18	0.27

It is at once apparent that in a general way the results of the retention tests are very similar to those of the immediate test. Grade one stands by itself again for reasons that have been mentioned. Table XXVII shows the results of Grades four, six, and eight combined.

TABLE XXVI

Showing the data of Table XXV on a relative basis

Method		1	2	3	4	5
Grade eight	Relative score	47.65	85.20	92.73	119.14	155.46
	P. E.	2.85	5.30	3.87	4.55	5.98
Grade six	Relative score	47.37	64.49	89.76	113.95	184.60
	P. E.	4.07	4.88	6.51	7.42	10.76
Grade four	Relative score	41.01	69.21	98.12	124.32	167.45
	P. E.	4.44	6.55	6.20	5.49	6.20
Grade one	Relative score	116.81	91.88	103.48	97.68	90.14
	P. E.	5.22	6.96	6.09	5.22	7.83

TABLE XXVII

Method	1	2	3	4	5
Relative score—average for Grades four, six, and eight	45.34	72.96	93.53	119.13	169.17
P. E.	1.19	3.32	1.34	1.92	4.47

Table XXVIII following shows the differences between the various methods computed from Table XXVII.

TABLE XXVIII

Showing the differences between the various methods with P. E.'s of the differences

Differences of methods	Differences of methods	Differences of methods
2—1 = 27.62 ± P. E. 3.51	3—2 = 20.57 ± P. E. 3.60	4—3 = 25.60 ± P. E. 2.34
3—1 = 48.19 ± P. E. 1.78	4—2 = 46.17 ± P. E. 3.87	5—3 = 75.64 ± P. E. 4.62
4—1 = 74.79 ± P. E. 2.25	5—2 = 96.21 ± P. E. 5.56	5—4 = 50.04 ± P. E. 4.84
5—1 = 80.75 ± P. E. 4.58		

The steps from Method One to Method Five are all large and reliable. Nearly four times as much is recalled when the learning was predominantly recitation (Method Five) as when it was entirely reading (Method One). As the amount of recitation increases the amount recalled becomes greater. This increase in the amount recalled is fairly uniform with the exception of the comparatively great difference between Methods Four and Five. An explanation for this was suggested earlier.

It will be recalled that in the immediate tests no differences were found between the performances of the grades (except Grade one) so far as the effects of the different methods of learning were concerned. In the recall tests, there seems to be a slight difference between Grades eight and five with respect to the superiority of Method Five over Method One. From Table XXVI the differences between Methods Five and One have been computed with results as follows:

Differences for grade eight is 107.81 ± P. E. 5.83
Differences for grade six is 137.23 ± P. E. 11.70
Differences for grade five is 126.44 ± P. E. 7.61

The superiority, in this respect, of Grade five over Grade eight is 18.63 per cent. ± P. E. 9.53. The P. E. of the average of Grade six is so large as to make comparisons with that grade meaningless. Although Grades eight and five do differ by twice the P. E., the exception in the case of Grade six and the possibility

TABLE XXIX

Showing the score obtained in the retention tests for sense material

Method		1	2	3	4	5	6
Grade eight	Relative score	9.59	11.60	15.29	15.77	15.51	14.53
Interval four hours	P. E.	0.37	0.50	0.46	0.54	0.47	0.59
Grade six	Relative score	8.13	8.61	12.36	13.43	12.99	11.13
Interval three hours	P. E.	0.43	0.40	0.60	0.60	0.62	0.57
Grade five	Relative score	7.17	8.20	10.51	12.28	10.79	11.62
Interval three hours	P. E.	0.27	0.29	0.39	0.44	0.36	0.39
Grade four	Relative score	7.66	9.14	9.67	11.23	10.36	9.90
Interval four hours	P. E.	0.49	0.44	0.47	0.61	0.47	0.47
Grade three	Relative score	4.75	5.83	8.16	9.40	8.89	8.70
Interval three hours	P. E.	0.36	0.39	0.41	0.43	0.44	0.34

of a more general unreliability of the data (see p. 46) for fine distinctions, casts doubt upon this apparent difference between grades.

In general, then, the results for the retention of nonsense syllables are similar to those found in immediate tests, with the important difference that the superiority of the methods involving recitation is much greater.

RESULTS AS REGARDS RETENTION OF SENSE MATERIAL

Tests for retention of the sense material were given from three to four hours after the learning tests, the time always being the same for each class. The names of the individuals whose biographies had been studied were written on the board and the pupils were asked to write all they could remember about each person. Ample time was given.

TABLE XXX

Showing the data of Table XXIX on a relative basis

Method		1	2	3	4	5	6
Grade eight	Relative score	79.58	96.26	126.88	130.87	128.71	120.58
	P. E.	3.06	4.14	3.81	4.47	3.89	4.89
Grade six	Relative score	74.31	78.70	112.97	122.76	118.73	101.73
	P. E.	3.93	3.65	5.48	5.48	5.66	5.20
Grade five	Relative score	71.06	81.26	104.16	121.70	106.93	115.16
	P. E.	2.71	2.90	3.89	4.42	3.60	3.90
Grade four	Relative score	79.29	94.61	100.10	116.25	107.24	102.48
	P. E.	5.04	4.53	4.84	6.28	4.84	4.84
Grade three	Relative score	62.33	76.50	107.08	123.35	116.66	114.17
	P. E.	4.71	5.10	5.37	5.63	5.76	4.45

Table XXIX shows the results for the various grades in the form of averages with P. E.'s computed as before. Table XXX gives the same data on a relative basis. Figure 3 gives the data of Tables XXX and XXXI in graphic form.

TABLE XXXI

Showing the average of the results for all grades

Method	1	2	3	4	5	6
Average	73.31	85.44	110.23	122.98	115.65	110.82
P. E.	1.93	2.45	2.83	1.43	2.46	2.27

Table XXXII following shows the differences between the various methods with the P. E. of the differences, computed from the data of Table XXXI.

TABLE XXXII

Showing the differences between the various methods with the P. E. of the differences

Methods	Methods	Methods
2—1 = 12.13 ± P. E. 3.12	3—2 = 24.79 ± P. E. 3.74	4—3 = 12.75 ± P. E. 3.16
3—1 = 36.92 ± P. E. 3.42	4—2 = 37.54 ± P. E. 2.83	5—3 = 5.42 ± P. E. 3.74
4—1 = 49.67 ± P. E. 2.40	5—2 = 30.21 ± P. E. 3.46	6—3 = 0.59 ± P. E. 3.63
5—1 = 42.34 ± P. E. 3.12	6—2 = 24.38 ± P. E. 3.34	5—4 = −7.33 ± P. E. 2.84
6—1 = 37.11 ± P. E. 2.98		6—4 = −12.16 ± P. E. 2.68
		6—5 = −4.83 ± P. E. 3.34

From these tables it is clear that the general results for retention are very similar to those obtained in the immediate test. Method Four is the best, while Method One is the poorest. In the immediate test (see Table XXIV) Method Four showed a superiority over Method One of 26.64 per cent. ± P. E. 2.5 while in the retention test (see Table XXXII) the difference is 49.67 per cent. ± P. E. 2.4 or nearly twice as great. The earlier finding that Methods Five and Six are somewhat inferior to Method Four is borne out by the similar result in the retention test. On the whole the results of the immediate and the delayed tests are similar except that the differences between methods are more pronounced in the retention tests. In the immediate test Method Four was superior to Method One by 26.64 per cent. ± P. E. 2.5; while in the retention test the superiority is 49.67 per cent. ± P. E. 2.4.

Differences Among Grades

When the differences between grades are considered, the correspondence of the results for immediate and delayed memory is not so close. The finding in the immediate tests, that the difference between the best method and the poorest method was much greater for the lower grades, is not borne out by the results for retention.

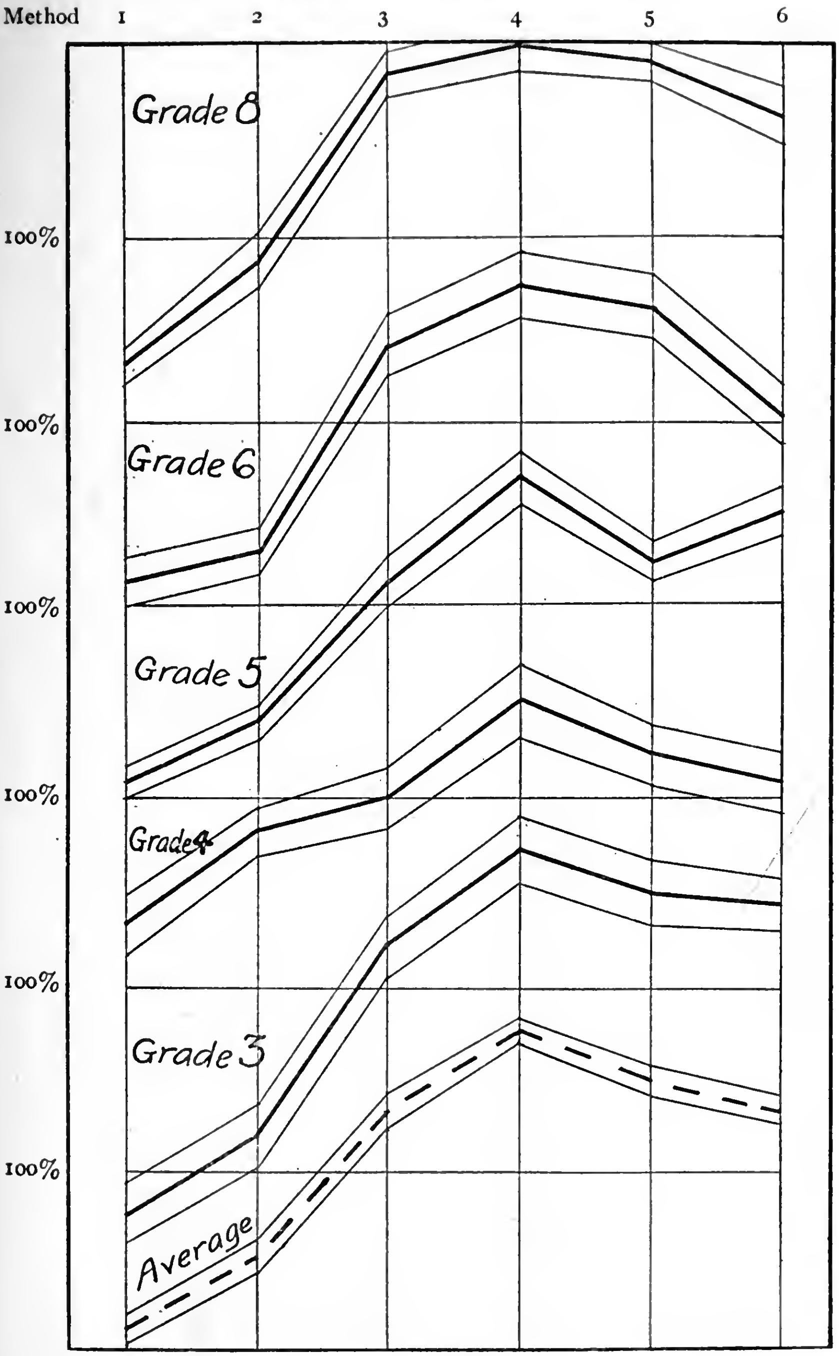
The average superiority of Method Four over Method One

For grade eight	= 51.29 per cent.	± P. E. 5.38
For grade six	= 48.45 per cent.	± P. E. 6.70
For grade five	= 50.64 per cent.	± P. E. 5.19
For grade four	= 36.96 per cent.	± P. E. 8.06
For grade three	= 61.02 per cent.	± P. E. 7.28

The magnitude of these differences shows no correspondence to order of grades.

FIGURE 3

Based on the data of Tables XXX and XXXI



Another difference between grades previously found, namely, that the methods introducing recitation very early worked a hardship upon the lower grades but not on the upper ones, is not shown by the results for retention.

The superiority of Method Four over Method Six, for example, is

For grade eight	=	10.29 per cent.	± P. E. 6.63
For grade six	=	11.03 per cent.	± P. E. 7.55
For grade five	=	6.54 per cent.	± P. E. 5.83
For grade four	=	13.77 per cent.	± P. E. 7.93
For grade three	=	9.18 per cent.	± P. E. 7.14

The differences between grades in this respect are nil. In fact, all of the differences between the two methods are so small in comparison with the P. E. as to be of very doubtful significance.

A third difference found on the immediate tests—that for the upper grades a method involving sixty per cent. reading (Method Three) was quite as good as any other, while for the lower grades this method was distinctly inferior to the methods giving more time to recitation—is quite clearly indicated in the findings for retention.

The superiority of Method Four over Method Three

For grade eight	=	3.99 per cent.	± P. E. 5.83
For grade six	=	9.79 per cent.	± P. E. 7.74
For grade five	=	17.54 per cent.	± P. E. 5.83
For grade four	=	16.15 per cent.	± P. E. 7.93
For grade three	=	16.27 per cent.	± P. E. 7.68

For Grades six and eight, Method Four shows no real superiority over Three, but in the case of Grades five, four, three real differences appear.

Summary of Results for Retention of Sense Material

In general, the results are similar to those found for immediate tests, the differences between the best and poorest methods being somewhat greater. The superiority of Method Four over Method One in the immediate test was 26.64 per cent. ± P. E. 2.5 as compared to 49.67 per cent. ± P. E. 2.4, or very nearly twice as great.

In some respects, the findings for the retention tests have not borne out the earlier results concerning differences between the various classes. But, as was explained before, certain possible sources of error in the data from the retention tests render these results of doubtful value for fine comparisons. The findings in the immediate tests are probably better indications of real distinctions between grades.

RESULTS FROM EXPERIMENTS UPON ADULTS WITH NONSENSE MATERIAL

In addition to the experiments carried out with children as subjects, tests were made upon adult students, using materials and methods similar in most respects to those previously described. The purpose of this extension of the work was threefold: *first*, to furnish a basis of comparison of the present results with those of earlier investigations; *second*, to permit a comparison of the work of children with adults in similar tests; and *third*, to obtain more detailed information concerning the nature of the particular functions employed in the two methods of learning.

The subjects for the tests to be described were for the most part graduate students, members of classes in experimental psychology. The materials used were qualitatively the same as those employed with the children.

Fifteen such students of psychology at Columbia were given, on three different days, ten-minute tests with series of twenty nonsense syllables. Each day one of the three different methods of study was used; *first*, 10′ reading; *second*, 5′ L + 5′ R; *third*, 2′ L + 8′ R.

The group was divided into three squads, and practice effects, individual differences and differences in tests were neutralized in the total by employing a method in all essentials the same as that described on p. 26 ff. The records of the individual students, however, are fairly reliable as such for the reason that all of these subjects had just completed a series of experiments on the learning process and memory extending over three months. Each had learned during this time several hundred nonsense syllables as well as much other material and were thus fairly highly practised subjects. Detailed introspective accounts of the factors involved in the several methods of learning were requested. These will be considered later. Each person acted as subject for his or her regular laboratory partner, who kept the time, and noted the number of repetitions made. Later the two reversed positions, the former experimenter now acting as subject. Each used a different series of syllables, and six different texts were used altogether.

The data were scored by giving a grade for each correct letter when there were two or more correct, and an additional credit when the syllable was in correct position. The highest possible score for the twenty syllables would thus be eighty.[7]

[7] See Lyon, D. O., 'The Relation of Quickness of Learning to Retentiveness', *Archives of Psychology*, 1916, No. 24, p. 27.

Table XXXIII shows the records of the various individuals with the average, the A. D.'s, for the number of repetitions and the P. E.'s for the obtained average. The P. E.'s were computed by the formula P. E. $= .6745 \frac{\sigma \text{ dis.}}{\sqrt{n}}$

The differences between the methods as shown by the average results are large and reliable, the P. E.'s being small. In general Method Three results in more than twice as much material learned

TABLE XXXIII

Showing the scores obtained by adults in learning nonsense syllables. Highest possible score eighty

Subject	Method 1. L10′		Method 2. L5′ R5′			Method 3. L2′ R8′		
	Repetitions	Score	Repetitions L	Repetitions R	Score	Repetitions L	Repetitions R	Score
	Day one		*Day two*			*Day three*		
Hm.	62	22	26	21	60	6	25	78
Dn.	54	32	22	18	49	5	21	80
Bm.	23	16	13	11	58	5	10	74
Bs.	22	32	15	10	60	4	9	80
Tr.	50	23	24	23	38	6	15	68
	Day two		*Day three*			*Day one*		
Py.	55	7	20	6	19	6	9	47
Gl.	68	9	17	5	40	7	13	49
Bn.	26	38	23	14	52	7	16	78
Jy.	26	28	9	7	32	6	14	42
Hr.	38	36	12	10	56	5	16	50
	Day three		*Day one*			*Day two*		
Mk.	37	24	18	14	48	6	16	60
Sn.	50	29	24	20	58	9	29	74
At.	36	56	10	8	56	4	18	63
Ly.	23	12	25	15	43	6	11	62
Wp.	29	23	10	14	30	4	15	50
Average	40 A.D.14	25.8 P.E.1.2	18 A.D.5.2	13 A.D.5.1	46.6 P.E.1.4	5.7 A.D.1.0	16 A.D.4.0	63.7 P.E.1.4

as Method One, Method Two stands slightly above the average for the other two. Minor differences among individuals will be apparent on observation, but it will be noticed that in but one case is Method Three inferior to Method Two and in no case is Method Two inferior to Method One, although in one case Method Two is equal to and in two other cases very slightly superior to Method One. The superiority of Method Three over the others, is somewhat greater than the data show in two cases. Subject *Bs.* had completely learned the series in 8′42″ under Method Three and Subject *Dn.* had completed the learning some time (exact amount not known) before the end of the ten-minute period.

Great individual differences appear with regard to the number of repetitions made during the ten-minute study period. The average number of repetitions when the learning was entirely by means of reading was forty, with a mean variation of fourteen. The extreme rates were those of Subject *Bs.* with twenty-two repetitions and *Gl.* with sixty-eight, or three times as many as *Bs.* Method Two shows similar individual differences in the learning by reciting as well as by reading. The average number of repetitions for five-minute reading being eighteen with a M. V. of 5.2 and for five-minute recitations the average number is thirteen with a M. V. of 5.06. The average figures also show that the rates of repetition were less for learning by reciting than for learning by reading, although as far as this test is concerned, the difference may be taken to mean merely that the repetitions in the last half of a period of learning are longer than those of the first half. That the former interpretation is more likely to be the correct one is indicated by the fact that the sum of the repetitions for the all reading test (Method One) is greater than for the half reading, half recitation test (Method Two), *i. e.*, forty as compared to thirty-one. This greater speed of repetitions in the reading portion of Method Two is shown by fourteen of the fifteen individuals. Method Three shows the same situation, the total number of repetitions here being twenty-two, with rather wide differences among individuals.

More Intensive Work with Nonsense Syllables

Somewhat more extensive work was done with two graduate students, more skilled in introspective observation. Each of these subjects was given several preliminary tests to insure an acquaintance with the procedure and to eliminate practice effects to some extent, before the main experiment was begun. Series of twenty nonsense syllables were studied for eight minutes according to six different methods. Three tests were made by each method, and in each case the number and duration of the repetition were noted by the writer who kept the time with a stop watch. But one test was made on a single day. A recall test was made after approximately six hours for Subject *S* and after twenty-four hours for Subject *T*. The following table gives the results in detail.[8] The data were

[8] The durations of the repetitions are not presented here, but will be mentioned in a late section.

scored in the manner described on p. 53, eighty being the highest score possible. The 'natural method' gave the subject liberty to study in any way he might choose.

The results for Subjects *S* and *T* are very much the same as the average results just found for the larger group. The differences between Methods Four and Five for both subjects are too small to

TABLE XXXIV

Showing the average results for three trials of each subject

Combinations	Method One 8'L		Method Two 6'L 2'R			Method Three 4'L 4'R		
	Repetitions	Score	Repetitions		Score	Repetitions		Score
	L		L	R		L	R	
Subject S, Average	17	24	8	2.3	34	6	4.3	48
Subject T, Average	32	16	9	3.0	27	13	10.0	34
Subject S, recall after six hours. Average		7.6			10.3			17.6
Subject T, recall after twenty-four hours. Average		8.6			11.3			13.3

Combinations	Method Four 2'L 6'R			Method Five 1'L 7'R			Method Six Natural Method		
	Repetitions		Score	Repetitions		Score	Repetitions		Score
	L	R		L	R		L	R	
Subject S, Average	3.5	5.6	49	2	8	52	2.6	9	51
Subject T, Average	4.0	18.0	49	3	22	51	3.0	20	50
Subject S, recall after six hours. Average			27.3			26.3			29.0
Subject T, recall after twenty-four hours. Average			22.6			24.0			21.3

be of significance and the 'natural method' produces results that are quite as good as any other. This means that trained subjects are capable of discovering and employing the best methods of attack. Subject *S* began to attempt to recite in the 'natural method' after two, four, and two repetitions respectively or at about the same stage at which recitation was begun in Methods Four or Five. The case is similar for Subject *T*. Subject *S* in Methods Four, Five, or Six learned about twice as much as in Method One, while Subject T learned about three times as much. The recall tests after six or twenty-four hours bear out the findings for the learning test, being somewhat more emphatic. For Subject S under the optimum methods shows an amount recalled three times as great as under Method One while for Subject T the ratio is nearly four to one.

The speed of repetitions varies considerably for the two subjects and for the same subject at different times, but in nearly all cases reading seems to be done at a higher speed than recitation, although as will be found later the duration of repetitions during reading are very uniform while those during recitation are very irregular.

In addition to the data here presented, a few additional experiments, somewhat more specialized in nature, were performed and are presented in a later section (pp. 71 and 72), in which fourteen adult subjects participated in two five-minute periods of studying sixteen nonsense syllables, according to two methods: *first*, in which only reading was permitted; and *second*, in which recitation was permitted from the first. The average results show a score 16.4 for the reading method and 32.85 for the recitation method, or exactly twice as much. A similar test (p. 81) with eleven subjects gave similar results, 5.54 syllables being correctly recalled in the reading test as compared to 11.4 in the recitation test.

Summary of Results for Adults with Nonsense Syllables

1. Several different experiments upon adult students in learning nonsense syllables produce results similar to those found for children. The advantage of methods affording an optimum amount of recitation over the reading methods is very great, the two methods showing in general a ratio of about two to one.

2. Although considerable individual differences were found, no subject was discovered who did not obtain better results with recitation than without it.

3. Great individual differences were found in the rate at which the series were read or recited, but in general the durations of recitations are longer than the durations of readings.

4. The advantage of the methods combining recitation with reading in the learning period is more pronounced in delayed than in immediate recall.

EXPERIMENTS UPON ADULTS WITH SENSE MATERIAL

Non-connected Sense Material

Two graduate students, S and T, acted as subjects for a few tests, in studying for eight-minute periods series of thirty words of four letters each, according to several different methods. Two preliminary trials were given in each case before the actual series were started. Two series of tests were given, the order of methods being reversed in the second series. No word was repeated in the series of lists used. The data were scored by giving a credit of two for a correct word and an additional credit if it were in the correct position. Thus the highest possible score would be ninety. Table XXXV gives the results:

TABLE XXXV

Combinations	Method one 8′ L		Method two 4′L 4′R			Method three 2′L 6′R			Method four Natural method		
	Repetitions	Score	Repetitions		Score	Repetitions		Score	Repetitions		Score
	L		L	R		L	R		L	R	
Subject S											
First series	12	30	7	4	49	3	10	53	3	11	51
Second series	15	35	6	6	58	2.5	11	60	2	11	54
Average	13.5	32.5	6.5	5	53.5	2.7	10.5	56.5	2.5	11	52.5
Subject T											
First series	14	27	8	7	39	3	14	45	3	13	50
Second series	16	34	9	7	42	3	13	47	3	15	47
Average	15	30.5	8.5	7	40.5	3	13.5	46	3	14	48.5
Subject S. Recall after six hours											
First series		15			30						20
Second series		12			17			30			
Average		13.5			23.5			30			20
Subject T. Recall after six hours											
First series		12			25						
Second series		14			29			30			
Average		13			27			30			

While this experiment is far from being extensive enough to be decisive, it is suggestive. The subjects were well habituated to this kind of learning, having previously learned nearly thirty series of nonsense syllables. Both agree in showing that lists of words can be more readily learned by a method which permits recitation, but the difference between the methods is not so pronounced as was found with nonsense syllables. For Subject *S*, in learning nonsense syllables, Method Four was related to Method One as two to one; for Subject *T* the ratio was nearly three to one; while for lists of meaningful words the ratios of the corresponding methods are for Subject *S* about one and seven-tenths to one, for Subject *T* one and five-tenths to one. The retention tests for series of words show a similar ratio, although the data are too few for reliable results.

Experiments with Connected Sense Material

Subject T endeavored in six different tests of ten minutes each to learn twenty-line stanzas of poetry from Goldsmith's 'Deserted Village', according to three different methods. Recall of the material was attempted after six hours. The results show the number of words learned or remembered.

TABLE XXXVI

Combinations	Method One 10′ L		Method Three 5′L 5′R			Method Four 2½′L 7½′R		
	Repetitions	Score	Repetitions		Score	Repetitions		Score
	L		L	R		L	R	
First series	11	78	6	4	93	3	7	84
Second series	12.5	86	5	5	106	3	8	108
Average		82			99.5			96
Recall after six hours								
First series		44			66			
Second series		53			58			58
Average		49.5			62			58

The advantage of the methods including recitation over the reading method is apparent although not very large in both the immediate and the delayed memory test. The method employing fifty per cent. reading and fifty per cent. recitation seems to be quite as good as the method permitting seventy-five per cent. recitation. The results, of course, are too few to be more than suggestive, although they do seem to be quite in harmony with the findings for children.

The biographical sense material used with the school-children (see p. 26) was studied by fifteen graduate students under three different methods, as shown in Table XXXVII. The fifteen subjects were divided into three groups, each employing a different method on the different days. Other details were the same as those described on page 53.

TABLE XXXVII

Showing the number of details or facts recalled

Combinations	Method One 8'L	Method Two 4'L 4'R	Method Three 2'L 6'R
Subjects	*Day one*	*Day two*	*Day three*
Bm	41	48	52
Ws	14	28	32
Ky	36	39	45
Tr	9	20	22
Sn	13	18	19
	Day two	*Day three*	*Day one*
At	39	47	49
Rs	40	47	39
Gl	10	20	19
Py	8	22	20
Sa	8	18	21
	Day three	*Day one*	*Day two*
Mn	14	21	26
Ce	23	27	31
An	18	24	27
Sn	19	18	21
Ms	18	30	26
Average	20.6	28.4	29.9
P. E.	2.1	1.9	1.9

The average results for Methods Two and Three are distinctly superior to those for Method One, and this is true practically without exception for all of the individual cases. The difference, in the average results, between Methods Two and Three is small and being no larger than the P. E. is unreliable. The general result of this test upon adults is the same as that obtained from the older children with the same sort of material. The value of recitation as compared to reading is not so great as it is when nonsense material is used, and no particular advantage is obtained by introducing the recitation very early in the learning.

GENERAL SUMMARY

Nonsense Material

1. In general, recitation, after a few initial readings, is of much more value in learning than more reading.

a. Under the conditions of the present experiment a method devoting the first twenty per cent. of the time to reading followed by eighty per cent. recitation will result in learning for immediate reproduction twice as much material as will a method of reading only.

b. As measured by recall three to four hours later, the difference between the two methods is about twice as great; four times as much being recalled under the recitation method as under the reading method.

2. After a certain amount of initial reading (one minute and forty-eight seconds or twenty per cent. of the total time in this experiment) the more quickly the recitation is introduced the better the results as measured by either immediate or delayed recall.

3. No conspicuous differences appear between the results for adult subjects and children or between the various grades with the exception that the findings for the first grade differ from all others.

Sense Material

1. In general the best results are obtained from a method devoting about forty per cent. of the time to reading followed by an equal amount of recitation.

2. In general, the optimum combination of reading and recitation produces in immediate tests results superior by about twenty-seven per cent. to those obtained from reading only.

a. The difference shown by recall three or four hours later is nearly twice as great as that shown in the immediate test.

3. In most respects the results for adults and for the various grades are very similar.

4. In certain respects differences between the grades were found on the basis of the results of immediate tests.

a. The advantage of the best methods over the poorest is much greater in the lower grades than in the upper, *e. g.*, the average advantage for grades three and four of the best method over the poorest is 35.99 per cent. as compared to 18 per cent., the average for Grades six and eight.

b. Introducing recitation very early in the study period has a disadvantageous effect upon the learning of the lower grades, but has little or no ill effect upon the work of the upper grades.

c. The upper grades, in comparison with the lower, learn more effectively under the methods involving a relatively large amount of reading.

5. With the exception of (*c*) above, none of the differences between grades were evident in the results of the retention tests.

a. This was believed to be due, in the main, to unavoidable errors which crept into the retention tests (see p. 46).

Results from Tests on Adults

1. The advantage of recitations over reading is greater the more senseless and unconnected the material.

a. Advantage is greatest for nonsense syllables, less great for lists of words, and still less great for connected prose or poetry.

2. Great individual differences appear in the tempo of studying by reading or recitation, some individuals completing a perusal on the average in one-third of the time taken by others.

3. As a rule, the tempo is considerably quicker in reading than in recitation, for most individuals.

4. Usually, a given individual during a single sitting, reads and rereads at a very uniform speed, while the rates for consecutive recitations are very variable.

COMPARISON WITH RESULTS OF OTHER INVESTIGATIONS

The general findings in the present experiment upon children as well as upon adults are in harmony with the results of most of the earlier investigations, which were presented in Chapter II. It will be necessary here to recall but briefly the conclusions obtained in some of the more important of the earlier works.

Katzaroff found, by combining the results for three subjects, four tests each, that fifteen readings of nonsense syllables—the test being made seventy-two hours later—produced a score of six as compared to twenty obtained from eight readings and seven recitations. Other individuals in similar tests, showed even greater differences. Witasek, Knors, and Abbott also verified the greater effectiveness of recitation in learning nonsense syllables under various conditions, although the quantitative determination of the superiority of recitation has differed considerably.

The work of Kühn, being more akin to the present experiments, is of more value for comparative purposes. In immediate tests, the superiority of recitation over reading found by Kühn is very similar to that found in the present work, for each of the several materials used. Kühn's conclusion (p. 422), "By the majority of people [adults] recitation is much better than readings, and the relative advantage is greater, the more senseless the material," is verified by the present results with children as well as adult subjects.

With regard to the present finding that the superiority of recitation over reading is greater when measured by delayed than by immediate recall, but little evidence has been produced by the earlier studies. But the results that are available seem to be in harmony with the present findings. For example Kühn found (see p. 8) that a lesson, although learned in very much less time by means of recitation than by reading alone, was retained much better and that the superiority of recitation in this respect became greater the longer the retention test was delayed.

The matter of individual differences deserves consideration. Abbott in experiments upon five subjects found one among these for whom reading was a better method of learning than recitation and Kühn found the same in the case of three out of thirteen subjects. Both investigators found that such learners employed a 'purely mechanical' form of learning or were of very strong visual type—such that best results were obtained when the subject simply 'looks at a word and lets it soak in'. Abbott concludes, "We must go back to the type of the individual to explain the processes and relative efficiency in recall." This matter of learning types will be taken up in more detail in the next section. For the present, while there is no intention of contending that such extreme types as those found by Abbott and Kühn do not exist, the present work indicates that they are in no wise numerically so prominent as their findings would suggest. While Kühn found three among thirteen subjects, and Abbott one among five, in the present work, tests upon fifty or more adults made under less artificial conditions have not produced a single case of such 'mechanical' or 'strongly visual' types. In no case has the method of learning by reading given better results than a method in which recitation was also a factor. Unfortunately, the data of the children cannot be employed on this point with assurance, for the reason that the effects of a particular method in the case of any individual may be marked by practice effects, differences in texts, and the like. However, in spite of all these differences, an examination of the individual data shows that exceptions to the general rule that recitation is more effective than reading are very, very rare. This fact has a very important pedagogical significance, since it gives assurance that such applications as follow from a study such as the present one, may be made by the teacher to her pupils as a whole without working a hardship on more than a very few if any individuals.

Further considerations of interest to pedagogy, such as the optimum point of introducing the recitation in the case of various materials, and the efficacy of various minor functions employed in learning, will be treated in more detail in the next two sections.

V

AN ANALYSIS OF READING AND RECITATION AS FACTORS IN LEARNING

The previous section, from an objective point of view, gave us certain facts concerning two very broad and complex functions, *e. g.*, learning by reading and learning by recitation. It was found that the results, measured in terms of the amount of material learned in a given time and the amount retained after a given time, differed considerably according to the proportion of time allotted to one or the other of these two functions. One is interested to discover, if possible, in just what manner these two broad functions differ, since the result of their exercise is so markedly different. It is likely that the best method of discovering these differences is to analyse each of the complex functions, as far as possible, into their elements, finding just what minor functions are operative and in what manner they combine to make up the gross functions of learning in each case. If such an analysis can be successfully accomplished, the result should be a much better understanding of the two functions as a whole and the production of valuable suggestions with regard to the selection and combination of constituent functions for the most economical methods of study.

But such a reduction of the complex functions into constituent processes that shall be typical is by no means an easy or a certain matter. Some of the elementary functions can be observed from the outside and can be verified by objective tests, but most of the facts can be observed only by the learner and we are forced to limit ourselves to his reports upon them. Indeed, most of our analysis is of the introspective, or more accurately retrospective sort, subject to the limitations of this form of evidence.

In the present work an effort has been made to get reports as full as possible, and as free from suggestion as possible, from subjects believed to be reliable and capable. About forty subjects in all were used, and they were subjects whose experience seems to have fitted them for the retrospective work. Nearly all had had several months' practice in introspective reporting, each having learned, previous to the experiments, a large number of series of nonsense syllables and other kinds of material and having had considerable practice in describing their mental imagery in various sorts of mental tasks. After each test in the present experiments the subjects wrote a full account of the functions employed in the learning, such as the kinds of imagery employed, the kinds of 'aids' used and how, their attitude toward the work, the satisfyingness and annoyingness of different methods, the fatigability of different methods, and the like. Reports from the children were secured on many of these points also. Wherever practicable the introspective accounts were checked up or tested by manipulation of the data already at hand or by new experiments devised to fit the case, and the results of other studies have been freely drawn upon.

Before proceeding to the results, a few cautions should be indicated. In the first place, individual differences play a large role. No single individual at any time is likely to make use of all the minor functions that will be described. Some subjects place more emphasis upon certain functions, some upon others, and the same individual usually changes his method to some extent according to the nature of the material and the like. More constant differences among individuals due to earlier training in learning methods or to memory types will be mentioned. But just as we found in the preceding section no very sharp differences in mental type, and no definite cases in which reading proved to be superior to recitation, so we shall find that typical methods of learning contain the main functions employed by nearly all learners.

A second caution is that wholesale conclusions from results obtained mainly from adults should not be made to apply to the learning of children. Necessarily the introspective accounts are largely those of adults, but the reports of children have also been considered to some extent, and where possible, introspective accounts have been verified by objective data obtained from children. That the minor functions employed by children should correspond closely to those of adults has already been indicated by the fact that the results of the exercise of the two general functions have been very similar for both classes of subjects.

With these precautions in mind, a consideration of the various activities, aids, and attitudes involved in learning and recalling any material will now be taken up, special attention being given to the differences that appear according to whether the method of learning is reading or recitation.

Nearly all reports, in the first place, agree in emphasizing the fact that learning even a series of sixteen or twenty nonsense syllables is far from a simple mechanical task. The number and variety of associative aids is remarkable. Where adults go to their wit's end for such associations it can hardly be doubted that they assist learning. A consideration of such aids is perhaps a good place to begin.

In general it may be said that such aids to learning may be of two sorts: one which is found in the material itself, needing only to be noted and employed, and another sort which is worked into the material by the learner. Of either sort some may be marked off as motor in character and others as perceptual.

ARTICULATION

Although the tests for learning were always written, the majority of adult subjects reported that practice in accurate pronunciation of the material was an aid in learning. This was found to be particularly true in the case of nonsense material which was difficult to articulate. Subjects report that the motor and auditory elements of the words were secured better from reciting, especially when the material offers great difficulty in pronunciation. The learner is likely to begin by carefully articulating the material to himself while reading, but if the reading is prolonged too long, these functions are likely to be neglected. In many cases the explanation given for this is that they were able to move down the series of syllables more easily without articulating, depending more upon a visual imprinting of the data. In recitation this is rarely the case. When they attempt to recite the material, the articulation is a most natural and in most cases an essential act. The reproduction and practice of the motor act is an aid to learning. The school children found considerable difficulty in pronouncing the syllables, and for them actual articulation was more essential. The members of the sixth and eighth grades in answer to the question: "Why are the syllables so hard to learn?" wrote, many of them, "Because they are hard to say." They also reported that they liked the recitation because it gave them a better chance "to see if they could say them." Movements of the lips, sometimes without, although generally with whispering, especially in the lower grades, were very marked in the recitation part of the learning period.

ACCENTS AND RHYTHMS

Articulation is usually accompanied by accenting or stressing certain syllables or words, according to the report of nearly all subjects. The following serve as samples. Subject *Rs* in one test reported accenting syllables one, five, nine, thirteen, and seventeen in the series, syllables five and nine being more strongly accented than the others. Subject *Py* accented every third syllable. Subject *Sn* accented every fourth syllable strongly and every second syllable less strongly. Subject *Bn* reported an increasing accent within groups of four syllables, the last being most strongly accented, followed by a drop to the minimum on the fifth. Sometimes these accents are obvious to an observer who may notice the accompanying motor activities such as nodding the head, tapping the finger, or thumping the foot. That the children employ such accents was usually evident from such signs and was usually indicated by the whispering which accompanied the learning.

The value of such accents lies in the fact that a syllable comes to be associated with its accent and the act of accenting tends to call up the syllable. Although individuals show great differences in their choice of accents and the same individual may often employ different accents according to the material being studied, in any one lesson the accentuation is usually constant and assists learning through this tendency to repeat the same motor activity which acts as a frame-work upon which the syllables may be affixed.

Such accentuation should, however, be considered in connection with the almost universal employment of rhythm in the learning of a series of syllables. Müller and Schumann,[1] Meumann[2], and others have shown the value obtained from the employment of rhythm in learning. The kind of rhythm, like the kind of accentuation, varies with individuals and materials. In learning a series of twenty nonsense syllables, subject *At* divides the material into feet of three syllables, the first being long and accented, the two following unaccented and short —́ ∪ ∪ | —́ ∪ ∪ | with a pause between groups. Subject *Py* uses an identical rhythm. Subject *Sn* employs a trochaic measure with two pairs combined into a measure of four by placing greater accentuation on the third and seventh than on the first and fifth — ∪ —́ ∪ | — ∪ —́ ∪ |. Subject *Tr* employs a measure of four feet, a long accented syllable followed by three short unaccented syllables with a pause between measures, —́ ∪ ∪ ∪ | —́ ∪ ∪ ∪ |.

According to the reports of most individuals, the employment of such rhythms is the most natural thing in learning by recitation, but in reading they are not so frequently or easily used. Some report, in the case of learning by reading, that they begin by arranging the material for rhythmical perusal with accents and pauses but abandon the method before the lesson is over because it seems to be of no avail. It seemed that a method employing more visual factors and less motor would work better; their efforts were directed to 'looking hard' at the syllables to assist them to 'soak in'. Several subjects, however, reported that they did use a rhythmical division of the material throughout the reading, and their opinion that it did not prove to be of great value was usually borne out by the meagre results of the final tests. "With my eyes on the paper,"

[1] 'Experimentelle Beiträge zur Untersuchung des Gedächtnisses', *Zeitschrift für Psychologie*, 1894, **6**, pp. 81–191.

[2] *The Psychology of Learning*, translated by Baird, 1913.

says subject *S*, "it is hard to do more than just read hard and think about the individual syllables. I knew the rhythms and other aids would be of more value if I could only look away from the list."

Auditory and Visual Types of Learners

It appears that there are certain differences in method according to whether the learner relies more upon auditory-motor elements or upon visual elements in learning. In some cases in which the learning is predominantly of the auditory-motor type, imprinting consists in forming a series of auditory or vocal images of the whispered words or a series of successive innervations of the vocal muscles, which are often accompanied by sensations or images of movements. The subject learns the sounds, muscular feelings, and rhythmic sequences of the syllables which he memorizes. Reproduction may be a sort of melody in which the various syllables assume their proper rhythmical positions. Usually in reproduction the subject cannot get the whole series in consciousness at once. He must start the series off and let it run its course. Now many of these subjects report that reading is of value to a certain point, but if no opportunity for recitation is afforded, the latter part of the process of learning is very much hampered and complete learning seems impossible. The presence of the words to the eye precludes the subjective innervations which are essential for learning. A different process seems to be involved when the visual stimuli are absent.

Some subjects reported that they made use of visual imagery to a much greater extent. They were not so greatly hampered by lack of recitation. But no one was found who relied entirely upon visual imprinting, auditory and motor elements being always employed as well. Of those who relied to the maximum upon visual imagery, most employed a rhythmical division of the material to some extent. Such subjects divided up the material into measures, with a motor stressing of certain syllables coupled with a visualization of all of the syllables, especially those that were accented. They differed from the auditory-motor learners, apparently, only by relying somewhat more upon visualization and less upon the auditory and motor factors. None used visual imprinting alone. In the learning by reading these subjects employed the visual factors to the utmost, with the corresponding neglect of the auditory and motor elements. While their results, as a rule, differed less for the two methods than did those of the auditory-motor learners, in no case were they so efficient in tests permitting no recitation as in the tests in which recall was a factor.

Just as there was, among about forty different adult subjects, none that could be called a purely visual learner, so there was none that seemed to rely entirely upon auditory-motor factors. Visualization to some extent usually entered into the learning of the latter. The differences were merely those of emphasis upon one or another factor, and, indeed, among the subjects were many who seemed to be able to employ now some factors, now others, according to the situation to be met. In general, learning by reading seemed to throw the emphasis upon the visual method.

LOCALIZATION AND NOTING OF POSITIONS OF ITEMS

A number of aids to memorizing which are more of a perceptual than a motor sort are usually employed. They are closely connected with the motor aids of articulation, rhythm, etc. Some of these depend upon peculiarities or divisions found in the material itself, while others are worked into it by the subject.

One important matter is the noting of the positions of certain syllables. Such localizations seem always to be an aid to memory. Sometimes localization is greatly aided by peculiarities within the text, but often more arbitrary methods of obtaining a localization schema are employed. Some report that they simply localize a certain few 'head-liners' in the series by noting their positions in visual space. Although they are not able to visualize all of the items, a few are made to stand out plainly, serving as landmarks to which others are attached. Other subjects divide the list into a certain number of parts, a few syllables thus being denoted by their numerical positions. A few report these localizations to be determined by modulation of the voice or dependent upon the rhythm that is employed. *But all report that these localizations are an aid in memorizing and that they were more easily employed in recitation than in reading.* On the introspective side such reports as these are found: (subject *T*) "In reading it was so easy to glide through the series that I did not take the trouble to note any special points of interest. It seemed that I could do more if I just looked hard at the syllables, covering up my ears so that I could do nothing but look. But when I began to recite I found that I had to note certain syllables specially, which I afterwards used as starting and stopping places." Evidently, recitation tests the value of the different aids and generally leads the learner to recognize the value of those which serve the purpose desired.

In order to obtain some objective data on the matter of localization, a test was given for that purpose. Fourteen graduate students whose status and introspective training have been described, acted as subjects. Lists of sixteen nonsense syllables were used as material. All the subjects studied at the same time, half of them by the reading method first and half by the recitation method first. Later another experiment was given in which each used the other method. Five minutes were devoted to the study and the syllables were written down immediately afterwards. The subjects were then asked to indicate those whose positions they felt certain were correct, those which were doubtful, and those which they were sure were incorrect in position. They were then asked to describe the means or cues by which they made their judgments. The results are shown in Table XXXVIII.

It should first of all be noted that almost exactly twice as high a score was obtained by the recitation method, and this introduces a factor which tends to produce a better showing in the matter of accurate localization for the reading method. It will be noticed, for example, that many subjects in the reading series were certain of the positions of only two or three syllables, which were in nearly every case the first, or the first and second, and the last. It is well known that the first and last syllables are the easiest to learn and to localize. In the reading series these two or three syllables form

TABLE XXXVIII

Results given in the absolute number of syllables

Subject	Reading					
	Judged			Really		
	Correct	Doubtful	Wrong	Correct	Wrong	Score
Ln.	2	2	0	1	3	7
Sa.	3	2	0	2	3	13
Sn.	3	4	1	4	4	17
Ms.	2	1	2	2	3	13
J. M.	7	2	0	6	3	19
Tr.	3	2	1	2	4	11
Wr.	2	2	2	3	3	10
Ce.	4	2	1	5	2	16
Py.	9	0	0	5	4	21
Gl.	4	4	0	3	5	14
Mn.	6	4	0	7	3	23
An.	6	3	0	4	5	18
At.	11	2	0	9	4	32
Rs.	6	2	0	5	3	16
Average	4.85	2.79	0.50	4.14	3.5	16.4
Per cent.	59.6	34.3	6.3	54.2	45.8	

TABLE XXXVIII—*Continued*

Subject	Recitation					
	Judged			Really		
	Correct	Doubtful	Wrong	Correct	Wrong	Score
Ln.	6	1	3	4	0	26
Sa.	7	1	0	8	0	24
Sn.	16	0	0	16	0	48
Ms.	7	3	0	10	0	23
J. M.	8	1	2	8	3	30
Tr.	13	2	0	14	1	46
Wr.	4	5	0	4	5	24
Ce.	12	2	0	13	1	42
Py.	8	0	0	8	0	24
Gl.	10	0	0	10	0	30
Mn.	8	2	0	8	2	26
An.	7	2	0	7	2	25
At.	15	0	0	15	0	44
Rs.	16	0	0	16	0	48
Average	9.64	1.34	0.34	10.07	1.0	32.85
Per cent.	85.1	11.8	3.1	90.9	9.1	

a larger portion of the whole number written down than in the recitation series.

In spite of this advantage, the subjects, after learning by reading, felt certain of the positions of but fifty-nine and six-tenths per cent. of the syllables written down, as compared to eighty-five and one-tenth per cent. in the recitation series. The reading series is conspicuous with respect to the number of 'doubtful' cases, which amount to thirty-four and three-tenths per cent. as compared to eleven and eight-tenths per cent. for the recitation series, or three to one. So it is quite clear that the subjects are more confident of their opinions in the recitation series. The data also show that a much larger number of the syllables learned by recitation are in correct position (ninety-one per cent.) than in the reading series (fifty-four per cent.). Another important fact appears, namely, that the judgments after learning by recitation are not only more accurate but also more conservative than after learning by reading. As to the first point, while fifty-nine and six-tenths per cent. of the syllables written down after learning by reading were judged to be in correct position, only fifty-four and two-tenths per cent. actually were, while in the recitation series, of those written a larger percentage (eighty-five and one-tenth) was judged to be in correct position, and a still larger percentage (ninety and nine-tenths) actually was. An examination of the original data showed that in the reading series, of those judged 'doubtful' nearly all were really in an incorrect position, as were also nine of the sixty-eight certified as 'correct' in position, while in the recitation series some of those judged 'doubtful' were really in a correct position, while only two cases out of the total of 127 judgments of 'correct in position' were wrong. It thus appears that after learning by recitation, the subjects are both more accurate and more conservative in their judgments.

NOTING UNUSUAL CHARACTERISTICS OF THE MATERIAL

The remark was just made that the noting of unusual words and characteristics in the material was often an aid in localization. This function is of value because it serves to break up the material into units that can be more easily handled. A peculiar word or syllable becomes a center around which other syllables are grouped, or it may serve as a starting and stopping place within the series.

The kinds of peculiarities noted are myriad. Sometimes it is the sound—the children especially are attracted by 'funny sounding' syllables. Sometimes a syllable stands apart by having the consonants each standing above or below the line, *e. g.*, *gop*, *lib*. Sometimes the fact that one letter was printed light, or that the whole was blotched or blurred, or that a mark appeared on the page opposite it, is noted. More often the associations are meaningful, and these will be considered more fully in the next section.

Subjects report that all such peculiarities are brought out more clearly by reciting the material. They are not so effectively brought into play when one is reading because the words before the eyes render such aid unnecessary. The thing to do is simply to 'look hard and try to avoid distractions'. Subject *Fx* reports: "After the reading period was over [four minutes out of eight], I could remember only three syllables. I had a hazy idea of some of the others but I couldn't quite get them. But by picking out two queer looking syllables, the sixth and the tenth, I was soon able to fill in those between."

MEANINGS OF TERMS AND RELATIONS OF PARTS

Subjects report that the nonsense syllables take on more meaning during recitation. Some feel that in merely reading they take the syllable as it stands; they may notice its form and position but they do not try so hard to make it mean something. The meanings come out more clearly when they are forced to reconstruct it in recall. The kinds of meanings are various. Sometimes it is a farfetched resemblance to some familiar word, such as *toq* = *toque*, *soy* = *say*, etc. Sometimes two words are combined to form a single word, such as *sor-dit* = *sordid*, *jor-kih* = *jerky*. Often a resemblance to a familiar foreign word is seized upon, *qos* = Latin *quos*, or a word is associated with a foreign equivalent, *dit* = French 'he says'. Again a syllable is employed as part of some familiar word, as *gov* in governor, and still more common were associations between the syllables and the 'nicknames' of known persons. Sometimes the recurrence of words having a similar look or sound is noted, such as *toq* and *doc*, and occasionally the first letters of successive syllables are combined to form a new word. Sometimes the associations are less definite; the syllable merely feels big, or dull, or bright, or buzzy, *e. g.*, *viz* feels 'buzzy', likewise *zop; dit* is short and snappy, *qos* seems to be 'such a mouth full'.

In the case of sense material, recitation leads to a more thorough understanding, both of the minor details and of the meaning of the thing as a whole. They size up the men described more definitely. One subject reports, "In reading I was dealing more with a lot of details, which I handled mostly in a verbal way. There was no flesh and blood about the men. But during recitation, I could really picture them as men of [such and such age, size, etc.]." It appears that this better grasp of the meaning of the material is an aid to memory. In this connection Meumann writes:[3] "In the case of coherent and meaningful material the chief memorial support consists in the apprehension of the meaning and the logical context."

CHANGING METHODS DURING A STUDY PERIOD

A few subjects reported that they believed one advantage of recitation was to be found in the fact that they could shift from one kind of imagery to another more readily. In reading they were more likely to depend on visual imagery, or, as they reported, to use no imagery at all, but simply look at the syllables, while in recitation they would employ now one sort of imagery, now another, or more accurately emphasize different sorts of imagery at different times. One subject reports: "Sometimes I tried to recall by seeing the words in my mind's eye, and sometimes by trying to remember how it sounded, and again by trying to say several words quickly without imagery. I think this helped since it made the work more interesting and allowed me to resort to different methods when I got stuck." This shifting from one method to another may have made the work more absorbing, but its general value as an aid in

[3] *The Psychology of Learning*, p. 297.

learning may well be doubted. At all events, the greater freedom to employ any method that seems desirable is a notable characteristic of learning by recitation.

PATTERNS AND GROUPINGS

Closely connected with the previous finding that recitation leads to better articulation, accentuation, pauses, vocal inflections, use of melody and rhythms, as well as to better localization, noting of peculiarities and meanings in the material, is the finding that recitation tends more toward a division and grouping of the material. In reading, the syllables are handled more as isolated terms; the learner tries to imprint each by itself. In recitation more of an attempt is made to make the material over into some sort of pattern, a more or less highly organized structure. The patterns differ greatly among individuals and vary according to the list of syllables used. Very often the structure is decidedly of a rhythmic character, associations being formed between accented terms, their positions and pauses, as we have seen. In these cases the associations between members of a given foot are particularly strong, and the feet, although they are in the beginning relatively independent, are bound together in various ways. Sometimes the groups are of unequal length, being determined by the location of syllables which for various reasons stand out prominently. More often, of course, the groups are of equal size, including from two to six syllables, usually three or four.

Subjects report that this active process of dividing up the material and making it over into groups is more easily done in recitation. It is, however, very often done in reading also, but it is then more difficult to do; the divisions cannot be made so sharply, and the ease of reading down the series defeats their purpose. For example, one subject (*Bn*) whose results were very poor in the reading tests, said: "A certain amount of reading is valuable to get acquainted with the material and to frame up a method of attack, but thereafter it seems to do me no good. I simply can't learn by more reading, except by taking a small bit of the series, giving it special attention at one time and later going through it very hurriedly. The desire to look away from the paper to see if I can recite the material is well nigh irresistible." This 'going through it very hurriedly', which the subject speaks of, is probable a very close approach to recitation.

It thus appears that in the reading series the material is handled more by separate items than by groups. Less effort is used to build up a structural whole—there is less organization of the material. Subjects *S* and *T* show in another way an advantage of recitation which is dependent upon better organization of the material. The following figures give the number of seconds for each of a number of repetitions in several tests.

Subject T reading

15, 12, 11, 12, 12, 12, 14, 13, 11, 15, 16, 13, 11, 9, 15, etc.
14, 13, 10, 12, 11, 15, 14, 17, 13, 13, 17, 9, 14, 13, 11, etc.
16, 10, 12, 13, 14, 12, 11, 16, 14, 9, 11, 12, 16, 12, 11, etc.

Recitation

37, 45, 62, 20, 45, 12, 45, 36, 6, 50, 35, 4
27, 5, 47, 52, 46, 8, 31, 45, 33, 12, 6
47, 27, 53, 12, 34, 5, 34, 2, 26, 53, 35

Subject S reading

24, 24, 18, 24, 32, 22, 25, 34, 26, 30
18, 14, 16, 18, 14, 20, 25, 23, 24, 19

Recitation

82, 90, 42, 72, 12, 87, 36, 12, etc.
72, 80, 36, 8, 46, 90, 42, 6, 45

In the first place it will be noted that the rates for readings are very uniform. The subject reads and rereads in much the same way, giving as we have seen about equal attention to all syllables. But in the case of recitations, the rates of the repetitions are varied, the average rate being slower with a much higher mean variation. The subjects were able to account for this, in part at least. Usually the material was divided into groups, different ones being featured at different times. To begin with, the first group was hit hard, perhaps also the last group, with the result that these two groups were earliest learned. When these were fairly well under control, attention was given to the second group, and so the learning progressed. The variations in the total time for repetitions are due to the varied treatment of some of the groups. Usually a group was perused very slowly when it first became an object of attack and once having been fairly well mastered was passed over very rapidly, except that now and then a more lengthy and more thorough review might be given.

Of special interest and importance are the very short repetitions of four, five, six, eight, etc. seconds which occur at various intervals, being more numerous near the end of the study period. The subjects reported that these amounted to very hasty reviews of the whole series. In the beginning they served the purpose of providing a better acquaintance with the material as a whole, while later on they usually amounted to very hasty surveys of the material already learned, either with or without much attention to the unlearned syllables. They served a two-fold purpose of economizing time and of working over the lesson as a whole. In the latter capacity they served the purpose of building up associations between the various groups of items and perfecting the organization of the whole structure.

G. E. Müller,[4] who has made an extensive study of learning methods, describes in the course of memorizing series of digits, nonsense syllables, etc., several stages in the organization and grouping of the material. With simultaneous presentation, the first stage is a 'collective apprehension' of the row of items. This stage affords opportunity to secure an acquaintance with the material generally and to observe such near-lying cues as there may be that can be employed in dividing up the material for further learning. A second stage is called 'collective successive apprehension', which consists of 'a speedy perusal of the individual members of the complexes with attention'.[5] The result is that 'the two successive members of one and the same group are bound together by associations stronger than the associations between successive members of different groups'. This is followed by a third stage, which consists of an 'inner reconstruction' of the earlier apprehended groups. Usually recitation is the chief constituent of the third stage. The subject endeavors to reproduce the material without looking at it, and this leads to the employment of the various kinds of aids that have been previously mentioned. The learner must select the bonds that are requisite to reproduction and exercise them until, once set into operation, they will run their course without external assistance. Of course, during the recitation, references may be made to the text for purposes of prompting as well as for review of material already partly learned. But the 'inner reconstruction' of the material is the important function. Kühn observed as the most serious deficiency of learning by reading the almost unavoidable tendency to neglect many of the functions which are essential to recall, functions which as a rule can operate only in voluntary recall. He writes:[6] "Therefore we come to the conclusion that recitation is better because it leads to a more fundamental, many-sided working-over ('Verarbeiten') of the material."

The typical learner, we have seen, breaks up the material into smaller groups which are dealt with as units. Similar to the present findings, Kühn noted that such manipulation of the material was more characteristic of recitation. He states: "By learning with recitation the construction of groups can be carried on more readily than through reading. Many persons say, in fact, that in really pure reading such a construction of groups is impossible."[7]

[4] 'Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufes', *Zeitschrift für Psychologie*, 1911, Supplementary vol. 5, pp. 253–403.

[5] *Ibid.*, p. 254.

[6] 'Über Einprägung durch Lesen und durch Rezitieren', *Zeitschrift für Psychologie*, 1914, 68, p. 443. [7] *Ibid.*, p. 440.

The manner in which these groups are built up, the determination of their number, size, and distribution, has already been described. In general, the nature of the grouping depends upon the kind and length of the material and upon a host of peculiarities which may be found within it. Great differences are also found among different individuals and for the same individual at different times.

The value of such groupings of the material as an aid in learning has been pointed out by Müller. They are in brief:

1. Although it is impossible to grasp in one span of attention a whole list of items, the smaller groups can be utilized as units for attention, thus leading to economy of time and energy in apprehending the whole group.

2. The factor of localization comes more effectively into play. One cannot remember the positions of each member of a series of twenty nonsense syllables, but he can remember the position of four or five groups, each being treated as a unit.

3. Each group comes to have its own individuality and thus serves as a center of attack.

4. Groupings assist rhythmical and melodic perusal.

5. Groups as such are more interesting than a series of single items which the learner soon becomes familiar with, as such, and then permits attention to flag. The groups, as interesting problems to be mastered, arouse and direct attention.

When the series is quite long, it is not enough that the individual groups should be mastered, but the series of groups must be bound together by additional associations. Sometimes the localization of the groups in visual space or numerically is sufficient, but very often other associative or mnemonic aids are employed.

Our previous analysis of the learning process would fit very nicely into Müller's scheme of three stages. That the reading method should be employed to some extent in the beginning has been pointed out by Müller—in fact, the first two stages are entirely dependent upon reading. The third stage of 'inner reconstruction' is, as its name implies, primarily a stage of attempted recitation.

To limit the learner entirely to the reading method precludes the possibilities of the active stage of 'inner reconstruction' and thus greatly hampers the learning. The natural tendency of the learner to resort to this latter method of study is shown in the oft repeated statement that the desire to do so was 'well nigh irresistable' and the like. Most subjects can, to varied degrees, continue to learn by reading, but there are some, perhaps, who can advance only to a limited extent. Kühn found,[8] in fact, that after a certain number, additional readings may prove not only to be of no value for imprinting, but may be positively harmful. For example, one subject (*Got.*) required after

[8] *Op. cit.*, p. 477.

40 readings,	17 additional recitations to learn
25 readings,	9 additional recitations to learn
12 readings,	6 additional recitations to learn
2 readings,	5 additional recitations to learn

Similar results were found for three other subjects. Such tests, however, have been tried with several subjects in the present study, but in no case were such negative results found, although two subjects were found who were unable by reading alone to completely master a long series of nonsense syllables.

HELPS MORE CONSTANT AND MORE NATURAL IN RECITATION

It was pointed out earlier that recitation leads more successfully to the employment of various sorts of aids, such as modulations of the voice, rhythms, pauses, meaningful associations, and the like. An additional point very often reported is that such aids not only come into play more readily in recitation but that they are more constant. During reading, some report that they emphasize now one syllable, now another; they now use one rhythm, later another; the sight of the word suggests now one association, later another. In recitation, when once adopted, the aids are more constant. This is partly due to the fact that most learners do not like to refer to the text unless it is absolutely necessary, and since recall is entirely dependent upon the use of some association, a connection once initiated is likely to be invariably employed. During reading, since the syllable in each case is present to the eye, the previously observed association, being less essential, is not so deeply impressed; other connections, depending upon the attitude of the subject at the moment, are likely to overrule it with the result that a new association is substituted. This, in essence, is what many report: "It is hard to keep my mind on the work in reading. Different influences seem to come in continually that give the material a new look. First a syllable means one thing and later I associated it with something else." Subject *Tr* says: "I first thought of *fab* as part of *fable*, *wab* as *Weber*, etc., but it was often difficult to remember some of them because I didn't have to depend upon them." A similar situation was found in the case of many subjects by Kühn, who concluded:[9] "The helps in recitation seem to be more natural, while in reading they appear manifold and artificial."

TESTING THE LEARNING

In an earlier section, evidence was found that there was a greater certainty as to what was known when recitation was employed in the learning. This, of course, is not only true at the completion of

[9] *Op. cit.*, p. 440.

the learning but during the various stages. In addition to the objective evidence already presented (p. 71 f.) are the reports of many subjects that when they read only, they are not at all certain how much of the material is known or how well it is known. They may have a general feeling that they can recite a certain part of the material, but they cannot be sure until they have tried. The recitation, of course, constitutes the test.

One of the values of recitation is that it gives exact knowledge of the results that are being produced and serves to throw into relief the efficacy of the different aids that are being employed as a means to learning the lesson. Recitation leads more surely to the selection and repetition of the desirable bonds and to the elimination of the unfit. In other types of learning, Judd has shown that knowledge of results of practice is essential to improvement.[10] He found that practice in locating the continuation of sloped lines, part of which was concealed from the subject, produced no improvement when the results of the practice were not disclosed, but improvement immediately resulted when the subject was permitted to view briefly the results of his efforts.

In an earlier section (see p. 71 f.) it was found that recitation leads not only to better localization of the syllables but it also leads to a more accurate knowledge of the correctness of the position of syllables. In learning by recitation, out of 127 judgments of 'correct in position' but two were wrong, while in the reading series nine out of sixty-eight such judgments were wrong.

In order to find if there is a greater certainty with respect to the form of the syllables without regard to their position, another similar experiment was made. Eleven graduate students acted as subjects in two tests of five minutes each, one by the reading method and one by the recitation method. Half of the subjects took the former and half the latter test first, the order being reversed for the second test. Table XXXIX gives the results.

In the first place, a greater number of syllables are written down after the recitation test than after the reading test (twelve and two-tenths as compared to eight). The absolute number judged correct in the recitation series is about twice the number so judged in the reading series, eleven and three-tenths as compared to six syllables. Likewise, the number of syllables that were actually correct was about twice as great for the method including recitation, eleven and four-tenths as compared to five and fifty-four one-hundredths. Of the total number of syllables written down in the reading series seventy-five per cent. were judged to be correct, while in the recita-

[10] 'Practice without Knowledge of Results', *Psychology Review Monographs*, 1905, **7**, pp. 185–198.

tion series ninety-two and six-tenths per cent. were judged to be correct. That is to say, there was a greater assurance of correctness when the learning involved recitation. Moreover, in the recitation tests, of those written down ninety-three and four-tenths per cent. were actually correct as compared to sixty-nine and two-tenths per cent. for the reading series, indicating again that there is less certainty about the knowledge of results during reading. It should be noted that in the reading series there is a considerable discrep-

TABLE XXXIX

Results given in number of syllables correct in form without regard to position

Subject	After five minutes reading			After five minutes recitation		
	Number written	Number judged correct	Number actually correct	Number written	Number judged correct	Number actually correct
At.	13	13	11	16	15	14.5
Rs.	8	8	6	16	16	16
Py.	9	7	7	10	9	9.5
Tr.	7	5	5	15	14	14
Gl.	8	4	5	10	10	10
Mn.	11	6	7	12	10	10
An.	8	4	5	12	11	11
E. M.	5	4	4	10	9	9
J. M.	7	6	4	11	9	10
Sn.	7	5	4	16	16	16
Sa.	5	4	3	7	6	6
Average	8.0	6.0	5.54	12.2	11.3	11.4
P. E.	1.4	1.5	1.3	2.0	2.3	2.2
Per cent. of number written		75.0	69.2		92.6	93.4
Per cent. of number judged correct			92.3			100.0

ancy between the number of syllables 'judged correct' and the number 'actually correct'; while for the recitation method these two figures are almost identical. This means that after you have studied a lesson by the recitation method you are practically certain how well you know it, but after you have studied by reading you are not only uncertain about your knowledge but your honest opinion is likely to be an overestimation of your attainment. A closer examination of the table, however, will reveal the fact that individuals differ in this respect. Under the reading method, three people correctly estimate their knowledge (*i. e.*, the number of syllables 'judged correct' equals the number 'actually correct');

five people overestimate their knowledge; while three underestimate their knowledge. Confronted by this general uncertainty of results, some subjects are likely to be very conservative in their judgments and others much less so, the general result being an overestimation of attainment.

In the case of the recitation series, eight subjects correctly estimate their knowledge, while one overestimates and two underestimate their knowledge. The sum of the differences between the number 'actually correct' and the number 'judged correct' is two syllables for the recitation series and eleven for the reading series.

On the whole then, learning by reading makes it very difficult to estimate one's attainment, while learning through recitation leads to very accurate knowledge of results. This should be thought of in connection with the fact that in our tests the amount learned by recitation is about twice as great, a fact which can only emphasize the greater accuracy in that case. Other things being equal, we should expect twice as many errors of judgment in the recitation results.

Some evidence can be obtained from the children's data to indicate a similar result. From the data of several classes was computed the total number of syllables written down, and the total number of syllables that were correct in form. From the various methods of study including recitation, certain ones were chosen in order to make practice effects, etc., balance up with the reading series. The following is a sample result, based on forty pupils of the sixth grade.

	Reading		Recitation	
	Written down	Correct	Written down	Correct
Number of syllables	7.12	4.5	11.12	9.06
Relative number	100.0	63.2	100.0	81.4

The results show a clear superiority in favor of the recitation method of learning.

A similar result was found with sense material, a sample of which follows, showing in the case of thirty-nine eighth-grade pupils the number of details of facts written and the number correct.

	Reading		Recitation	
	Written down	Correct	Written down	Correct
Number of facts	27.2	22.6	28.7	26.1
Relative number	100.0	83.1	100.0	90.9

Objective data, thus, support the introspective opinion, previously given, that one has better knowledge of results in learning by recitation and that this is an aid in learning. Some of the concrete ways in which this knowledge may be of assistance may be briefly considered.

First: There is a feeling of satisfyingness in the certainty of progress, in knowing that headway is actually being made. Conversely, it is annoying to be uncertain whether the study is bringing returns. The satisfyingness results in better attention and better application to the work, while annoyingness is distracting and hampers learning. Subject *An* gives a typical report: "It [reading] was discouraging because I did not feel that I was making much progress during the last part of it. There was no way to tell."

Second: A certain saving of energy may result from knowing what parts of the material are known and what are not known. *a*. Overlearning of certain portions may be prevented. Usually the first and last syllables are first learned and when the subject knows that these are mastered, they can be passed over hastily in subsequent perusals, a very slight amount of review being sufficient to keep them intact. Subject *Rs* says: "I saved time during recitation by skipping hurriedly over the words I already knew." *b*. An opportunity is afforded to direct special attention to those portions that are still unlearned. Subjects report that certain syllables offer special difficulty which is often not suspected until they endeavor to recite. *c*. The two factors together, easing down on familiar or learned portions and attending more intensely to unfamiliar or especially difficult portions, result in a saving of energy in the long run. It makes the work more absorbing, and also makes possible short periods of relaxation of attention or breathing spells, which may result in a rebound of energy for learning the more obstinate portions.

ERRONEOUS RECALL

It is obvious that an attempted recitation may result not only in a failure to recall a certain syllable, but it may also result in erroneous recall, neither of which could occur during reading in the strict sense. If the errors are too numerous or if they are not discovered in the case of recall, they become a harmful rather than a beneficial factor in learning. Failures to recall are very frequent in some cases in which the recitation is introduced too early, with the result that time is frequently lost in unfruitful endeavor to recall items that are not as yet sufficiently fixed in mind. Erroneous recalls, under the same conditions, are also frequent, but they seldom occur without some feeling or indication of incorrectness.

Many subjects report that they profit much by these mistakes. Noting and correcting an error helps to fix the proper item in mind; it receives better attention at that moment and will receive special attention on the next repetition. Subject *Rs* said as a sample concerning a test in which recitation was begun after five minutes' reading: "Twice I failed on *puv*, each time saying *poy* [evidently confused with *soy* which followed]. But after twice correcting it, I had it so well in mind that I will probably remember it longer than any other in the series."

In connection with the matter of unsuccessful and erroneous recall, Katzaroff pointed out, as an advantage of recitation, a kind of growing satisfyingness in the task. Successful recall is satisfying and failure is annoying. As we proceed, the proportion of satisfyingness becomes steadily greater, toning up the learner and enabling him to keep up interest and application in spite of growing fatigue. He states:[11] "The learner is active, he has to seek, he rejoices when he has found and is irritated at the syllables which evade his call. Here crowd sentiments of affection for certain syllables, of antipathy for others, which contribute to enrich the associative bonds and favor conservation and recall." One of the workers in the present study similarly said: "In reading, it is the last part of the test that is most wearisome, but in recitation, it becomes almost a pleasure as I approach a mastery of the whole bunch of words."

UNINTENTIONAL RECITATION DURING THE READING TESTS

A great many of the subjects found it difficult to resist their natural tendency to recite, during the reading series; in fact, the reading was nearly always combined with more or less recall of an unintentional, practically unavoidable sort. The effort to avoid reciting acted as a positive disturbance and source of annoyance, thus distracting attention and consuming energy to no purpose. Subject *Md* speaking of the reading method said: "Very difficult and disagreeable, because I was constantly inhibiting the tendency to test what I had been trying to learn."

SATISFYINGNESS AND ANNOYINGNESS IN READING AND RECITATION

According to the introspections of many adult subjects and the reports of many school children, one conspicuous difference between reading and recitation lies in the greater satisfyingness of the latter. That the matter of satisfyingness and annoyingness of mental work is important has been emphasized by Meumann.[12] "The emotional condition in which we find ourselves during the performance

[11] *Op. cit.*, p. 257.
[12] *The Psychology of Learning*, p. 281.

of a mental task is of profound importance for the accomplishment of the task. In general, it may be said that an emotion of pleasantness facilitates the function of memory, and that unpleasantness has a very detrimental effect upon memory." Thorndike is more cautious:[13] "No one probably doubts that interest in the exercise of a function favors improvement at it," and "such statements appeal to our common sense as probably true, though they have not been fully verified."

It shall be our purpose, first, to inquire as to what differences appear between recitation and reading as producers of satisfyingness and annoyingness, and then to consider briefly in what way or by means of what minor functions these effects are brought about.

That there is greater satisfyingness in studying by the recitation method is indicated by the witness of nearly every subject, child or adult. At the close of the experiments with the school children they were asked to state what method of learning they liked best. For ease of selection the cases considered were three: one in which they read all the time, one in which they read about half of the time, and another in which they recited nearly all the time. The following table gives the distribution of opinion.

With nonsense material

	All reading	Half and half	Mostly recitation
Grade eight	3	2	29
Grade six	2	10	27

Sense material

	All reading	Half and half	Mostly recitation
Grade eight	4	17	20
Grade six	2	10	28

It is clear that the children strongly preferred the methods in which recitation was included.

The reasons for their preferences are varied and not very specific. Such statements as, "It isn't such hard work," "I learn better that way," were common. Some explained their preferences as follows: "I knew I was learning them when I recited"; "I get so tired when I read"; "When I recite, it's fun to see if I can say more every time than I ever did before."

The introspective accounts of adults are even more emphatic. Among the subjects listed in Table XXXIII, fourteen reported that

[13] *Educational Psychology*, vol. II, p. 219.

Method Three (two minutes reading and eight minutes recitation) was most satisfying, one that Method Two (half and half) was most natural and satisfying, and all reported that Method One (all reading) was least so. In fact, most of them report that the last four or five minutes in the reading test were positively 'annoying', 'monotonous', 'tiresome', 'very fatiguing', etc. Subject *Hn* declared: "Without a doubt, trying to learn a series of nonsense syllables in this way is the most monotonous work I have ever done. The syllables came to have absolutely no connection or association, and the typewritten letters became, after four or five minutes, so many stupid hieroglyphics."

Many of the actual ways through which reading becomes annoying and recitation satisfying have already been indicated, and they will receive but brief mention here. *a.* There is satisfaction in the realization that progress is actually being made. We have seen earlier that this is the case during recitation. Conversely, it is annoying to be uncertain of one's progress in the learning. *b.* Recitation is satisfying because it offers the learner more freedom to employ such aids, and work with such methods, as he may desire. Reading becomes annoying because it hinders or prohibits the exercise of many of the desired functions. *c.* The facts of *a* and *b* taken together explain other sources of satisfyingness in recitation. For example, it is satisfying, as Katzaroff pointed out, to attack portions of the lesson that offer special difficulties—difficulties that are often not realized until one begins to recite. Again, the opportunity that recitation affords the learner to ease off on familiar portions, and strike hard at difficult portions, seems to be a good remedy for boredom and fatigue. *d.* Annoyingness attends the constant effort exerted by many in resisting the natural tendency to recite during the reading series.

IS THERE GREATER ACTIVITY IN RECITATION THAN IN READING?

The early investigators on this subject gave great emphasis to the conclusion that recitation, as compared to learning by reading, produced a greater activity on part of the learner, and to this greater expenditure of energy was attributed in large measure the better results obtained. For example Katzaroff says:[14] "In the readings, the subject is passive, calm, indifferent; in recitation he is active."

The introspections and observations from the present work do not lead to exactly this conclusion. The distinction seems to be one of kind rather than one of quantity. It appears that recitation does not always, in fact, does not generally result in greater activity, effort, or expenditure of energy on part of the learner, but the indica-

[14] *Op. cit.*, p. 257.

tions are that the energy is expended in a different way. Certainly the conscientious learner by the reading method is not 'calm and indifferent'. An apparent, but certainly not a real lack of activity is indicated by the already mentioned fact that many of the motor functions such as articulation, accentuation, the use of rhythm, etc., are much less prominent in reading. In the next section, however, it will be seen that the subjects declare that every internal symptom indicates that reading is more consuming of energy than recitation.

FATIGUE EFFECTS OF READING AND RECITATION

Other things being equal, we should expect, were it true that recitation results in greater activity and expenditure of energy than reading, that it would also be more fatiguing. Unfortunately, in this study, we have no indisputable measure of fatigue, but it nevertheless appears, in so far as one is able by a subjective judgment to estimate fatigue, that recitation is much less fatiguing. The findings reported in the section on the satisfyingness and annoyingness of the two methods bear strongly on this point. In so far as feelings of fatigue, boredom, monotony, and the like are indices of real fatigue, there can be no doubt that recitation is less fatiguing than learning by reading. Whether or not they are measures of *real* fatigue, they are at least very important from the point of view of work in the school-room. In the face of such statements as those following, there can be no doubt that recitation is to be preferred to learning by reading in this respect. Subject *Hn:* "Reading is the most monotonous work I have ever done." Subject *Dn:* "Reading is most fatiguing because there is no variation." Subject *Fx:* "Reading most fatiguing—monotonous—took all my energy to keep up interest." Subject *Sn:* "This method very tiresome—effort seemed to be fruitless." Subject *Mk:* "Very tiresome and disagreeable." On the other hand, learning by recitation may, as Subject *Py* said, "be almost a pleasure," or, as Subject *Mk* states, "much more satisfying," or, as Subject *Rs* says, "not so bad as reading, that's certain."

Subjects also report that the after-effects of learning by reading are greater than learning by recitation. Subject *Bn* reports: "I couldn't apply myself to work for an hour after the experiment.' Subject *T:* "I felt tired all the rest of the afternoon."

From a practical point of view, it should also be considered that the fatigue, based on the amount learned, rather than the time spent, would be relatively very much greater in the case of learning by reading. The subjects report that they are very much less fatigued by ten minutes of study by recitation than by ten minutes study by reading, yet they have learned twice as much. If the study by reading were continued until the amount learned was equal to that learned by the recitation method, the fatiguing effects of the former would doubtless be still more marked.

SUMMARY AND CONCLUSIONS AS TO THE NATURE OF READING AND RECITATION AS FUNCTIONS IN THE LEARNING PROCESS

Our analysis of learning has shown the memorization of any material, especially of nonsense material, to be a complex process involving the formation of a host of bonds. It has appeared, moreover, that many, in fact, most of these bonds can be properly formed only by means of recitation. Our subjects have reported that it was 'difficult', 'unnatural', 'annoying', 'fatiguing', or 'impossible' to establish most of the essential bonds during reading. Efforts to learn the material by rote, to memorize it mechanically or by means of 'visual imprinting' during a series of readings proved to be futile. Memorization was possible only by means of establishing bonds between items and their pronunciation, sound, or look: between items and accents, pauses, or elements of a rhythm; between items and their position in a series; between an item and other items which it may be considered a part of, similar to, or somehow related to and the like, as well as additional bonds between characteristics of successive groups of items. We have found that it is to the formation of just these bonds that recitation leads, and that it is just these functions that it is difficult or impossible to exercise adequately during reading in its pure form. Consequently it seems to be a justifiable conclusion that complete learning is possible only by means of some form of recitation. *Pure* reading alone will scarcely enable one to completely learn a lesson which exceeds the memory span by any considerable length, yet it serves an important function in the learning process as we have seen.

The considerations of the present chapter have shown that reading and recitation are very broad functions made up of many minor ones. Economical learning consists not only in selecting and exercising those more minute functions which are essential and eliminating those that are valueless, but also in exercising them in proper sequence and each for an optimum time. It will be necessary here to review but briefly some of the essential functions, indicating to which of the two broader functions they belong and the order in which they are customarily exercised.

The first stage of the learning consists, as we have seen, in looking over the whole material with the purpose of obtaining an idea of its general make-up, noting the individual items in the group, getting the pronunciation or look or sound of the terms to some degree, and noting outstanding 'aids' which may be employed in breaking up the material to further the learning. Much may be done in this stage to determine upon a method of attack. The length of the material, its apparent difficulty, its peculiarities, the possibilities for rhythmical division, and its ready-made associations are considered. The material may be thus perused for several times until the reader feels 'familiar' with it and a method of procedure is tentatively adopted. This stage is the reading stage, including what Müller has termed the stages of 'collective' and 'successive apprehension'. That the functions of reading are essential and satisfying here has been indicated by all introspective data.

The optimum duration of this stage depends upon many factors, such as the length and difficulty of the material, the age, training and capacity of the learner, and the like. These considerations will be taken up on a later page.

Following this stage, new functions may be introduced and the original functions may be employed in a somewhat different manner. This is the stage of recitation. It consists essentially in the final selection of the bonds requisite to recall and the exercise of these bonds until they are firmly established. What these bonds are, it was the purpose of the preceding sections of this chapter to point out. Thus it appears that memorizing is in no essential way different from any other form of learning. The bonds selected are exercised, those found to be unfit are eliminated, and new bonds are added as the case demands, the period of practice being continued until, once initiated, the series of desired responses runs off in the proper order.

Like other processes of learning memorizing may be explained in physiological terms. An adequate explanation of this sort would make the difference between the functions of reading and recitation more intelligible. The learning of a series of nonsense syllables, like the formation of any habit, involves two things: a sensori-motor response or the formation of a bond between a situation and a particular response, and a sequential connection between the various situation-response bonds in serial order.[15] The following diagrams illustrate in a very rough way, what physiological actions and changes are involved in the learning of a series of nonsense syllables or any other material.

[15] A standard treatise in English upon the physiological aspects of learning is Ladd and Woodworth, *Physiological Psychology*, New York, 1911. For an abbreviated but excellent account, see also Thorndike, E. L., *Educational Psychology*, New York 1913, vol. 1, chapter XIV. The illustrations used in the present article are similar in some respects to those employed by Bair to explain the development in skill in typewriting. Compare 'The Practice Curve', *Psychological Review Monographs* 1902, No. 19, pp. 1–70.

In Figures 4, 5, and 6, S, S^1, S^2,—represent the stimuli, that is, the sight of the syllables, occurring in serial order as they would in reading. These stimuli are conducted to the sensory centers A, A^1, A^2,—, which discharge respectively into the higher centers B, B^1, B^2 —, and these cells in turn discharge into M, M^1, M^2 —, the effectors which produce the motor responses of writing or speaking the syllables.

Let us consider a case of *pure* reading; *pure* in the sense of being entirely devoid of all elements of recall, waiving for the moment

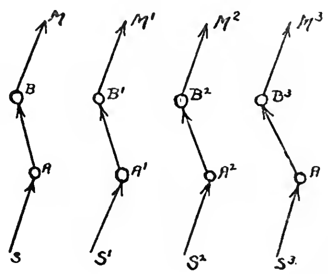


Fig. 4—'Pure' reading.

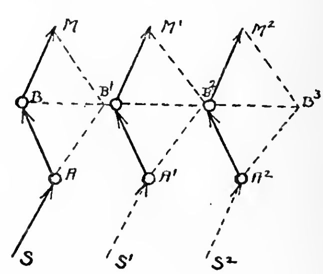


Fig. 5—Reading with formation of associations.

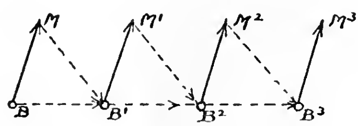


Fig. 6—Recitation.

the question of whether such reading actually exists. *Pure* reading would consist in the exercise of bonds S A B M, $S^1A^1B^1M^1$, etc., as distinct units. The more often these bonds are exercised, the more definite becomes the connections and the more automatic the response. But it is obvious that however firmly these bonds become fixed, they cannot of themselves make possible voluntary recall, since S, S^1, S^2, —, (the sight of the syllables being learned) is an essential link in the process.

Figure 6 illustrates roughly the requirements for voluntary recall. In this case the expression of the syllables (designated as M, M^1, M^2, —,) are produced in the absence of the stimuli, S, S^1, S^2, —, of the visible words. What is required here is that bonds should have been formed between the various higher units. Connections between B and B^1, M and B^1, or both, must be established. The result is that once the series is started, the physiological processes which produce the first syllable act as the stimuli for the production of the processes which bring about the response of the second syllable and so on.

Perhaps few would doubt that this illustration would account, in a very rough way, for the process of recall, but many might be unwilling to admit that Figure 4 is a correct representation of the processes involved in reading. The doctrine of association by contiguity might insist that the mere repetition of the syllables one after another would result in the establishment of bonds between them. Bair,[16] consequent to his study of the development of skill in typewriting, concluded, although he really gave very little weight to it, that "connections are formed between cells that for a number of times have been stimulated or discharged—in succession." Woodworth has pointed out the inadequacy of this doctrine. To quote:[17] "contiguity is a necessary condition of association. But is it a sufficient condition? There is little in the experimental work on memory to indicate that it is sufficient, and much to indicate that it is not usually depended on to accomplish results. The things to be connected must be together, in order to arouse the reaction connecting them; but, unless they arouse some such reaction, they do not become connected, except it be very weakly." Professor Woodworth has shown some convincing experimental evidence[18] in support of his view and doubtless much more could be discovered by search through studies already in print,[19] but space will not permit us to go into the matter here.

In attempting to learn by reading, the subject does not rely entirely upon mere repetition of the syllables—upon the alleged efficacy of contiguity—alone, but in most cases, tries to form the serial associations upon which he must rely to recall the series when the time comes. Why are these bonds not definitely formed? The reason is that the presence of the printed words (S, S^1, S^2, etc.) makes it so unessential, during reading, to connect B with B^1, or M with B^1, that the learner's purpose to strengthen these bonds is defeated. Since, according to prescription, the learner must,

[16] *Op. cit.*, p. 51.

[17] 'A Revision of Imageless Thought', *Psychological Review*, 1915, **22**, pp. 1–27, especially pp. 16–22.

[18] It may be well to give the following sample test, in the words of the author. "I read a list of twenty pairs of unrelated words to a group of sixteen subjects, instructing them beforehand to learn the pairs so as to be able to respond with the second of each pair when the first should be given as a stimulus. But, after reading the list three times, I told them that they should, if possible, give also the first word of the following pair on getting the second word of the preceding pair as a stimulus.—The results were most definite: the second members of the pairs were correctly recalled in seventy-four per cent. of all cases, but the first members were recalled in only seven per cent. of the cases." 'A Revision of Imageless Thought'. *Ibid.*, p. 18.

[19] For example, see Hollingworth, H. L. 'Characteristic Differences between Recall and Recognition', *American Journal Psychology*, 1913, 24, pp. 532–544. Also 'The Influence of Caffein on Efficiency', *Archives of Psychology*, 1912, No. 22, p. 17.

on completing the series $S\ A\ B\ M$, then read S^1, the connections $S^1\ A^1\ B^1\ M^1$ being by previous practice better established, are thrown into action before the incipient bonds $B–B^1$, $M–B^1$, are awakened. The response follows directly upon the stimulus provoked by seeing the word.

How is it possible, then, as shown by our quantitative results given earlier, that some memorization does result from reading? In all probability *pure* reading is a fiction; recall, to some degree, being always present. Nearly all subjects were able, introspectively, to discern this fact. Figure 5 shows, roughly, the physiology of this situation. The dotted lines $S–A$, $S^1–A^1$, etc., indicate that these bonds between the sight of the word and its expression are more feebly exercised; are less depended upon than is the case in Figure 4, which illustrates the hypothetical pure reading. The manner in which the items (S^1, etc.) are required to play a minor role are various. Sometimes the subject pauses between the series, $S\ A\ B\ M$ and $S^1\ A^1\ B^1\ M^1$, etc., thus permitting the bonds $B–B^1$, $M–B^1$, etc., to be thrown into action before S^1 is observed. That is to say, the subject anticipates the next word in the series, more or less, before he reads it. Sometimes the syllables are read in a hazy, inattentive way, in which case the subject relies partly upon the exercise of the serial bonds as well as upon the objective stimulus of the printed word. In these and other ways, *actual* reading departs from *pure* reading and in consequence leads more effectively to memorization. In short, the actual reading which the subject practices is a sort of hybrid between the hypothetical pure reading and recitation.

A more accurate picture of the anatomical substrata of memorizing would undoubtedly be much more complex than our simple diagrams. As we have seen, consciousness of the meaning and form of the material is a prominent factor in learning. Consequently, the diagram should contain at least a symbolic representation of the centers upon which, presumably, such consciousness depends. Thus, in Figure 7, P, a 'psychic' center may be added, in which elaborations of the sensory data take place.[20] Probably in learning a passage, as well as during the recall for some time, P is called into activity, discharging into B. As practice continues, in all likelihood, pathway $S\ A\ B\ M$ becomes relatively more and more permeable, until finally conduction through $A\ P\ B$ ceases almost entirely—the process becomes practically unattended by consciousness of meaning. To illustrate this condition for reading (illus-

[20] For an account of such distribution of functions, see Ladd and Woodworth, *Physiological Psychology*, chapters IX and X. The 'association' or 'psychic' areas are given special treatment on pp. 251–263.

trated by Figure 7) we may repeat the statement of one of our subjects: "The typewritten letters became, after four or five minutes, so many stupid hieroglyphics."

In recitation also, the 'psychic' center *P* is involved. Figure 8 is merely a suggestion of the possible connections of this center with others. *P* may have connections with B^1, as well as with P^1, which in turn, is connected with B^1, through which the discharge into the organ of expression takes place. If now we consider that each of these possible connections indicated by a straight line in the figure is a representation of hundreds, perhaps thousands of different neurones that may be employed, the complexity of the

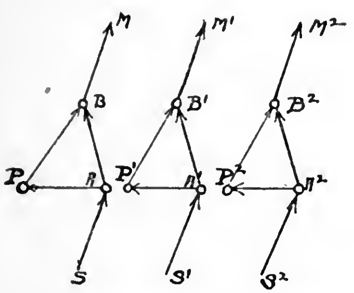


Fig. 7—Reading with awareness of meaning.

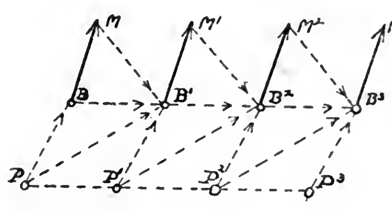


Fig. 8—Recitation with awareness of meaning.

neural substrate involved in learning is suggested. But recitation of the series of syllables may become short-circuited to a nearly mechanical activity, such that once initiated, the series of responses occurs automatically while attention is occupied with other matters. This may be typified by considering that the connections through *P* and P^1, etc., and perhaps even the connections B–B^1, etc., drop out, so that the connection of *M* with B^1, which leads directly to M^1, is so close that once the series of responses is started, each follows its predecessor with mechanical precision.

Recitation, in brief, differs from reading physiologically by the fact that it selects and exercises the bonds upon which the established habit depends, while reading calls into action some bonds that are not strictly needed for recall, omits some that are requisite, and does not so well exercise the remaining few, needed for recall. Recitation is for memorizing what practice is for other habits.[21] The physiological basis is the same.

[21] Such a physiological explanation, for example, has been worked out in detail by J. H. Bair for typewriting. See 'The Practice Curve', *Psychological Review Monographs*, 1902, No. 19, pp. 1–70.

Some of the differences between reading and recitation, which appeared from the introspective analysis of the two functions, may appear with more clarity when considered from the physiological side. For example, many subjects reported that the associative aids adopted during recitation were more constantly employed than those adopted during reading. Recitation results in the continued exercise of particular bonds, as we have seen, and of course each repetition strengthens those bonds, with the result that the nervous impulse once initiated flows along the most frequently traversed pathway. In reading, none of the serial bonds receives adequate exercise, with the result that none has a great advantage over any other, and now one, now another pathway may be traversed.

Annoyingness and fatigue in the case of prolonged reading may be considered as largely due to a check placed in the way of the exercise of the bonds desired. "When any conduction unit is in readiness to conduct, for it to do so is satisfying. When any conduction unit is not ready to conduct, for it to do so is annoying. When any conduction unit is in readiness to conduct, for it not to do so is annoying."[22] After the preliminary exercise of the conduction units *S A B M*, further exercise of that bond becomes annoying; the serial bonds are then ready to conduct. For them to do so, under the conditions specified in reading, is practically impossible, because the stimuli S, S^1, S^2, etc., by virtue of their firmer establishment, cause the conduction to take the habitual course, *S A B M*, etc.

Space will not permit further illustrations of this sort. By way of summary of this section, it is only necessary to repeat that reading and recitation are relatively distinct yet essential functions of the learning process. Each has its proper place, and as we have seen, introducing recitation too early or withholding it too long retards learning. The important matter is to determine the optimum point at which to introduce it, a matter which will receive consideration in the following section.

THE OPTIMUM TIME AT WHICH TO INTRODUCE RECITATION

The quantitative results presented in Chapter IV indicated that the optimum time at which to introduce recitation varied somewhat according to the age and training of the different groups of school children, and the data obtained from adults pointed to differences among individuals of approximately the same age and training. It is obvious that the determination of the optimum time at which to introduce recitation is a most important matter for purposes of economizing time and energy in learning. That the quantitative

[22] Thorndike, E. L., *Educational Psychology*, vol. II, pp. 1–2.

determination of the best combinations of reading and recitation made in the present study apply only to the particular conditions here employed, has already been pointed out. The optimum time for the beginning of recitation will doubtless vary not only according to the age, training, and capacity of the learner but also according to the kind of material, the length of the lesson, and the purpose in view, *i. e.*, whether the lesson is to be learned *verbatim*, whether the substance without the exact form is to be reproduced, or whether a less definite mastery is all that is desired. Consequently, a quantitatively precise rule cannot be made.

The present study, however, has produced some results that are suggestive. In general, all the evidence, quantitative, introspective, and interpretative seems to imply that recitation should be introduced early. Only a very small percentage of the total time required to learn should be devoted to reading. However, it seems to be a natural tendency of many adult subjects to make too early an attempt at recitation. Some of the reasons for the disadvantageous effect of introducing recitation may be pointed out. *First*, The bonds between the words and syllables and their correct pronunciation are not sufficiently well formed to permit successful recitation. *Second*, The advantageous effect of a preliminary determination of a line of attack is foregone in whole or part. *Third*, The amount of data that can be recalled at so early a moment is insufficient. The learner is likely to waste time in fruitless endeavor to recall syllables that are simply not as yet forthcoming. *Fourth*, Too much time is wasted looking on and off the text, 'finding the place' and the like. *Fifth*, Too frequent failures in attempted recalls break attention and may develop an unpleasant attitude on part of the subject. *Sixth*, Too many erroneous recalls may be made. When the learner has such slight acquaintance with the material as a whole, errors once made are likely to be repeated. Later these undesirable bonds must be broken down before the correct bonds can be formed.

Just as introducing the recitation too early has a deleterious effect, so does introducing it too late retard learning. The abundant quantitative evidence for this has been presented in Chapter IV.

The optimum combination of the two functions can be best expressed in this way. Reading should be continued until the learner is fairly familiar with the material as a whole and with the items of which it is composed. The learner should have decided meanwhile upon his general method of attack. Enough of the material should be clearly in mind so that the learner's first attempts at recall will meet with some success. Just how much is enough will depend largely upon the learner. As a guiding principle one may consider that the first few recitations should not result in too great a distortion of the material, nor should it cause a waste of time in fruitless endeavor to recall. The capacity of the learner to quickly judge the status of his knowledge is of prime importance; he should be able to know at once whether continued effort to recall this particular syllable will end in success or not, and in the latter case time should not be wasted before reference to the text is made. The early stages of learning will thus employ both reading and recitation, the relative amount of the former decreasing as the learning progresses. Economical learning would consist, in part, in employing recitation, after it is once introduced, to the full, coupled with the capacity to speedily resort to reading where it is essential.

THE EFFECTIVENESS OF RECITATION IN LEARNING NONSENSE MATERIAL AS COMPARED TO LEARNING SENSE MATERIAL

The quantitative results of Chapter IV seemed to indicate two things: *First*, that the optimum time for introducing recitation was considerably earlier for nonsense than for sense material; and *Second*, that recitation seemed to be a more fruitful method of study in the case of nonsense material than in the case of sense material.

The first result is apparent rather than real. While it is true that recitation introduced very early produced richer returns for nonsense than for sense material, this should be considered in connection with the fact that the amount of material forming the lesson in the former case is but a small fraction of that used in the latter. Yet the amount of material should be considered only in connection with the difficulty of the material. While the nonsense material was much less in amount, it was very much more difficult to learn. A further consideration of this point is unnecessary since the factors which influence the introduction of recitation, just considered, are the same in either case.

The point with regard to the value of recitation as dependent upon the kind of material is important and demands further consideration.

The results have shown clearly that equal amounts of recitation produce richer returns in the case of senseless non-connected material than when connected senseful material is used. The reasons why this should be the case have been given in the previous sections of this chapter, and it is only necessary here to summarize the factors upon which this difference depends.

In the first place, it was found that recitation was of great service in assisting the subject to organize the material into some sort of compact and connected whole, such an organization being essential to a thorough mastery of it. The particular means of accomplishing this organization were: the formation of bonds between the items and accents, modulations of the voice, pauses, and elements of a rhythm; the formation of bonds between items and their meaning, immediate or distant; the noticing of peculiarities in the text and the formation of bonds between items and their position for the purpose of breaking up the material into groups; the noticing of bonds between items and characteristics of the groups and so on. In short, recitation rendered great service in creating usable associations within the material where there was none, or in more adequately noticing and exercising those that were already present. In nonsense material these bonds between items are absent, and this process of organization and creation of associations is difficult and essential; learning of such material consists in accomplishing just this organization. In the connected sense material such as that used in the present experiment, most of these associations are already present; the material is already organized, the items are connected by serial connections of meaning, rhythms, and the like, by means of which the various elements are firmly knit together. The function of recitation for the formation of these bonds is not required. What is needed is that the ready-formed associations be noticed and exercised, although, in most cases, bonds in addition to those found in the material will be required.

A second reason for the better results obtained by reading in the case of sense material is closely related to the first and lies in the fact that reading is less 'pure' in studying sense material. As was remarked earlier, after a certain number of perusals the reading of either kind of material is probably not pure and becomes less and less so as the subject becomes more familiar with it. The more easily the material can be grasped, the less pure the reading becomes, as a rule. Nonsense material is always rather hard to articulate and hard to work with generally, and as a consequence, there is less of a tendency to depart from reading when it is prescribed. But in the case of connected sense material, the reader is usually already familiar with the words and phrases as such; only the combinations are new and doubtless not all of them. The greater fluency and greater familiarity of the material results in combining recitation with reading; only certain key words need be noticed, the gaps being filled in by recall. The learner can glance along the lines, scarcely seeing more than an occasional word which suggests the context.

The physiological explanation that was applied to reading and recitation in general can be equally well utilized to illustrate these points. In recitation the connections between the items (represented in Figures 7 and 8 by P–P^1, P–B^1, B–B^1, etc.) are, in considerable degree, already given in the material. In fact, the serial associations between the words of familiar phrases are already fixed in one's nervous system through earlier practice. Recitation, as a factor making possible the formation of many connections, is consequently not needed. In other cases where the connections are less definitely formed, only a small amount of practice is required to stamp them in. The result is that in so far as the connections are ready-formed, reading amounts in all essentials to recitation. The eye neglects many of the words as such, fixating only occasional points. Reading thus becomes far from pure and approaches recitation, in all likelihood, more and more closely as the learning advances.

VI

CONCLUSIONS AND PEDAGOGICAL IMPLICATIONS

A detailed summary of results will not be attempted at this point. Only a few of the results which are of practical importance for the work of the school-room will be repeated. The reader who wishes a more detailed account of the findings may refer to the summaries that are to be found at the close of the previous chapters.

The compilation of quantitative and introspective evidence has shown that reading and recitation are relatively distinct functions in the process of learning. Each has its proper office to perform, and to restrict the learning entirely to one or the other results in loss of time and energy. Reading, as the introductory function, should be employed until the learner is fairly well acquainted with the material as a whole; until a method for further attack has been tentatively adopted; and until the first attempt at recall will meet with some success without too great a distortion of the material. The optimum point for introducing it, thus, occurs early in the process, but to introduce it too early, as well as to introduce it too late, will have a detrimental effect. In determining the exact moment at which recitation can best be introduced, one must take into account the length of the lesson, the difficulty of the material, the kind of learning that is desired, the age, training, and general capacity of the learner.

The function of recitation, as we have seen, is similar to that of practice in any form of sensori-motor learning. Memorization consists in selecting certain essential bonds, eliminating the unfit, and exercising the former until the connections are so well formed that once initiated, the series of responses will occur in proper sequence. The laws of use and disuse apply here as in other forms of learning; the physiological basis is the same.

Since recitation is equivalent to practice in other forms of learning, we should expect as a matter of course that any restriction upon its employment during the process of memorization should result in retarding improvement. Our experiments upon this point have shown that this is the case. This was true for all subjects, except children so young as to be unable to meet the requirements of the test, and for all materials employed, although, as might reasonably be expected, minor differences are to be found. In general, a method in which recitation is introduced at the optimum time, in comparison with a method in which the learner is entirely restricted to reading, enables the learner to reproduce immediately after a short period of study approximately twice as much material. The advantage of recitation as one should expect, is much more pronounced in delayed recall. After an interval of three or four hours, recitation makes possible the recall of four times as much material as does reading. This is to be expected, since recitation is understood as a process of adequate practice, while reading, whose function is introductory, restricts or inhibits the exercise of the bonds upon whose strength recall depends. In reading, while many of the bonds may be well enough established for immediate use, the neural connections rapidly disappear with disuse.

As the nature of reading and recitation now appears, the question is not so much—How is it that reading produces such poor results?—but rather—How is it that reading permits of any memorization at all? The evidence that has been gathered makes it doubtful whether *pure* reading would result in memorization. But there is little doubt that *pure* reading is a fiction; more or less recitation is always present in any prolonged effort to learn.

The fact that reading is seldom if ever pure can be most clearly illustrated in the case of learning sense material, and this fact helps us at the same time to understand why reading as a method of learning is more fruitful when applied to such material than when employed with non-connected senseless material. Nearly all of the subjects admitted that their learning, especially of sense material, was not limited to pure reading. The eye moved along the line actually seeing only occasional words. Other words, in fact whole phrases, were filled in by recall. The text served only to suggest groups of words or ideas which were for the most part filled in by the learner. In so far as this subjective reproduction of the material was carried on, to just that extent the learner was reciting rather than reading, and without doubt this sort of recall was at all times considerable, becoming more and more so as the learning progressed. Consequently, it appears that the memorization of the material, technically speaking, must, after all, be attributed to recitation.

The findings of Chapter IV were to the effect that reading was much more productive when the material was senseful and connected than when senseless and non-connected. The previous paragraph explains in part why this should be so. Reading of senseful connected material is far from pure, while with senseless material, on account of its less fluency and lack of senseful serial associations, the learner finds it less unnatural to actually see and read each item. No associations are present in the material which enable the learner to fill in the gaps when only occasional syllables are read. The bonds between items must be built up by the learner himself, and it is in this process that recitation is of the greatest value. These two factors together, namely, that the bonds between items in nonsense material must be worked in by the learner and that reading is much more pure with this material, explain the relatively greater advantage which recitation brings about with nonsense as compared to senseful material.

In addition to the fact that recitation as compared with reading enables the learner to form the requisite bonds more quickly and more permanently, the results of Chapter V have indicated other advantages of recitation as a form of learning. It was found that recitation leads to greater certainty of one's knowledge. It enables the learner not only to know but to be aware of how well he knows. Fewer blunders and erroneous recalls are made. The material is better organized; it is in more usable form. The meaning of the material is better obtained, and the relations among parts become more clear. In addition to this, as Katzaroff found (see page 5), material learned by means of recitation can be more promptly recalled; the recitation time is less.

From every point of view the superiority of recitation over reading, beyond the few perusals required to furnish the initial grasp of the material, is very clear. It holds for all materials and for practically all subjects. Consequently, the applications of the results to pedagogy are direct and manifestly important.

For the improvement of methods of study among school children, it is first of all necessary that the teacher should be aware of the value of recall in learning and that she should endeavor to impart this information in a practicable way to the pupils. That the pupils cannot be depended upon to discover economical methods of studying by themselves has often been discovered by inquiry. Miss M. J. Baldwin,[1] for example, found for grammar and high school pupils "that eighty-two per cent. studied words rather than thoughts, that they study in a mechanical sort of way which enables them to say that they have studied the lesson and spent the required time. They read the words over and over and doubtless get more confused the more they read."

It is perhaps not sufficient, however, that the pupils should be merely aware of the fact that attempted recitation is an essential process in learning. The teacher must devise means by which the pupils may be induced to study by trying to recall the material rather than by merely continuing slavishly to read and reread the words. The determination of these means, of course, does not lie

[1] 'How Children Study', *Archives of Psychology*, 1909, No. 12, p. 70.

within the scope of this study. It has been the purpose here merely to show that the recitation method *can* be employed, and employed very effectively, by pupils from the third grade up.

A few things which may induce the pupil to rely more upon recall will occur to anyone. Any method which requires the summarization of the facts of the lesson brings recall into play, since in such a process the pupil must think over the whole material, cull out the essentials, and state in his own words the main points. The teacher should encourage the pupil to react to the lesson in this way and reward him for successful attempts. Condensations of the ideas in written form, or even better, if possible, summarizing the content of the lesson mentally, is almost certain to bring rich returns. It brings into play the beneficial factors involved in recitation, develops power to distinguish the essentials from the unessentials, and may develop confidence and satisfaction in the pupil, since it enables him to be more certain of his mastery of the material. The pupil who has reasonable assurance that he has the lesson in hand can approach the recitation before the teacher in a more effective frame of mind.

The fact that recall is of such great importance in learning has a significant bearing on the nature of the recitation period in the school-room. As Colvin has pointed out:[2] "The fact that the recitation, as such, is largely ignored in higher grades of instruction is doubtless a serious pedagogical defect, which can be remedied only by accustoming the student to practise on his own initiative recall in his learning." The recitation should be regarded not merely as an opportunity afforded the teacher to find out what the pupils know, how hard they have studied, and what grade they should be given. Instead of an inquisition it should become a period of instruction. It should offer the pupil an opportunity to recite material he has previously more or less completely mastered. If the silent pupils could be induced to recall the material as well as the pupil who is orally reciting to the teacher, the period could become a valuable opportunity for review. Its most admirable function would consist in affording the pupils an opportunity to discover where their knowledge is hazy, inexact, and uncertain.

In addition, the teacher should make of the recitation a means of discovering the methods of studying employed by the pupils and of suggesting improvements in that respect. The unprepared student should not simply be met by the remarks: "How many times did you read your lesson?" and "Go read it again!" but more detailed inquiry into the cause of failure, followed by more valuable suggestions with regard to methods of study, should be the pro-

[2] *The Learning Process*, New York, 1913, p. 165.

cedure. In a word: "It should be remembered that instruction in the technique of learning is perhaps as important as instruction in the content of the subjects of the school curriculum."[3]

More advanced students may profit by the knowledge of the indispensible value of recitation. The college student is confronted by a situation in which the 'absorbing' of knowledge seems paramount, and where reaction is too little required. Listening to lectures and reading the texts require most of his time; recitations are few and far between. That they 'read lots but learn little' is a stock criticism, and it is indeed not seldom true that the college student is quite as ignorant of economical methods of study as the grammar school pupil. Recently the writer heard the case of a college student who came to a professor of psychology for an examination of what he believed to be a very poor memory. The student asserted that he could read a lesson over a dozen times and still not know it. A brief examination showed his memory not to be below par, but all the evidence indicated entirely inadequate methods of study. The student relied upon impression with little or no effort at expression; recall of the main points of his lesson was seldom tried. Yet for the college student who is so seldom called to account for his acquirements, recitation is more than usually essential. Frequent reviews, thinking the matter over by one's self, writing briefs of the main points, conversation with other students, and the like, are valuable because they throw into relief the portions that are hazy, inexact, and confused as well as because they fix more clearly in mind the material that is rehearsed.

Various opinions have been expressed with regard to methods of taking notes during lectures.[4] Doubtless the method must be varied somewhat to suit the material that is presented, but the findings in the present study suggest a method which, although seldom employed, should bring good results. Instead of making of one's self a mechanism for transferring spoken words to paper with but little heed to the meaning, the student devotes his attention to a thorough understanding the material presented, selecting the important points, organizing them into a systematic whole as the lecture progresses, and for the most part, delaying to a later hour the writing of the notes. Later in the day or evening, the lecture is rehearsed and an outline written down for future reference. While some disadvantages, or more likely, inconveniences, of such a method may appear, certain advantages of an important nature are obvious. First of all, the student may develop better habits

[3] Colvin, *op. cit.*, p. 178.

[4] This subject will be found discussed at length in two recent books: G. V. N. Dearborn, *How to Learn Easily*, Boston, 1916, Chapter II, and Harry D. Kitson, *How to Use Your Mind*, Philadelphia, 1916, Chapter II.

of attention during the lecture. He forces himself to pick out the essentials, to grasp the relations of ideas and to unify and organize the material presented. The will to remember, which Meumann so strongly emphasizes, comes into play. The student must actively grasp the meaning of the lecture in order to be able to reproduce it later. Secondly, the writing of a brief of the lecture at a later hour combines the advantage of a recitation, which the copious note-taker too seldom practises, with the well known benefits to be derived from the distribution of learning periods.[5] A few students who have tried this method speak enthusiastically of its effectiveness.

Finally, a word with regard to a more technical application of the results of this study. Individuals, when permitted to study by their 'natural method,' were found to employ various methods, not only for different materials, but for lessons of the same material and of the same length, at different times.[6] The quantitative results consequently vary considerably, according to whether the subject does or does not happen to employ an optimum combination of reading and recitation. In experimental work on memory and learning in which successive tests under constant conditions are required, it would seem to be an important precaution to specify the time at which the learner should change from reading to attempted recall, with instructions to employ thereafter the recitation method until learning is complete.

[5] See Jost, A., 'Die Assoziationfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen', *Zeitschrift für Psychologie*, 1897, **14**, pp. 436–472, or Ebbinghaus, H., *Memory*, translated by H. Ruger and C. Bussenius, New York, 1913.

[6] But little of the actual data bearing on this point has been presented in this paper. For the most part, such data were obtained from the practice tests conducted preliminary to those here presented.

MENTAL FATIGUE

MENTAL FATIGUE

DURING CONTINUOUS EXERCISE OF A SINGLE FUNCTION

BY

THOMAS RUSSELL GARTH, Ph.D.

WEST TEXAS STATE NORMAL COLLEGE

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. Woodworth

No. 41, AUGUST, 1918

COLUMBIA UNIVERSITY CONTRIBUTIONS TO PHILOSOPHY AND PSYCHOLOGY

Vol. XXVI, No. 2

NEW YORK

THE SCIENCE PRESS

Agents: G. E. Stechert & Co.: London (2 Star Yard, Carey Street, W. C.); Leipzig (Koenigstr. 37); Paris (16, rue de Condé)

THE writer makes the following acknowledgments: To Dr. R. S. Woodworth for valuable suggestions in the handling of the data and criticism of the calculations; to Dr. E. L. Thorndike for use of his addition sheets and suggestions as to the conduct of the experiment; to his wife, Ethel Nadine Garth, who corrected some of the sheets and assisted in the tabulation and without whose encouragement the enormous task of correcting the 12,355 papers and that of their tabulation and the many and various calculations might never have been brought to completion. My thanks are due also to Dr. J. A. C. Chandler, Superintendent of City Schools of Richmond, Va., and his first assistant, Mr. A. H. Hill, for access to the classes of the schools of that city; to Mr. F. M. Martin for the same courtesy in the Petersburg, Va., schools, and to the principals and teachers who rendered valuable assistance in giving the tests; and to Mr. S. P. Duke, Head of the Department of Education of the State Normal School of Farmville, Va., for much appreciated assistance while tests were being given in Richmond.

INTRODUCTION

A SURVEY OF THE SUBJECT OF MENTAL FATIGUE DURING CONTINUOUS EXERCISE OF A SINGLE FUNCTION

So far the investigations of various experimenters in the field of Mental Fatigue during continuous exercise show that continuous work under the lash of keen motivation reduces the efficiency only slightly, even though the subjects worked from one to two hours. Burgerstein, *'91*, found in an experiment with a large group of boys, aged eleven to thirteen years, in which the pupils performed easy examples in addition and multiplication in ten-minute periods with five-minute rests between, that the number of examples increased as the test proceeded, but there were more errors and corrections made by the pupils as the work went on.

Höpner, *'94*, undertaking to make the exercise more like that of the ordinary recitation, had a class of forty-six boys, of average age, nine years, write nineteen sentences from dictation. Each sentence was read aloud to the class once, then it was repeated once by a single pupil, next by the entire class, and then the pupils were required to write it from memory. The experiment extended over two hours. Höpner found that as the time proceeded, the pupils became less exact. Colloquial speech often took the place of the words in the dictated sentence. It is very evident that the materials for the experiment were not of equal difficulty. Neither were the pupils pressed to their utmost. (Offner & Whipple, *Mental Fatigue*, 57ff.)

Holmes, *'95*, had children from Grades Three to Eight add numbers for thirty-six minutes and broke the time into four periods of nine minutes each. There was an increase in errors of adding and copying. But when everything in the way of length of problem and speed is taken into consideration, the decrease in efficiency is not very great.

Cattell and Dolley found that simple reaction-time is not influenced by previous reactions to any great extent.

Thorndike describes an experiment of continuous mental work "of from three to eight hours" which "failed in the case of Dr. R. S. Woodworth, to produce any demonstrable diminution in efficiency." One part of the experiment was to mark the *e's* and *t's* in a book of 151 pages. This process continued for eight hours with less than nine minutes interruption. (See Thorndike's *Educational Psychology*, Vol. III, pp. 29–32.) Thorndike, '*11*, had fourteen students work from four to fourteen hours doing mental multiplication of three-place numbers by a like number. Six of these individuals took fifty-four per cent. more time for the end of the period on one day than they did at the beginning of the period for the preceding day. Six others had an average increase of fifty-nine per cent. For the other two, the increase was twenty-one per cent.

Arai, '*12*, had eleven students perform multiplication of two-place numbers mentally for two hours on two occasions. The intermissions were not the same for all participants in the experiment. The measure of efficiency was the time required for each example. It was found that the time was reduced. In performing the multiplications this means that there was an increase in efficiency.

Kraepelin worked out a method for continuous adding. A notebook is prepared with numbers printed in vertical columns. A bell is made to ring every five minutes and when the subject hears the bell he makes a mark. In this way the record is made according to the time worked. Kraepelin and his followers (Offner & Whipple, pp. 62–73) claim to have found certain interesting features in the work curve. These are practice, warming-up (Anlauf), swing (Anregung), spurt (Antrieb), initial spurt if it occurs at the beginning and 'spurt of change' if it occurs when a new kind of work begins, and 'terminal' spurt or end spurt (Schlussantrieb) when the subject realizes that he is nearing the end of the experiment or exercise.

Thorndike's absolutely fair treatment of the results and data of Kraepelin's work and of those who follow him brings into question the existence of such features in all work curves, and calls for further proof of their existence. He says: "Two hours or less of continuous exercise of a function at maximum efficiency produce a temporary negative effect . . . . of not over ten per cent. and in most functions still less than that."[1]

[1] *Educational Psychology*, Vol. III, pp. 68–69.

F. M. Phillips, *Journal of Educational Psychology*, 1916, had pupils of school grades of Fourth to Eighth work for ten minutes on arithmetic problems of addition, subtraction, multiplication and division. At the end of each minute the subject made a mark on the paper so that the experimenter knew what was done each minute by the subjects. In all the processes the best work was accomplished during the first minute. But the experimenter did not undertake to show whether there was fatigue or not. He found correlations between the first and last periods, in all the processes, ranging from .54 to .73, and says that the work of the first minute is less variable than that of later minutes, and that it is a fair index in all probability of one's ability in arithmetic.

W. S. Painter, *Journal of Educational Psychology*, 1916, did mental multiplication of four-place numbers after much mental and physical work which had resulted in a feeling of fatigue. He worked from 11 p. m. until 3:07 a. m. During the latter part of the experiment there was "a marked rise in time" per problem and the experimenter thinks that "the outcome suggests strongly that there exists a definite and relatively abruptly appearing point beyond which mental work becomes impossible." While the time in performing the last three multiplications is high, the errors are relatively few. Mr. Painter believes, however, that after this abrupt point all mental work, properly so called, was practically impossible.

CHAPTER I

THE ADMINISTRATION OF THE EXPERIMENT

The Subjects of the Experiment

There were seven hundred and eleven subjects in these experiments. Three hundred and sixty-eight of them were children in the Third and Fourth grades, and three hundred and forty-three of them were children of the Seventh and Eighth grades of the public schools of Virginia, all white children. A large part of the number was from the city schools of Richmond and Petersburg. The smaller part was from the Training School of the State Normal School at Farmville, Va., and the public schools of Farmville, Va. The boys and girls were about equally divided and their ages were those of the average for the grades in which the tests were given. There was absolutely no selecting of subjects, but the experimenter took the cases where he could get them. For this reason it may be taken as a fact that the results represent the work of the average child of these grades.

Materials Used

The materials used were the Thorndike addition sheets, such as those used by Kirby in the experiment described in his "Practice in the Case of School Children," Teachers College, Columbia University Contributions to Education, No. 53. There are seven different sheets. Each has forty-eight columns of one-place numbers, each column contains ten addends with the *1's* and the *0's* omitted, and each column is so arranged that any successive five of the columns are of a difficulty nearly, if not exactly, equal. These sheets were arranged in pads. For the Third and Fourth grades there were fourteen sheets, and for the Seventh and Eighth grade group there were twenty-one sheets in a pad. There was no possibility of a child's remembering the answers on a sheet which recurred, because there would be no way of identifying that sheet. Eye strain was reduced to a minimum because the type is so very large, and besides, the eyes got a rest every two minutes.

Conditions of the Experiment

Since the purpose of the experiment was to ascertain the behavior of the child mind during a recitation, both the materials and the conditions of the experiment were made as nearly like a recitation—or a drill recitation—as was possible under the circumstances. The materials were ordinary addition problems of one column, such as a teacher might use for a drill on adding and for improvement in number combinations. The experimenter gave all the tests in the presence of the regular teacher, and relied on his own teaching experience to put the child perfectly at his ease. There was nothing strained or strange about the experimenter's attitude but there was a constant endeavor to make the manner and atmosphere of the experiment entirely like that of an ordinary recitation-drill.

The subjects were told that the experimenter wished to find out who was the best adder in the class and that papers would be passed for this purpose, if the class was agreeable to it. The students were usually eager to try. They were told that the papers would be examined and that the name of the best adder would be announced later on by the teacher. The pads were then passed to the pupils, face downward. The experimenter used a stop watch. At a signal the pupils turned the pads face upward and added for two minutes when time was called. As quickly as possible with no intentional intermission except such as was necessary to keep the class together, the class received the instruction to "take the next sheet" and the signal "go" was given, when they again worked two minutes by the stop watch. The intermissions were only about ten seconds long. But in this time the subjects raised their eyes to the experimenter who tried to maintain a 'sympathetic' countenance such as would keep them working happily. Nothing more was said to the class by the experimenter after the initial motivation, excepting "take the next paper—get ready—go!" and after they had worked two minutes, "stop." The Third and Fourth grade pupils worked thus for twenty-eight minutes, and the Seventh and Eighth grade pupils for forty-two minutes.

The pads supplied a record of what was done every two minutes by each pupil. After the total time had passed the pads were collected. Each pad bore the name, age, and school and grade of the pupil on the back written by the pupil when he received the pad.

Method of Scoring

With fourteen sheets to every one of the 368 Third and Fourth grade pupils and 21 sheets to every 343 Seventh and Eighth grade pupils, the experimenter secured 12,355 papers. For the Third and Fourth grades there were fourteen periods of two minutes each and for the older group there were twenty-one periods of two minutes each. When these papers were examined, a record was made of what each child attempted each two minutes, and another record of the accuracy of each child for every two-minute period. If a child attempted six columns, he received a credit for that two minutes of six columns attempted. If only four of these were correct, he received credit on the accurate record sheet of only four columns. These two records, one of attempts and one of accurates, gave two tables, in the one case having fourteen columns 368 figures long, and in the other having twenty-one columns 343 figures long for attempts, and likewise the same for accurates. In short, we should say that the two groups of children—Third and Fourth grades and Seventh and Eighth grades—had two arrays of results, one for attempts and one for accurates.

If a child's record showed itself incomplete—nothing attempted after a certain time—except in the last period or so, his work was not counted in the experiment. One child had to stop because of a slight illness which she had when she came to school that morning. Another boy was convalescent from typhoid fever and had to stop in the midst of the experiment. If a lack-a-daisical attitude toward the experiment was evidenced by any great omission of effort in the written record, the material was thrown out. But it so happened that there was very little that had to be discarded. A child was given credit for what he did, both in quantity and quality.

Time of Day and Year

According to Heck's[1] experiments, the time of day makes little difference and so the experimenter made his tests at any time during the school day—early morning period, noon, and afternoon indiscriminately. Since there was a generous sampling of all times of the day, no one time of the day with its fatigue could assert itself

[1] W. H. Heck, *A Study of Mental Fatigue*, 1913.

to any great extent. The test was intended to show what happens during a recitation at any time of the day.

The time of the year was in the fall, in the winter, and in the spring. One class took its test when the snow was falling on the ground, a large part of the groups took their tests in November, and some of the classes worked with the windows open to the spring breeze. The tests were given during the time elapsing between December, 1913, and November, 1914.

CHAPTER II

ATTEMPTED AND ACCURATE PERFORMANCE

The Group Curves

Four curves are given herewith, plotted from the tables which accompany them. There are two work or attempts curves, one for Third and Fourth grades and one for Seventh and Eighth grades; and there are two accurates curves, respectively, for Third and Fourth grades and for Seventh and Eighth grades. In the Third and Fourth grade group there were 368 children participating and

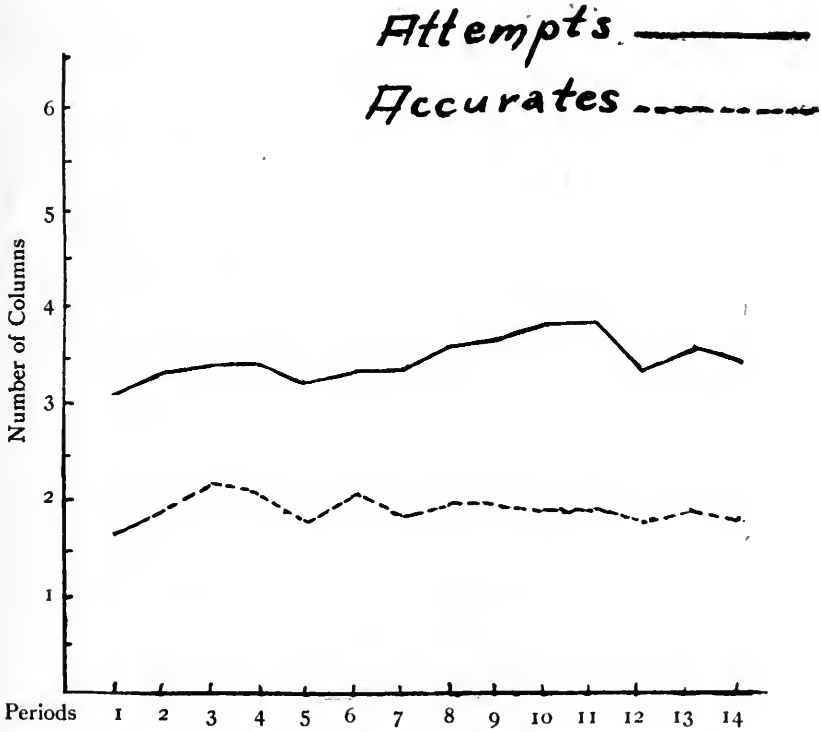


Fig. 1. Third and Fourth Grades

in the Seventh and Eighth grade group there were 343 who took part. The Third and Fourth grade group worked for twenty-eight minutes actually. This time was divided into fourteen periods of two minutes each. The horizontal line represents the time, in periods, during which the groups worked. The vertical line indicates the number of columns either attempted or accurately done. In the case of the work curve this vertical line indicates columns attempted whether right or wrong; in the case of the accuracy curve

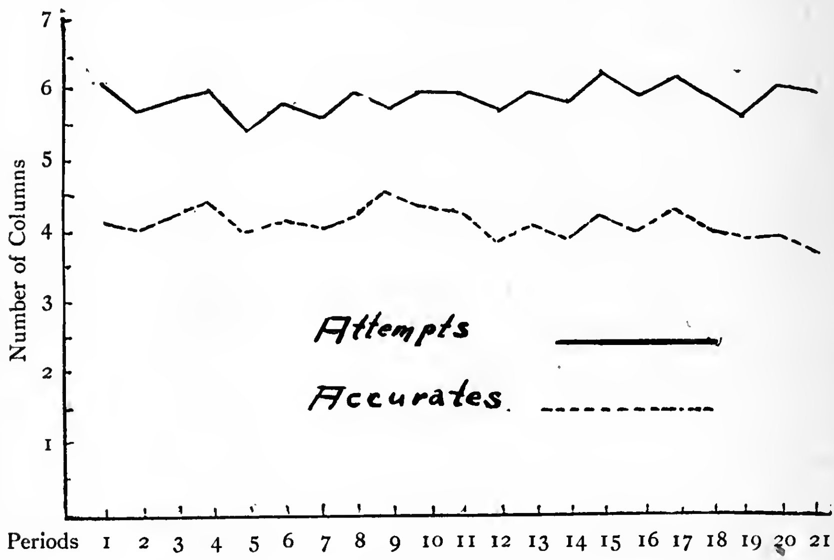


Fig. 2. Seventh and Eighth Grades

the vertical scale stands for columns correctly added. Every point on the curve was ascertained by finding the average attained by the group for that identical two-minute period of work. We shall call these the group curves.

These curves are given here for the sake of presenting the facts derived by the experiment. These are mere gross results. The attempts and accuracy curves derived herewith do not afford a true idea of measurement of fatigue of the individuals during the work period of twenty-eight minutes for the Third and Fourth grade group nor for the forty-two minutes during which the Seventh and Eighth grade group worked. If instead of giving the average we gave the quantity obtained by the group, this would emphasize the production by the group as a whole. These first curves give the changes in the group without regard to the changes of the individual. The curves following these will give the average of the individual curves and thus will regard individual changes.

TABLE I

Attempts of Third and Fourth Grades—Absolute Measures

Columns attempted for each period of two minutes

Period	*Average*	*A. D.*	*P. E.*
1	3.15	1.20	.058
2	3.35	1.17	.051
3	3.47	1.22	.053
4	3.48	1.34	.059
5	3.27	1.17	.051
6	3.37	1.25	.055
7	3.45	1.34	.059
8	3.54	1.33	.058
9	3.55	1.38	.061
10	3.57	1.39	.061
11	3.49	1.31	.057
12	3.28	1.31	.057
13	3.51	1.41	.062
14	3.38	1.36	.060

TABLE II

Accurates of Third and Fourth Grades—Absolute Measures

Columns accurate for each period of two minutes

Period	*Average*	*A. D.*	*P. E.*
1	1.70	1.27	.056
2	1.93	1.1	.048
3	2.11	1.32	.058
4	2.06	1.28	.056
5	1.88	1.16	.051
6	2.07	1.24	.054
7	1.86	1.21	.053
8	1.99	1.16	.051
9	1.98	1.34	.059
10	1.96	1.30	.057
11	1.94	1.27	.056
12	1.79	1.30	.057
13	1.82	1.21	.053
14	1.78	1.30	.057

TABLE III

Attempts of Seventh and Eighth Grades—Absolute Measures

Columns attempted for each period of two minutes

Period	*Average*	*A. D.*	*P. E.*
1	6.02	1.67	.075
2	5.62	1.39	.062
3	5.77	1.5	.068
4	5.92	1.61	.072
5	5.46	1.51	.068
6	5.78	1.55	.07
7	5.64	1.53	.069
8	5.81	1.69	.076
9	5.68	1.77	.079
10	5.98	1.52	.068
11	5.94	1.56	.07
12	5.65	1.69	.076
13	5.99	1.5	.068
14	5.83	1.68	.076
15	6.23	1.95	.088
16	5.91	1.52	.068
17	6.11	1.67	.075
18	5.95	1.84	.083
19	5.61	1.7	.077
20	6.02	1.78	.08
21	5.99	1.91	.086

TABLE IV

Accurates of Seventh and Eighth Grades—Absolute Measures

Columns accurate for each period of two minutes

Period	*Average*	*A. D.*	*P. E.*
1	4.22	1.85	.083
2	4.04	1.63	.074
3	4.19	1.50	.068
4	4.30	1.83	.082
5	4.01	1.57	.061
6	4.18	1.94	.087
7	4.03	1.59	.072
8	4.22	1.64	.074
9	4.64	1.65	.073

TABLE IV (*Continued*)

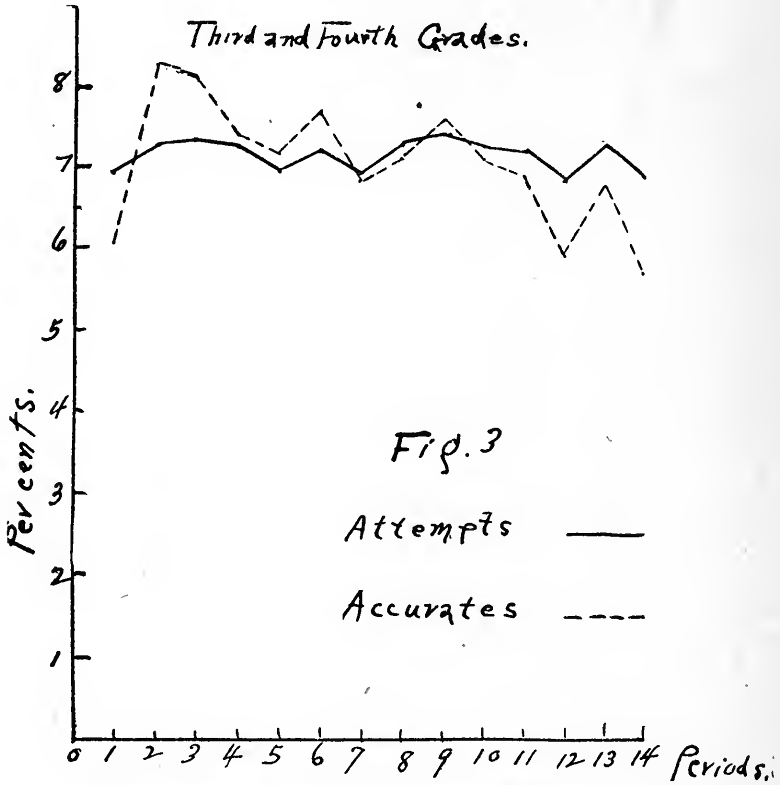
Accurates of Seventh and Eighth Grades—Absolute Measures

Columns accurate for each period of two minutes

Period	*Average*	*A. D.*	*P. E.*
10	4.31	1.68	.075
11	4.29	1.71	.076
12	3.89	1.76	.079
13	4.19	1.74	.078
14	3.88	1.76	.079
15	4.24	1.91	.085
16	4.03	1.71	.077
17	4.26	1.72	.077
18	4.02	1.72	.077
19	3.88	1.77	.080
20	3.93	1.83	.082
21	3.74	1.82	.081

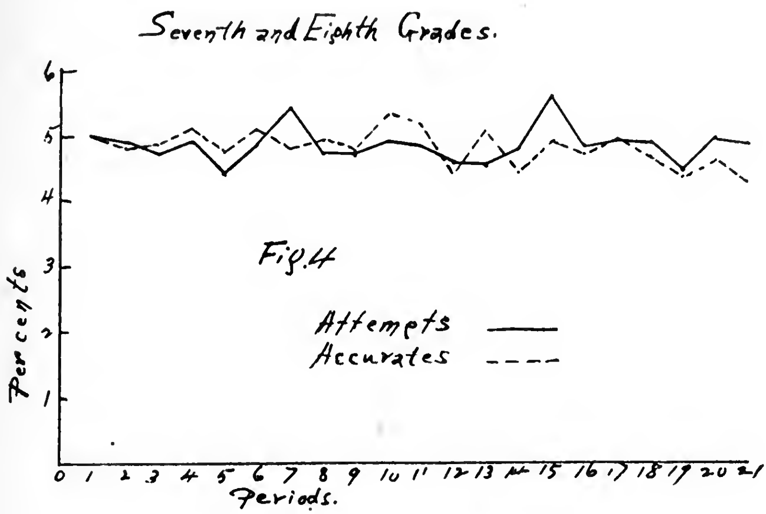
Average of the Individual Curves

The object in a study of this kind is to give every individual, whether his work is slow or fast, the same influence upon the final average. It is not fair to the poor or weak individual to put his small effort on a basis of absolute terms along with the effort of the better workers similarly expressed. The former may fatigue while the latter gain and a mean between them show essentially the curve of the better worker. The original data of the preceding tables fail to represent truly the facts which we wish to bring out because, as intimated above, the rapid workers with their larger measures of performance entirely swamp the changes that may occur in the work of the slower ones. But if everyone's work for the successive periods is reduced to percentages of his total, the slow count the same as do the fast workers in determining the general work curve. For this reason we have derived percentage curves which represent more truly than the foregoing curves the central tendency of the whole group with regard to fatigue. With the absolute numbers we can examine the question of fatigue from the standpoint of one interested in the total product of a group, whereas by the personal curves we disregard the total products and examine changes in the working ability of the individuals, and the general tendency of these individual changes.



Representing the Average of the Individual Curves.

The curves shown herewith (Figures 3 and 4) are, however, only slightly different in form from the preceding curves. Every number in the distribution is a percentage of the total accomplished during the entire period of work for each installment of time for each child. A child did so much every two minutes for twenty-eight and forty-two minutes in the Third and Fourth grade group and the Seventh and Eighth grade group respectively. Percentages of these distributions were determined by dividing each two minutes' work by the total amount added during the whole period of work by each individual. This gave fourteen per cents. in the younger group for each child, and twenty-one per cents. in the Seventh and Eighth grade group for each child of that group. The fourteen average per cents. for the Third and Fourth grade group were obtained,



Representing the Average of the Individual Curves.

and the twenty-one average per cents. for the Seventh and Eighth grade group likewise were obtained, and these averages give the accompanying curves. Every position on the curves indicates the central tendency of the relative value in attempts and accuracy of the groups for any two minutes of the time of working; and this should be taken as meaning the comparative working ability at that period. The attempts curve does not signify a pure mental performance, while the accurates curve here expressed in percentages does represent a more nearly pure mental process.

TABLE V

Data for the Curves Representing the Average of the Individual Curves—Third and Fourth Grades

Columns attempted in each period of two minutes, expressed as a per cent. of the total number attempted

Period	*Range Per Cent.*	*Average Per Cent.*	*A. D.*	*P. E.*
1	0–50	6.97	1.94	.09
2	0–25	7.24	1.52	.08
3	0–17	7.30	1.37	.06
4	0–17	7.24	1.40	.06
5	0–25	6.97	1.32	.06
6	0–14	7.20	1.32	.06
7	0–18	6.97	1.35	.06
8	0–25	7.29	1.43	.06
9	0–17	7.34	1.39	.06
10	0–12	7.28	1.44	.06
11	0–14	7.27	1.34	.06
12	0–13	6.78	1.47	.07
13	0–13	7.25	1.31	.05
14	0–16	6.90	1.62	.07
		100.00		

TABLE VI

Data for the Curves Representing the Average of the Individual Curves—Third and Fourth Grades

Columns accurate for each period of two minutes, in per cent. of total columns accurate

Period	*Range Per Cent.*	*Average Per Cent.*	*A. D.*	*P. E.*
1	0–100	6.13	4.53	.20
2	0–100	8.34	4.87	.22
3	0–50	8.16	4.34	.21
4	0–34	7.53	4.03	.13
5	0–100	7.27	4.20	.19
6	0–50	7.76	4.10	.18
7	0–100	6.94	4.01	.18
8	0–28	7.24	4.05	.18
9	0–100	7.62	4.57	.21
10	0–50	7.16	4.31	.19
11	0–29	7.06	4.27	.19
12	0–30	6.10	4.07	.18
13	0–34	6.89	4.30	.19
14	0–40	5.80	3.75	.17
		100.00		

TABLE VII

Data for the Curve Representing the Average of the Individual Curves—Seventh and Eighth Grades

Columns attempted for each period of two minutes, in per cent. of total columns attempted

Period	*Range Per Cent.*	*Average Per Cent.*	*A. D.*	*P. E.*
1	2–10	5.00	1.01	.05
2	0–10	4.86	.79	.04
3	0–9	4.67	.77	.04
4	2–10	4.79	.69	.03
5	2–8	4.37	.75	.03
6	3–8	4.73	.72	.03
7	2–8	5.30	1.04	.05
8	1–11	4.63	.79	.04
9	2–9	4.59	.75	.03
10	3–8	4.86	.75	.03
11	1–8	4.80	.68	.03
12	2–10	4.43	.74	.03
13	1–16	4.42	.88	.04
14	2–7	4.68	.83	.04
15	1–9	5.58	1.11	.05
16	1–9	4.74	.62	.03
17	0–9	4.80	.71	.03
18	0–9	4.75	.90	.04
19	0–8	4.40	.86	.04
20	0–9	4.84	.91	.04
21	0–11	4.76	.90	.04
		100.00		

TABLE VIII

Data for the Curve Representing the Average of the Individual Curves—Seventh and Eighth Grades

Columns accurate for each period of two minutes, in per cent. of total columns accurate

Period	*Range Per Cent.*	*Average Per Cent.*	*A. D.*	*P. E.*
1	0–20	5.00	1.74	.08
2	0–14	4.85	1.51	.07
3	0–30	4.85	1.51	.07
4	0–27	5.01	1.37	.06
5	0–11	4.50	1.37	.06
6	0–15	5.03	1.33	.06
7	0–20	4.70	1.46	.07
8	0–12	4.83	1.37	.06

TABLE VIII (*Continued*)

Period	*Range Per Cent.*	*Average Per Cent.*	*A. D.*	*P. E.*
9	0–12	4.73	1.30	.06
10	0–22	5.15	1.48	.07
11	0–17	5.10	1.28	.06
12	0–12	4.34	1.45	.07
13	0–12	5.03	1.39	.06
14	0–10	4.37	1.51	.07
15	0–13	4.90	1.48	.07
16	0–13	4.74	1.55	.07
17	0–11	4.84	1.42	.06
18	0–23	4.73	1.23	.06
19	0–13	4.40	1.35	.06
20	0–10	4.50	1.53	.07
21	0–12	4.30	1.69	.08
		100.00		

The Similarity between Attempts and Accurate Performance

The question arises as to whether the general tendencies of the attempts curve and the accurates curve are the same or different. Possibly the best evidence that they are rather similar is to be gotten by merely examining the curves themselves. (See Figures 1 to 4.) While the accuracy and attempts curves are not parallel they are decidedly similar in their general tendencies. To give further evidence of this similarity, correlations, though not as good evidence as the curves themselves, were gotten between the two kinds of curves.

The correlations between the Averages of Tables I and II, taken as a whole, is +.487. The correlation for the Averages of Tables III and IV, taken as a whole, is +.29. For Tables V and VI, and for VII and VIII, the correlations are respectively +.58 and +.32. The older group has the lower correlations. The curves derived from the per cents.—the curves called the average of the individual curves—give higher correlations than the curves derived from absolute measures.

Therefore, numbers for attempts and numbers for accurates show a positive tendency to vary together. They do not vary inversely, if we take the curves as a whole, though doubtless there are corresponding parts of the curves that do. Some of these divergencies will appear as we proceed in our examination of the curves.

TABLE IX

Absolute Measures

	Third and Fourth Grades		Seventh and Eighth Grades	
	Attempts	*Accuracy*	*Attempts*	*Accuracy*
I. Maximum period	Tenth	Third	Fifteenth	Ninth
II. Minimum period	First	First	Fifth	Second
III. Average first three periods	3.32 cols.	1.91 cols.	5.80 cols.	4.15 cols.
IV. Average last three periods	3.39 cols.	1.79 cols.	5.87 cols.	3.85 cols.
V. Per cent. loss or gain III and IV	2.0 gain	6.2 loss	1.0 gain	7.0 loss
VI. Average part one (or first third of curve)	3.36 cols.	1.95 cols.	5.74 cols.	4.14 cols.
VII. Average part two (or second third of curve)	3.44 cols.	1.96 cols.	5.84 cols.	4.20 cols.
VIII. Average part three (or third third of curve)	3.45 cols.	1.86 cols.	5.97 cols.	4.01 cols.
IX. VI as base	1.00	1.00	1.00	1.00
VII : VI	1.023	1.005	1.017	1.014
VIII : VI	1.026	.95	1.04	.97

TABLE X

From the Data of the Curves Representing the Average of the Individual Curves

	Third and Fourth Grades		Seventh and Eighth Grades	
	Attempts	*Accuracy*	*Attempts*	*Accuracy*
I. Maximum period	Ninth	Second	Fifteenth	Tenth
II. Minimum period	Twelfth	Fourteenth	Fifth	Twenty-first
III. Average first three periods	7.17%	7.54%	4.85%	4.90%
IV. Average last three periods	6.98%	6.26%	4.67%	4.40%
V. Per cent. loss or gain	3.0 loss	17.0 loss	4.0 loss	10.2 loss
VI. Average part one (or first third of curve)	7.19%	7.54%	4.82%	4.85%
VII. Average part two (or second third of curve)	7.15%	7.37%	4.63%	4.79%
VIII. Average part three (or third third of curve)	7.10%	6.60%	4.84%	4.63%
IX. VI as base	1.00	1.00	1.00	1.00
VII : VI	.99	.98	.96	.99
VIII : VI	.98	.88	1.004	.95

The accompanying tables, Nos. IX and X, give some of the data to be derived from the curves. Table IX is a summary of data gotten from the absolute measures and Table X is a similar summary gotten from the percentages. Using these data of the tables we shall discuss various factors which may influence the curves, as fatigue, warming-up, initial-spurt, practice, and end-spurt. Accordingly, the discussion will be from the two standpoints spoken of above, that of one interested in the total product and that of one interested in changes in working ability.

Maximum and Minimum Points on the Curve

The maximum point on a curve of continuous performance at least makes a good landmark for a preliminary survey of fatigue in that curve. Being the high point in the curve, the line must recede from it, be it ever so gradually, and this negative effect must be due to fatigue—granting that the workers are serious in their performance. The minimum must mean, if we still grant the seriousness of the workers, one of several things, *i.e.*, if it occurs before the maximum it must signify a lack of proper adaptation or a warming-up, if it follows the maximum it must be due to fatigue (from which there may be recovery). The maximum point in the curve does not tell us when fatigue begins to work in the performance; but it makes the presence of fatigue evident by the falling away that follows it. Consequently, it makes a good point at which to begin the discussion of fatigue.

If the decline in the curve following the maximum was not due to fatigue, it must have been due to ennui or loss of interest. It might better be said, shifting of interest. Now, because of the original strong motivation through appeal to the competitive impulse and because no other interest was permitted to arise in as far as lay within the power of the experimenter to prevent it—for the subjects really had no time for mind wandering, and were kept steadily at their tasks—for this reason the writer believes there was very little shifting of interest other than such as was due to the distraction accompanying fatigue.

For the group curves obtained from the absolute measures (See Table IX) the maximum of attempts was reached, in the case of the Third and Fourth grade group, in the tenth period or between the eighteenth and twentieth minutes. The maximum of accurates occurred in the third period or at about the sixth minute. As to the minima, they occur before the maxima and indicate lack of adaptation in consequence. The foregoing facts indicate that the maximum was reached earlier by fourteen minutes in the accurates curve than in the attempts curve. For the similar curves of the Seventh and Eighth grade group the same observation is made that the maximum of accurates preceded that of attempts by ten minutes. The older group showed more persistence, as the maximum of accurates occurred in the ninth period or about the eighteenth minute and that of attempts in the fifteenth period, or between the twenty-eighth and thirtieth minutes. Oehrn found that when his adult subjects added continuously for two hours, they attained their maximal speed about twenty-eight minutes from the start. (Whipple, *Manual of Mental and Physical Tests*, 335.)

From the curves representing the average of the individual curves the same thing is true as to the maximal attainment of accuracy occurring before that of work in attempts. The minima of accuracy since they are found following the maximal points in both groups register fatigue. (See Table X for the data from which these observations are made.)

The two standpoints—interest in total work and in changes in working ability—make practically no difference as to the time of occurrence of the maxima.

The general conclusions to be drawn here are then:

1. Fatigue is made evident in a curve of this sort by the line falling away from the maximum.

2. Workers attain their maximum of accuracy before they attain that point in attempting.

3. In both quantity and quality of work, the younger children reach their maximum as a group before the older ones, and therefore notably fatigue sooner.

4. The occurrence of the minimum after the maximum is in consequence of fatigue and not from lack of adaptation.

Fatigue

If we wish to find fatigue in an attempt or accurate performance curve we look for it certainly at the end of the performance because, though it may have been a factor nearly all along the way, it would here be more in evidence than elsewhere. Fatigue is made evident by a falling away from a previous performance. Now we must have something to measure by as well as something to measure. The basis of measurement or comparison taken here is the average of the first three periods of work. By taking our standard at the beginning we avoid practice and fatigue effects.

The question of novelty of stimulus is of slight importance here since the subjects were rather familiar with such material as that given in the experiment.

That part of the data of Tables IX and X given as III and IV, 'average of first three' and 'average of last three', forms the measures which are compared for the purpose of ascertaining the fatigue at the end of the total time of working. 'Average first three' signifies the average of the first three periods of two minutes each, and the average of the last three periods of two minutes each is signified by 'last three'. The facts under III give the basis of comparison. V, 'per cent. of loss or gain', indicates the results of the comparisons for the various curves. The results show that in absolute measures attempts gained slightly and accurate performance lost. The gain for attempts of Third and Fourth grade group was 2 per cent., for Seventh and Eighth grades, 1 per cent. The fatigue in accurates of Third and Fourth grade groups was 6.2 per cent. and for Seventh and Eighth grades was 7 per cent.

The result of similar comparisons in Average of Individual Curves made after the same fashion as that of Absolute Measures shows there was fatigue in both attempts and accurates in both the younger and the older groups. (See Table X.) Fatigue in attempts for the former was 3 per cent., for the latter was 4 per cent., but the accurates suffered a 17 per cent. loss for the younger group and 10.2 per cent. for the older.

If the writer's contention is accepted that the curves representing the Average of the Individual Curves best represent the facts, we may conclude that attempts and accurates alike suffer fatigue in an exercise of this sort.

General Form of the Curves

We first present Figures 1-a, 2-a, 3-a and 4-a, which are "smoothed" from Figures 1, 2, 3 and 4 respectively, by taking the averages of the first and second periods, of the third and fourth, etc.; *i. e.*, the two-minute periods of the experiment are here combined into units of four minutes each. Some of the chance irregularities of the original curves being thus removed, the general course of the curves can be better seen.

From Figures 1-a and 2-a it appears that the working power of the group as a whole remained very nearly constant during the experi-

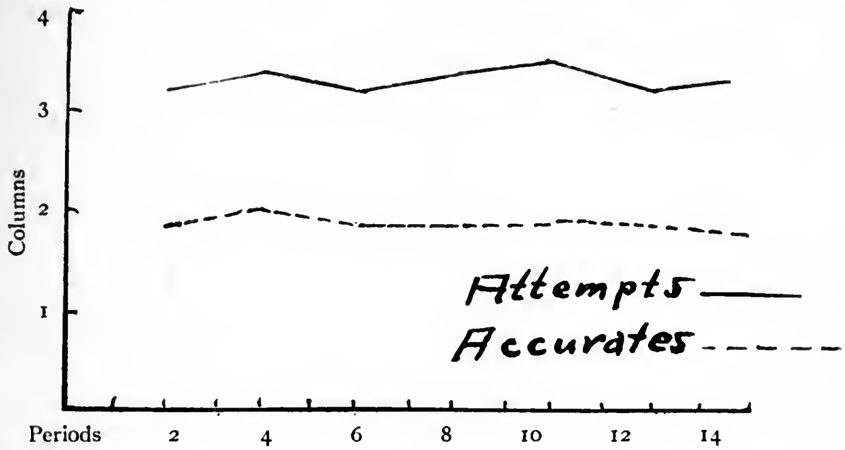


Fig. 1-a. Third and Fourth Grades

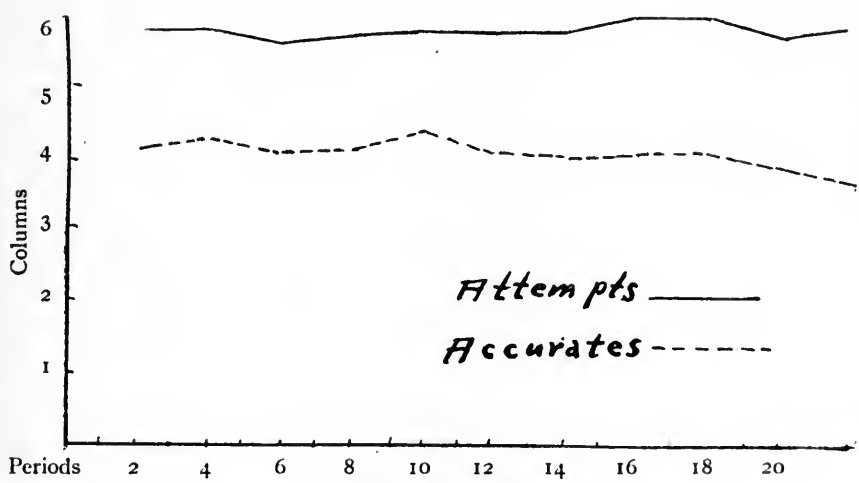


Fig. 2-a. Seventh and Eighth Grades

ment, though there is a slight but definite descent in the curves for accurate work, and this decline begins earlier in the younger than in the older group.

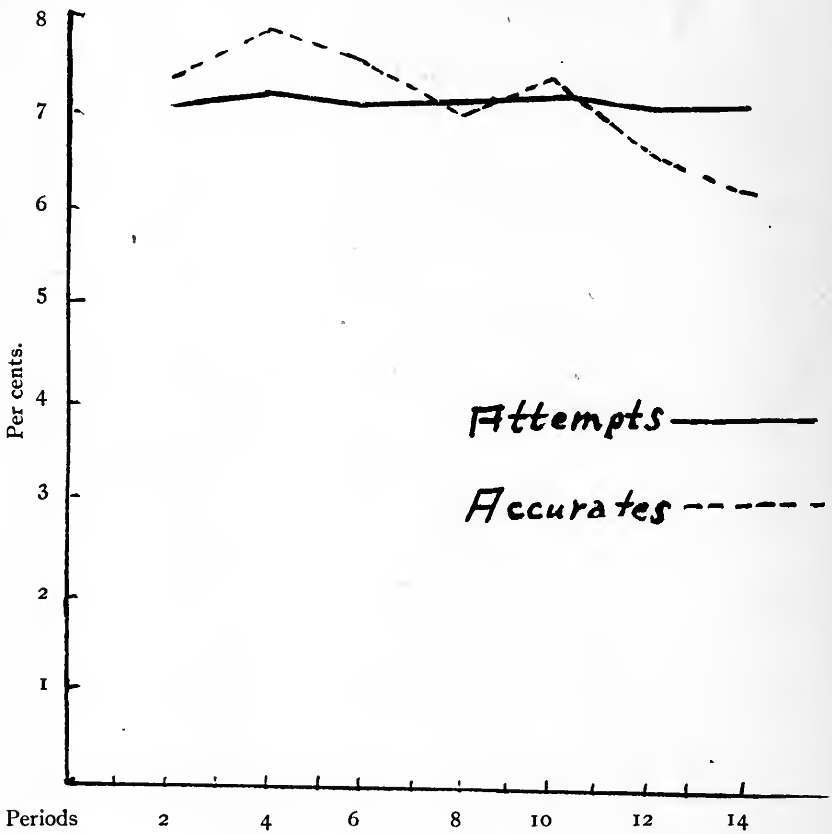


Fig. 3-a. Third and Fourth Grades

We next divide the curves into three parts, as nearly as possible thirds in the Third and Fourth group, and exact thirds in the Seventh and Eighth grade group. The former group is divided thus—four periods in the first part and five periods in the second and third parts. The fact that there were fourteen periods makes this unequal division necessary. The Seventh and Eighth grade group curve is divided into three parts, each containing seven periods. In the accompanying table, No. IX, for Group Curves, if the reader will refer to VI, VII, and VIII, he will find the averages and ratios to which we are going to refer. In the case of the Third and Fourth grade group the average amount attempted each two minutes for this part of the curve (eight minutes) was 3.36 columns, for the second part or ten minutes following the first division of time, the average was 3.44 columns and for the last ten minutes of

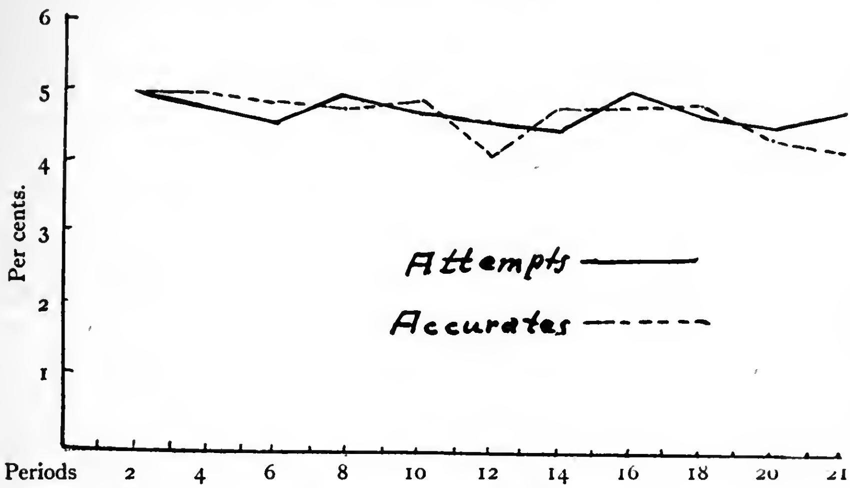


Fig. 4-a. Seventh and Eighth Grades

work 3.45 columns were attempted for two minutes on the average. This shows a steady rise, in the ratios of 1.00, 1.023, and 1.026.

In the case of the Seventh and Eighth grade group the three equal parts of the whole forty-two minutes show thus an average for the first fourteen minutes of 5.74 columns attempted, for the second fourteen minutes 5.84 columns were attempted and for the last fourteen minutes of time an average of 5.97 columns were attempted for two minutes. This likewise shows a rise. The ratios are 1.00, 1.017, and 1.04. These rises are so slight that too much significance should not be given them.

In the case of accurates, the first eight minutes in the Third and Fourth grade groups shows an average of 1.95 columns correctly solved for each two minutes. The next ten minutes of the total time produced an average of 1.96 columns correctly solved for each two minutes and the last ten minutes shows an average of 1.86 columns per period of two minutes each, correctly added. The ratios are 1.00, 1.005, and .95. With the older groups the three averages are 4.14, 4.20, and 4.01. The ratios are 1.00, 1.014, and .97.

The accurate performance curves here are seen to rise in the middle and decline at the end, but the attempts curves rise steadily in both groups. Practice was helping out very much in this, for them, quasi-automatic process.

Therefore, taking the group curves in three parts as we have just done shows (1) slight practice effect in the attempts process in both groups, (2) slight practice effect in the accuracy curve in the second third of the partition of the curve but fatigue in the last third in both groups, all of which holds alone for the absolute measures.

The above facts hold for the curves derived from absolute measures, but the corresponding parts of the curves showing average of the individual curves are somewhat different in their relations. The numbers in the curves are these (See Table X):

Third and Fourth Grades	*Attempts*
First Third, as base	1.00
Second third : first third	.99
Third third : first third	.98

Third and Fourth Grades	*Accuracy*
First third as base	1.00
Second third : first third	.98
Third third : first third	.88

Seventh and Eighth Grades	*Attempts*
First third as base	1.00
Second third : first third	.96
Third third : first third	1.004

Seventh and Eighth Grades	*Accuracy*
First third as base	1.00
Second third : first third	.99
Third third : first third	.95

The result is accordingly that the average individual falls off slightly in work attempted, with some recovery toward the end in the case of the older individuals (Seventh and Eighth grades); while in accurate work, the average individual shows a greater falling off, without recovery, and the falling off is greater in the younger child. The conclusion from this form of comparison is in general harmony with that derived from the immediately preceding comparison and with that derived from inspection of the curves.

'Warming-up' and 'Initial-Spurt'

We now proceed to examine these curves to see if there is 'Warming-up' or 'Initial-spurt'. "The best definition of 'Warming-up' as an objective act is that part of an increase of efficiency during the first twenty minutes (or some other assigned early portion) of a work period, which is abolished by a moderate rest, say of sixty minutes." Thorndike, *Educational Psychology*, vol. III, p. 66. "'Initial-spurt', if a real fact, will be found in an examination of the work, minute by minute, of the first quarter of an hour." *Ibid.*, p. 48. A sudden rise at the start will indicate 'Initial-spurt'.

The graphic representation of the attempted and accurate work affords perhaps the best objective material for determining the presence or absence of these phenomena in the performance under discussion, and so we call attention to the curves given in Figures 1 and 2, and derived from Tables I, II, III, and IV for Absolute Measures, and Figures 3 and 4 derived from Tables V, VI, VII, and VIII. The curves for the younger group show a fairly steady rise in the first six minutes in both attempts and accurate performances in all the curves representing them, and from this evidence we may say there was 'Warming-up' for them. If it should be objected that these rises were merely incidental and due to chance, we may resort to the tables showing the averages and their probable errors. By taking the difference between the averages of the successive periods, first and second, second and third, and bringing these into relation to the square root of the sum of their respective squared P. E.'s, we may determine the reliability of these differences. We give here these reliabilities. The chances that the gain of period two over period one, in Table I, is reliable are 91 out of 100; for period three over period two they are 75 out of 100. This 'Warming-up' holds there for attempts in the Third and Fourth grades. As to accurates the case is even better, for the chances that the gain of period two over period one, in Table II, is reliable are 96 out of 100, and of the gain of period three over period two are 92 out of 100, in the Third and Fourth grades. This holds for the Absolute Measures.

In the case of the curves representing the Average of the Individual Curves, we have the following facts as to 'Warming-up'. The chances that the gain of period two over period one is a true one are 87 out of 100. The chances that the gain of the third period over the second period (Table V) is significant are not so great, being only 32 out of 100 cases. In Table VI the gain of second period over the first, has a reliability of 84 cases out of 100. But there is no gain of the third period over the second, so that the 'Warming-up' disappeared by the time the sixth minute was reached.

Therefore, we may conclude, since the rise in the curves is found to be more or less reliable, that the 'Warming-up' found at the beginning of the work of the younger group was a real fact.

Now if there is here found in the younger group positive evidence of 'Warming-up', the fact excludes the possibility of 'Initial-spurt', which is in a measure the opposite of 'Warming-up'.

The curves for the older group do not show a 'Warming-up 'effect upon inspection. The rather sudden slight rise in the first period, followed by a fall in the second period, on the face of things indicates 'Initial-spurt' in both attempted and accurate performance. If we examine periods one and two (Table III) and take into account their probable errors we find that the loss of the second period over the first, has a reliability as a change of 98 out of 100 cases. This is for attempts. For accurates we take periods one and two, with their probable errors (Table IV), and find that the loss here is a difference whose reliability is 82 chances out of 100. These tables give the data for the Absolute Measures. The data for the curves representing the average of the Individual curves is found in Tables VII and VIII for attempts and accurates respectively. For attempts the chances that the change in going from the first period to the second is a real fact are 88 out of 100 cases and for accurate performances are 74 out of 100. The indications of 'Initial-spurt' are found thus to be fairly reliable.

Therefore, in so far as these data are concerned, the older children experienced an 'Initial-spurt' which is opposite to the 'Warming-up' experienced by the younger ones.

'End-Spurt'

Since the experiment was planned so as to avoid anything like 'End-spurt' and let fatigue do its worst at the end, it hardly seems worth while to examine these data with the intention of finding 'End-spurt', which, according to Kraepelin and others, is a sudden rise in efficiency, due to a knowledge on the part of the subjects of the approaching end of the task.

The experimenter here did all in his power to conceal the approaching end of the test. For this reason we do not think these data should be examined with the end in view of finding 'End-spurt'.

The writer has a conviction that interference operated at the beginning of the test because of the strong motivation. The desire to excel must have acted as a hindrance at this part of the game inasmuch as it probably acted as a distraction from letting the necessary associations play freely. Consequently, it would lower the score especially of accuracy. It also probably played strongly toward the end of the test and may account for some of the fatigue.

CHAPTER III

INDIVIDUAL DIFFERENCES

Individual Curves

In the preceding chapter, individual differences were eliminated and the general tendencies to fatigue, etc., of large groups alone considered. If we take the records of individuals, we shall, of course, find every variety of curve. Some do most of their work at the start and show fatigue considerably towards the end, some work steadily from start to finish, and some improve all the way. Such differences may be due to slight intercurrent causes, and are not, in the lack of further evidence, to be regarded as indicative of different types of workers. It may be well, however, to present the records of a few individuals, selected more or less at random from among those whose total performance was small, medium, or great. The data given in Table XI show for each individual the per cent. of his total accomplishment done in each successive period of two minutes. Data are given for twenty individuals in all, ten from the Third and Fourth grades and ten from the Seventh and Eighth, five in each group showing the results for attempts and five for accurate work.

For example, Case 22 from the Third and Fourth grade group attempted twenty-one columns in all, twenty-three per cent. of these being attempted in the first three periods, sixty-one per cent. in the first half of the time, and only nineteen per cent. in the last three periods—showing thus a considerable falling off in energy towards the end. Case 367, on the contrary, attempted only fourteen per cent. of his total of one hundred and ninety-one columns in the first three periods, forty-six per cent. in the first half of the time, and twenty-four per cent. in the last three periods.

These differences are found all along the way in going from the poorest total performance to the best. Some poor in total performance are steady workers. Some enthusiastic workers at the start do not do so well at the end, but do well in total performance. Those workers who profit by practice are found among the poorer workers as well as among the better workers as determined by total performances.

TABLE XI

Showing Changes in Working Ability—Third and Fourth Grades
Per Cent. of Columns Attempted

Period	*Case 22 Per Cent.*		*Case 138 Per Cent.*		*Case 276 Per Cent.*		*Case 344 Per Cent.*		*Case 367 Per Cent.*	
1	14		5		5		8		1	
2	14		8		5		6		5	
3	5		10		5		8		8	
		33		23		15		22		14
4	0		8		5		7		8	
5	14		7		7		7		8	
6	9		10		7		6		8	
7	5		5		5		8		8	
		61		53		39		50		46
8	6		7		9		8		8	
9	5		8		8		6		8	
10	5		5		8		8		8	
11	5		8		10		8		2	
12	5		5		10		8		8	
13	5		7		8		6		8	
14	9		7		8		6		8	
	100	19	100	19	100	26	100	20	100	24
Average per cent.	7.1		7.1		7.1		7.1		7.1	
Total columns attempted	21		40		60		79		191	

TABLE XI (*Continued*)

Third and Fourth Grades—Per Cent. of Total Accurate Columns

Period	*Case 142 Per Cent.*		*Case 246 Per Cent.*		*Case 329 Per Cent.*		*Case 350 Per Cent.*		*Case 368 Per Cent.*	
1	26		15		8		7		4	
2	16		15		8		7		7	
3	11		9		6		9		7	
		53		39		22		23		18
4	11		9		10		8		7	
5	0		3		5		6		7	
6	0		0		8		8		7	
7	10		9		8		6		9	
		74		60		53		51		48
8	5		15		8		7		6	
9	11		4		5		6		9	
10	0		12		8		8		8	
11	5		0		6		5		6	
12	0		0		6		9		8	
13	5		9		8		7		8	
14	0		0		8		7		7	
	100	5	100	9	100	22	100	23	100	23
Average per cent.	7.1		7.1		7.1		7.1		7.1	
Total accurate columns	19		33		50		60		110	

TABLE XI (*Continued*)

Seventh and Eighth Grades—Per Cent. of Columns Attempted

Period	*Case 32 Per Cent.*	*Case 52 Per Cent.*	*Case 215 Per Cent.*	*Case 298 Per Cent.*	*Case 336 Per Cent.*
1	8	4	3	4	5
2	6	6	2	5	4
3	5	3	3	5	4
	19	13	8	14	13
4	5	3	3	6	5
5	6	6	2	4	5
6	6	6	1	5	4
7	5	4	3	6	4
8	5	6	3	6	4
9	6	4	4	4	4
10	4	3	4	5	4
11	4	3	6	5	4
	60	48	34	55	47
12	3	6	7	4	5
13	6	6	8	4	4
14	4	4	8	5	4
15	5	3	6	4	4
16	5	3	6	4	4
17	4	6	6	5	4
18	3	6	6	4	4
19	3	6	5	4	4
20	3	6	5	5	4
21	3	6	4	6	4
	100 9	100 18	100 14	100 15	100 12
Average per cent.	4.7	4.7	4.7	4.7	4.7
Total columns attempted	77	89	133	160	223

TABLE XI (*Continued*)

Seventh and Eighth Grades—Per Cent. of Total Accurate Columns

Period	*Case 20 Per Cent.*		*Case 64 Per Cent.*		*Case 208 Per Cent.*		*Case 304 Per Cent.*		*Case 343 Per Cent.*	
1	10		10		2		5		4	
2	10		10		5		5		4	
3	12		4		2		5		5	
		32		24		9		15		13
4	7		6		3		5		6	
5	5		6		6		4		4	
6	3		6		6		5		5	
7	5		4		3		4		4	
8	7		8		6		5		4	
9	0		0		3		4		5	
10	3		6		6		5		4	
		62		58		42		47		45
11	5		4		6		4		6	
12	0		4		6		5		4	
13	3		5		3		5		6	
14	3		3		5		5		5	
15	3		4		5		5		5	
16	0		2		6		5		4	
17	4		4		5		5		5	
18	5		2		5		4		6	
19	3		4		5		5		6	
20	5		6		6		5		4	
21	7		4		6		6		4	
	100	15	100	14	100	17	100	16	100	14
Average per cent.	4.7		4.7		4.7		4.7		4.7	
Total of accurate columns	40		52		88		135		242	

Fatigue As Related to Total Performance

The purpose of this manner of handling the materials of the experiment is to see whether the work curve would be the same for those who accomplish much and for those who accomplish little in the total time; or whether the better workers fatigue more or less rapidly than the poorer workers.

Third and Fourth Grade Group divided into Tenths.

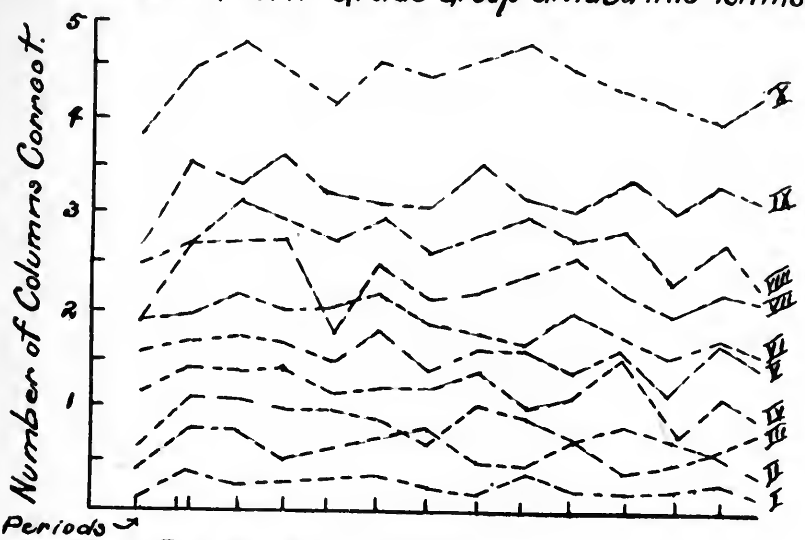


Fig. 5- Accurates

The two groups are divided into sub-groups on the basis of the total amount done in twenty-eight and forty-two minutes respectively. The 368 children of the Third and Fourth grades were arranged in rank order from lowest to highest, according to the total attempted in the twenty-eight minutes, and again according to the total accurate in the twenty-eight minutes, and this rank order arrangement was cut into ten parts. Each tenth of the 368 makes a sub-group, the sub-groups running from lowest to highest or upper tenth. The average of each sub-group was determined and curves were formed for these ten sub-groups. There are twenty curves for the Third and Fourth grades. Likewise, there are twenty curves for the Seventh and Eighth grades. The 343 children in this group were arranged in rank order from lowest to

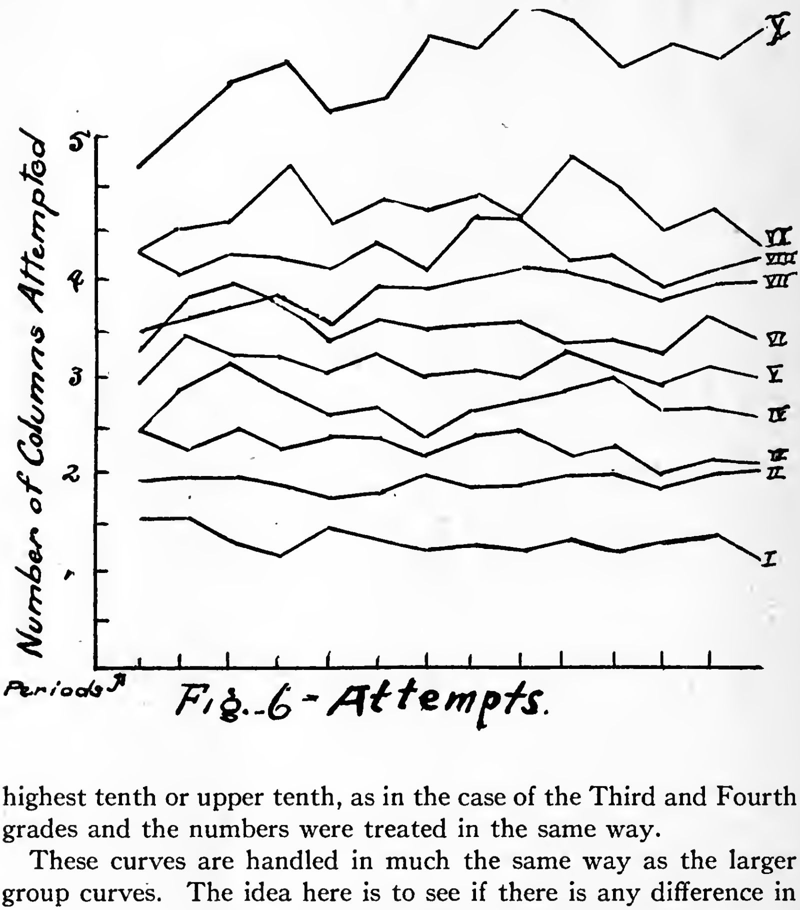


Fig.-6- Attempts.

highest tenth or upper tenth, as in the case of the Third and Fourth grades and the numbers were treated in the same way.

These curves are handled in much the same way as the larger group curves. The idea here is to see if there is any difference in the form of a curve for a poor total accomplishment from that of the better accomplishment.

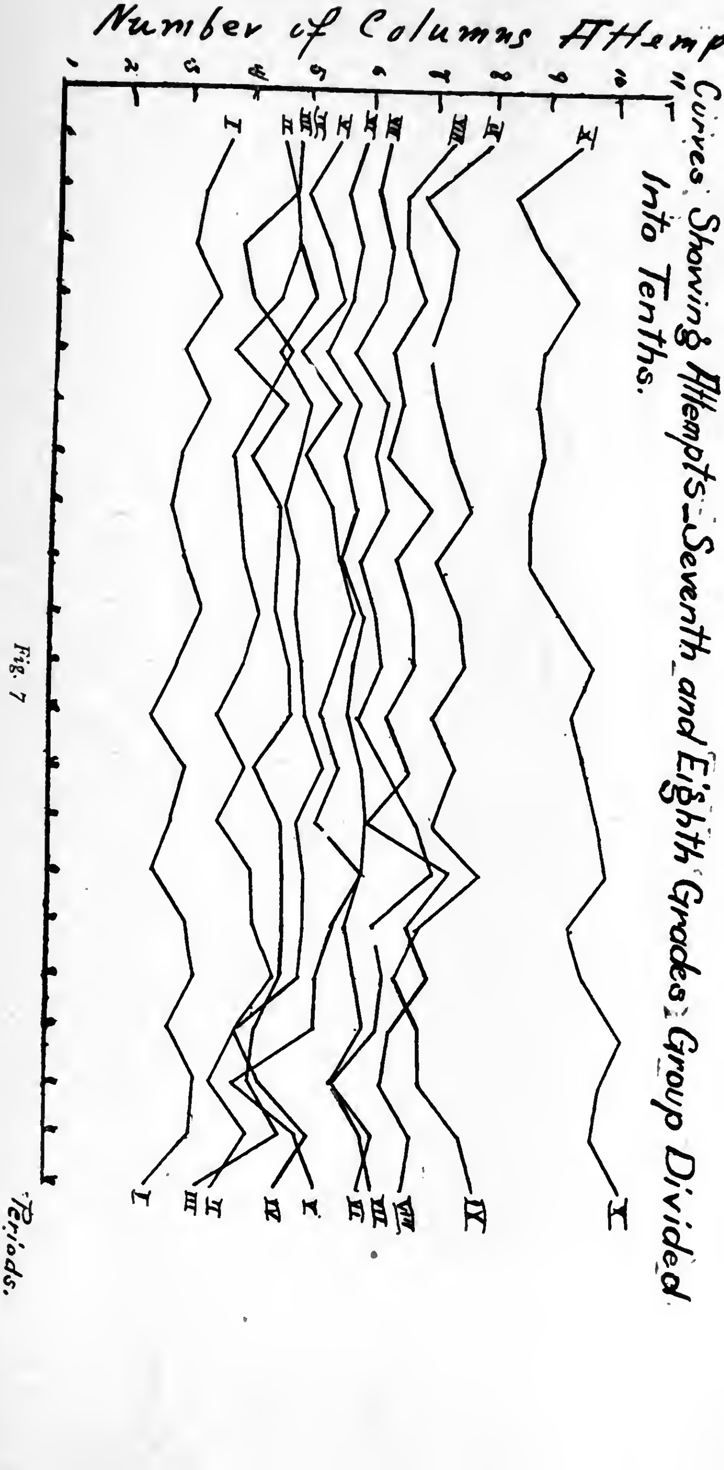

Number of Columns Attempted
Curves Showing Attempts—Seventh and Eighth Grades; Group Divided Into Tenths.
Periods.

Fig. 7

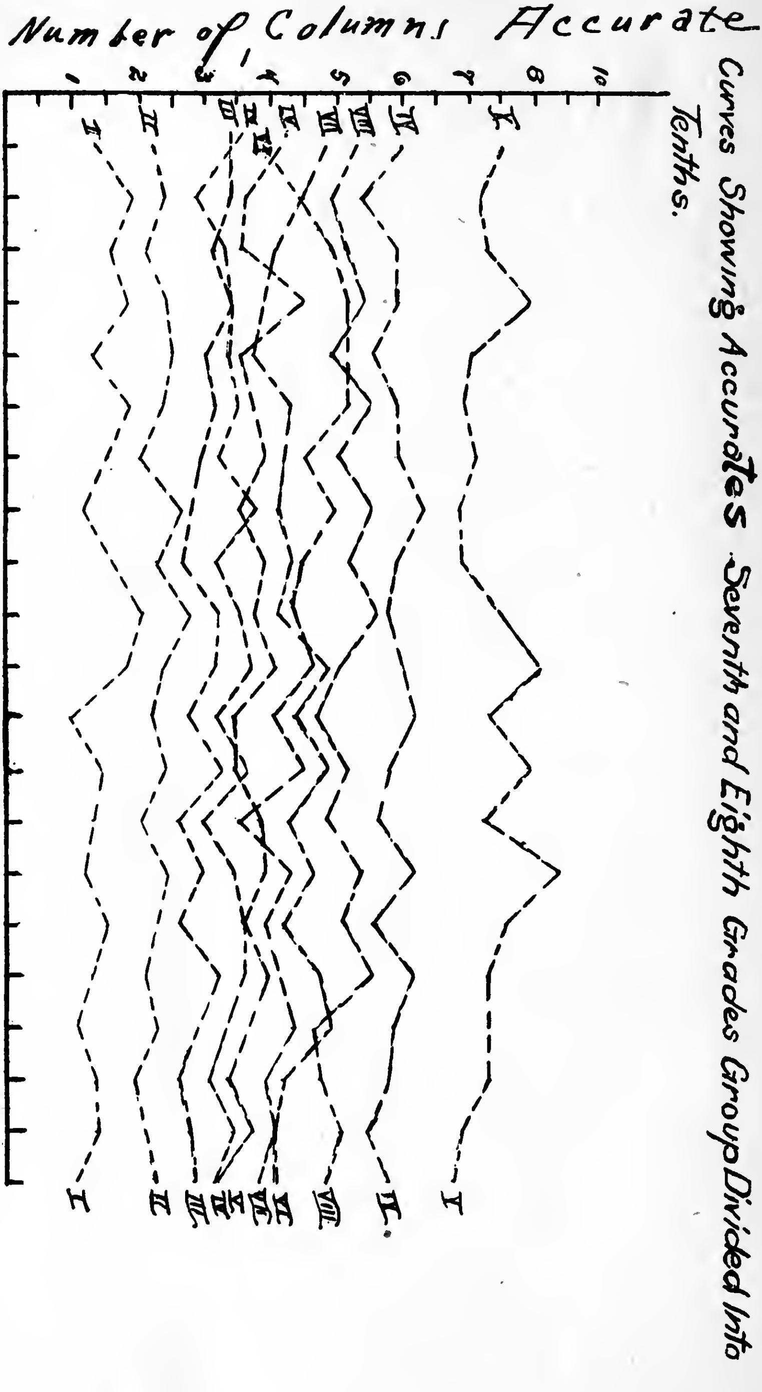

Number of Columns Accurate
Curves Showing Accurates Seventh and Eighth Grades Group Divided Into Tenths.
1
2
3
4
5
6
7
8
10
I
II
III
IV
V
VI
VII
VIII
IX
X

Fig. 8

TABLE XII

Attempts of Third and Fourth Grades

The 368 cases are divided into ten groups on the basis of total amount done in twenty-eight minutes, going from lowest tenth to upper tenth. Below are given the averages for each

First (Lowest) Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	1.52	.75	.11
2	1.5	.63	.09
3	1.27	.59	.08
4	1.19	.59	.08
5	1.44	.56	.08
6	1.36	.67	.09
7	1.22	.49	.07
8	1.33	.60	.08
9	1.22	.42	.06
10	1.36	.61	.09
11	1.22	.46	.06
12	1.33	.51	.07
13	1.38	.59	.07
14	1.08	.39	.05

Second Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	1.84	.55	.08
2	1.98	.41	.06
3	1.98	.31	.03
4	1.87	.41	.06
5	1.76	.51	.08
6	1.82	.46	.06
7	2.0	.27	.1
8	1.9	.42	.06
9	1.92	.5	.07
10	2.05	.45	.06
11	2.05	.3	.04
12	1.85	.46	.06
13	2.02	.36	.05
14	2.08	.6	.08

Third Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	2.43	.71	.1
2	2.29	.64	.09
3	2.49	.58	.08
4	2.37	.54	.08
5	2.4	.6	.09
6	2.52	.49	.07
7	2.27	.54	.07
8	2.4	.56	.08
9	2.45	.58	.08
10	2.29	.53	.07
11	2.37	.55	.07
12	2.02	.47	.07
13	2.21	.29	.04
14	2.08	.6	.08

Fourth Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	2.43	.66	.09
2	2.92	.55	.08
3	3.16	.47	.07
4	2.84	.6	.08
5	2.63	.61	.09
6	2.79	.54	.08
7	2.43	.6	.08
8	2.76	.47	.07
9	2.79	.56	.08
10	2.9	.47	.07
11	3.02	.52	.08
12	2.71	.79	.11
13	2.73	.54	.08
14	2.71	.59	.08

TABLE XII (*Continued*)

Attempts of Third and Fourth Grades

Fifth Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	2.95	.88	.12
2	3.41	.5	.07
3	3.29	.45	.06
4	3.24	.51	.07
5	3.05	.46	.06
6	3.35	.59	.08
7	3.	.37	.05
8	3.1	.52	.07
9	3.05	.46	.06
10	3.32	.58	.08
11	3.13	.51	.07
12	2.92	.5	.07
13	3.21	.54	.08
14	3.05	.75	.10

Sixth Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	3.43	.79	.11
2	3.64	.68	.10
3	3.7	.58	.08
4	3.79	.50	.07
5	3.43	.59	.08
6	3.62	.52	.07
7	3.51	.58	.08
8	3.57	.6	.08
9	3.64	.56	.08
10	3.37	.65	.09
11	3.45	.68	.10
12	3.27	.55	.08
13	3.76	.61	.09
14	3.32	.76	.10

Seventh Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	3.37	.76	.11
2	3.86	.61	.09
3	3.92	.40	.06
4	3.86	.53	.07
5	3.57	.62	.09
6	3.98	.41	.06
7	3.92	.62	.09
8	4.05	.46	.06
9	4.18	.5	.07
10	4.16	.52	.07
11	4.05	.46	.06
12	3.87	.55	.08
13	4.02	.47	.07
14	4.05	.46	.06

Eighth Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	4.29	.85	.12
2	4.02	.69	.10
3	4.34	.64	.09
4	4.29	.64	.09
5	4.16	.56	.08
6	4.40	.63	.09
7	4.13	.67	.10
8	4.56	.79	.11
9	4.64	.75	.10
10	4.29	.79	.11
11	4.32	.61	.09
12	3.94	.67	.093
13	4.16	.64	.089
14	4.34	.93	.132

TABLE XII (*Continued*)

Attempts of Third and Fourth Grades

Ninth Tenth—36 Cases				*Tenth Tenth—36 Cases*			
Period	*Average*	*A. D.*	*P. E.*	*Period*	*Average*	*A. D.*	*P. E.*
1	4.21	.85	.12	1	5.16	1.18	.17
2	4.61	.71	.10	2	5.75	1.11	.16
3	4.58	.79	.11	3	6.05	1.13	.16
4	5.22	.65	.09	4	6.25	1.13	.16
5	4.62	.61	.09	5	5.78	1.22	.17
6	4.89	.50	.07	6	5.39	1.09	.15
7	4.81	.65	.09	7	6.61	1.57	.22
8	4.95	.58	.08	8	6.58	1.47	.20
9	4.78	.72	.10	9	6.92	1.60	.22
10	5.36	.7	.10	10	6.75	1.56	.22
11	5.00	.55	.08	11	6.36	.94	.13
12	4.64	.73	.10	12	6.52	1.42	.20
13	4.78	.67	.09	13	6.41	1.48	.20
14	4.44	.78	.11	14	6.70	1.44	.20

TABLE XIII

Accurates of Third and Fourth Grades

The 368 cases are divided into ten groups on the basis of total amount done in twenty-eight minutes, going from lowest tenth to upper tenth. Below are given the averages for each group

First (Lowest) Tenth—37 Cases				*Second Tenth—37 Cases*			
Period	*Average*	*A. D.*	*P. E.*	*Period*	*Average*	*A. D.*	*P. E.*
1	.13	.26	.04	1	.37	.56	.08
2	.36	.49	.06	2	.81	.77	.11
3	.27	.40	.06	3	.81	.65	.09
4	.24	.36	.05	4	.51	.58	.08
5	.29	.43	.06	5	.67	.58	.08
6	.32	.43	.06	6	.70	.57	.08
7	.21	.34	.05	7	.78	.59	.08
8	.16	.27	.04	8	.59	.57	.08
9	.40	.56	.08	9	.48	.52	.08
10	.21	.34	.05	10	.72	.67	.09
11	.21	.34	.05	11	.40	.54	.08
12	.21	.33	.05	12	.59	.67	.09
13	.24	.52	.08	13	.62	.60	.08
14	.18	.31	.04	14	.43	.58	.08

TABLE XIII (*Continued*)

Accurates of Third and Fourth Grades

Third Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	.59	.64	.08
2	1.18	.81	.11
3	1.13	.58	.08
4	1.00	.7	.1
5	1.05	1.3	.18
6	.91	.59	.08
7	.75	.65	.08
8	1.08	1.00	.14
9	.97	.57	.08
10	.78	.72	.10
11	.82	.60	.08
12	.89	.74	.10
13	.64	.73	.10
14	.86	.56	.08

Fourth Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	1.26	.82	.11
2	1.40	.83	.11
3	1.37	1.1	.15
4	1.43	.66	.09
5	1.13	.73	.10
6	1.24	.65	.09
7	1.24	.71	.10
8	1.43	.74	.10
9	1.05	.56	.08
10	1.18	.78	.11
11	1.55	.76	.11
12	.72	.76	.11
13	1.21	.79	.11
14	.90	.67	.09

Fifth Tenth—37 Cases.

Period	*Average*	*A. D.*	*P. E.*
1	1.55	.74	.10
2	1.71	.78	.10
3	1.79	.78	.10
4	1.79	.77	.10
5	1.55	.74	.10
6	1.87	.76	.10
7	1.43	.66	.09
8	1.60	.74	.10
9	1.60	.90	.13
10	1.45	.79	.10
11	1.65	.68	.09
12	1.18	.63	.09
13	1.68	.61	.09
14	1.49	.85	.13

Sixth Tenth—37 Cases

Period	*Average*	*A. D.*	*P. E.*
1	1.90	1.04	.15
2	1.95	.88	.12
3	2.18	.91	.12
4	2.00	.70	.10
5	2.08	.64	.09
6	2.29	.80	.11
7	1.95	.77	.11
8	1.87	1.01	.14
9	1.68	.80	.11
10	2.02	.84	.11
11	1.73	.98	.14
12	1.63	.71	.10
13	1.73	.75	.11
14	1.63	1.03	.14

TABLE XIII (*Continued*)

Accurates of Third and Fourth Grades

Seventh Tenth—37 Cases

Period	Average	A. D.	P. E.
1	1.89	.88	.13
2	2.71	1.00	.14
3	2.73	1.03	.14
4	2.73	.72	.10
5	1.79	.61	.09
6	2.52	.84	.11
7	2.10	.62	.09
8	2.21	.83	.11
9	2.44	.88	.13
10	2.60	.79	.10
11	2.27	.92	.12
12	2.08	.91	.12
13	2.35	.84	.11
14	2.16	.80	.11

Eighth Tenth—37 Cases

Period	Average	A. D.	P. E.
1	2.49	1.04	.15
2	2.79	1.02	.14
3	3.16	.74	.10
4	2.98	.89	.12
5	2.79	.86	.12
6	3.00	.64	.09
7	2.68	.88	.12
8	2.82	1.01	.14
9	3.02	.79	.10
10	2.84	1.06	.14
11	2.95	.72	.10
12	2.29	.86	.12
13	2.79	.97	.14
14	2.29	.74	.10

Ninth Tenth—36 Cases

Period	Average	A. D.	P. E.
1	2.73	1.04	.15
2	3.57	.85	.12
3	3.33	.88	.12
4	3.63	.82	.12
5	3.27	.82	.12
6	3.10	.73	.10
7	3.13	.74	.10
8	4.08	1.36	.19
9	3.27	1.05	.15
10	3.16	.86	.12
11	3.45	.91	.12
12	3.08	.82	.12
13	3.40	.99	.14
14	3.27	1.01	.15

Tenth Tenth—36 Cases

Period	Average	A. D.	P. E.
1	3.84	.96	.13
2	4.5	1.16	.16
3	4.78	1.02	.15
4	4.41	1.19	.17
5	4.19	.93	.12
6	4.67	1.03	.15
7	4.47	1.02	.15
8	4.73	.82	.12
9	4.89	.96	.14
10	4.55	1.05	.15
11	4.33	1.14	.16
12	4.25	1.34	.19
13	4.08	1.36	.19
14	4.30	1.29	.18

TABLE XIV

Attempts of Seventh and Eighth Grades

The 343 cases are divided into ten groups on the basis of total amount done in forty-two minutes, going from lowest tenth to upper tenth. Below are given the averages for each group

First (Lowest) Tenth—35 Cases

Period	*Average*	*A. D.*	*P. E.*
1	3.71	1.13	.16
2	3.42	.89	.13
3	3.37	.65	.09
4	3.66	.87	.12
5	3.17	.68	.10
6	3.55	.73	.11
7	3.17	.61	.09
8	3.08	.63	.09
9	3.20	.56	.08
10	3.40	.57	.08
11	3.15	.57	.08
12	2.83	.64	.09
13	3.22	.55	.08
14	3.09	.68	.10
15	2.87	.64	.09
16	3.25	.70	.10
17	3.31	.77	.11
18	3.06	.70	.10
19	3.57	.98	.14
20	3.57	.84	.12
21	2.86	.65	.09

Second Tenth—35 Cases

Period	*Average*	*A. D.*	*P. E.*
1	4.65	1.24	.19
2	4.85	.89	.13
3	4.00	.74	.10
4	4.14	.63	.09
5	4.88	1.02	.14
6	4.20	.60	.09
7	3.95	.50	.07
8	4.00	.80	.12
9	4.11	.61	.09
10	4.33	.58	.08
11	4.22	.73	.10
12	3.83	.48	.07
13	4.17	.80	.12
14	3.8	.67	.10
15	4.34	.84	.12
16	4.31	.83	.11
17	4.58	.72	.10
18	4.11	.82	.12
19	3.75	.67	.10
20	4.31	.85	.12
21	4.85	1.05	.15

Third Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	4.88	.76	.09
2	4.74	.62	.12
3	4.83	.68	.11
4	4.68	.57	.12
5	3.88	.53	.11
6	4.68	.57	.07
7	4.21	.69	.11
8	4.77	.68	.09
9	4.71	.71	.07
10	4.98	.57	.08
11	4.79	.71	.10
12	4.38	.56	.08
13	5.00	.52	.10

Fourth Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	4.89	1.03	.14
2	4.86	.51	.07
3	4.94	.66	.19
4	5.21	.57	.08
5	4.74	.63	.09
6	5.11	.73	.10
7	5.02	.51	.07
8	4.94	.67	.10
9	5.00	.70	.10
10	5.06	.78	.11
11	5.06	.50	.07
12	5.16	.80	.11
13	5.52	1.09	.15

TABLE XIV (*Continued*)

Attempts of Seventh and Eighth Grades

Third Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
14	4.17	.63	.10
15	4.62	.75	.10
16	4.74	.49	.08
17	4.77	.76	.08
18	4.65	.86	.08
19	4.44	.72	.10
20	4.83	.86	.09
21	3.59	.61	.11

Fourth Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
14	5.17	.80	.12
15	5.26	.85	.12
16	5.29	.86	.12
17	5.21	.67	.19
18	4.12	1.04	.15
19	4.53	.99	.14
20	5.23	1.11	.16
21	4.86	.69	.10

Fifth Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	5.53	1.34	.19
2	5.09	.86	.12
3	5.41	.69	.10
4	5.62	.75	.11
5	4.97	.86	.12
6	5.59	.67	.10
7	5.05	.72	.10
8	5.53	.73	.11
9	5.68	.62	.09
10	5.89	.86	.12
11	5.68	.83	.12
12	5.45	.64	.09
13	5.71	.75	.11
14	5.38	.62	.09
15	6.29	1.18	.17
16	5.74	.84	.12
17	5.71	.70	.10
18	5.45	.72	.10
19	4.42	1.06	.15
20	5.17	1.04	.15
21	5.38	.81	.12

Sixth Tenth—35 Cases

Period	*Average*	*A. D.*	*P. E.*
1	5.98	1.29	.19
2	5.75	.62	.09
3	5.89	.73	.11
4	5.80	.58	.08
5	5.52	.75	.11
6	5.92	.63	.09
7	5.72	.94	.14
8	5.92	.84	.12
9	5.63	.56	.08
10	6.08	.85	.12
11	5.92	.48	.07
12	5.89	.56	.08
13	6.02	.49	.07
14	6.20	.68	.10
15	6.22	1.06	.15
16	5.78	.68	.10
17	5.92	.74	.11
18	6.17	.83	.12
19	5.80	.58	.08
20	6.36	.90	.13
21	6.22	.74	.11

Seventh Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	6.26	1.00	.14
2	6.17	.75	.11
3	6.32	.92	.13
4	6.26	.70	.10
5	5.74	.67	.10

Eighth Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	7.29	.98	.13
2	6.68	.92	.13
3	6.77	.59	.08
4	6.95	.72	.10
5	6.41	.62	.09

TABLE XIV (*Continued*)

Attempts of Seventh and Eighth Grades

Seventh Tenth—34 Cases				*Eighth Tenth—34 Cases*			
Period	*Average*	*A. D.*	*P. E.*	*Period*	*Average*	*A. D.*	*P. E.*
6	6.38	.86	.12	6	6.62	.62	.09
7	6.33	.74	.11	7	6.45	.65	.09
8	6.35	.84	.12	8	7.05	.66	.10
9	5.92	.52	.07	9	6.71	.59	.08
10	6.23	.81	.12	10	6.80	.61	.09
11	6.32	.77	.11	11	6.83	.61	.09
12	6.00	.76	.11	12	6.59	.64	.09
13	6.48	.88	.13	13	6.74	.62	.09
14	6.98	.97	.14	14	6.38	.99	.14
15	7.23	1.14	.16	15	7.55	1.00	.14
16	6.44	.87	.13	16	6.95	.78	.11
17	6.58	.93	.13	17	7.21	.89	.13
18	6.41	.72	.10	18	6.69	.78	.11
19	5.68	.86	.12	19	6.62	.70	.10
20	6.26	1.19	.17	20	7.05	.78	.11
21	6.35	1.02	.15	21	6.92	.98	.14

Ninth Tenth—34 Cases				*Tenth Tenth—34 Cases*			
Period	*Average*	*A. D.*	*P. E.*	*Period*	*Average*	*A. D.*	*P. E.*
1	7.83	1.13	.16	1	9.29	1.26	.18
2	6.83	.93	.13	2	8.61	1.15	.17
3	7.59	.78	.11	3	8.86	1.18	.17
4	7.50	.79	.11	4	9.44	1.31	.19
5	7.14	.71	.10	5	8.92	1.39	.20
6	7.29	.59	.08	6	8.89	1.09	.16
7	7.41	.83	.12	7	8.98	1.15	.17
8	7.76	.92	.13	8	8.95	1.12	.16
9	7.26	.69	.10	9	8.83	1.04	.15
10	7.65	.82	.12	10	9.38	1.25	.18
11	7.74	.78	.11	11	9.27	1.30	.19
12	7.25	.78	.11	12	9.52	1.61	.23
13	7.62	.79	.11	13	9.77	1.60	.23
14	7.45	.88	.13	14	9.95	1.41	.20
15	8.27	1.48	.21	15	10.11	1.60	.23
16	7.02	.62	.09	16	9.68	1.73	.25
17	6.77	.79	.11	17	9.98	1.50	.22
18	7.28	1.05	.15	18	10.44	2.18	.31
19	7.29	1.01	.14	19	10.11	1.56	.22
20	7.92	.65	.09	20	10.14	2.09	.30
21	8.14	1.25	.18	21	10.79	2.59	.37

TABLE XV

Accurates of Seventh and Eighth Grades

The 343 cases are divided into ten groups on the basis of total amount done in forty-two minutes, going from lowest tenth to upper tenth. Below are given the averages for each group

First (Lowest) Tenth—35 Cases

Period	*Average*	*A. D.*	*P. E.*
1	1.43	1.42	.07
2	1.86	1.37	.06
3	1.77	1.19	.05
4	1.80	.96	.04
5	1.49	1.01	.05
6	1.89	.98	.04
7	1.67	1.03	.06
8	1.52	.92	.04
9	1.80	.88	.04
10	2.17	.85	.04
11	1.83	.83	.04
12	1.02	.66	.03
13	1.54	1.04	.06
14	1.45	.97	.04
15	1.40	.97	.04
16	1.58	.82	.04
17	1.48	1.04	.06
18	1.22	.84	04
19	1.37	.95	.04
20	1.31	.93	.04
21	1.14	.71	.03

Second Tenth—35 Cases

Period	*Average*	*A. D.*	*P. E.*
1	2.22	1.00	.14
2	2.37	.98	.14
3	2.20	1.09	.16
4	2.40	1.00	.14
5	2.55	.62	.09
6	2.49	.52	.07
7	2.08	.96	.13
8	2.69	1.23	.17
9	2.40	.85	.12
10	2.75	.73	.10
11	2.43	.93	.13
12	2.20	.77	.11
13	2.45	.94	.13
14	2.00	.85	.12
15	2.45	.92	.13
16	2.31	1.15	.16
17	2.28	1.04	.15
18	2.34	1.02	.15
19	2.02	.83	.12
20	2.05	.81	.12
21	2.20	1.08	.16

Third Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	3.32	1.06	.18
2	2.95	.88	.12
3	3.26	.80	.12
4	3.41	.82	.12
5	3.08	.99	.14
6	3.05	1.06	.15
7	3.00	.94	.13
8	2.89	1.09	.16
9	2.74	.91	.13
10	3.17	1.05	.16
11	3.17	1.09	.16
12	2.54	1.13	.16
13	3.14	.88	.13

Fourth Tenth—34 Cases

Period	*Average*	*A. D.*	*P. E.*
1	3.35	1.10	.16
2	3.36	.78	.11
3	3.17	1.10	.16
4	3.48	.82	.12
5	3.44	.79	.10
6	3.50	.82	.12
7	3.23	.84	.12
8	3.71	.92	.13
9	3.23	.87	.12
10	3.53	1.05	.15
11	3.68	.94	.13
12	3.17	.95	.14
13	3.75	1.27	.18

TABLE XV (*Continued*)

Accurates of Seventh and Eighth Grades

Third Tenth—34 Cases

Period	Average	A. D.	P. E.
14	2.69	1.08	.16
15	2.74	.89	.13
16	2.66	1.05	.15
17	3.08	1.06	.15
18	2.83	.84	.13
19	2.44	1.02	.15
20	2.71	.93	.13
21	2.77	1.02	.15

Fourth Tenth—34 Cases

Period	Average	A. D.	P. E.
14	2.95	.95	.14
15	3.38	1.01	.15
16	3.53	.99	.14
17	3.59	.99	.14
18	3.33	1.09	.15
19	3.11	.76	.11
20	3.14	1.30	.19
21	3.11	1.18	.16

Fifth Tenth—34 Cases

Period	Average	A. D.	P. E.
1	4.09	1.10	.16
2	3.68	1.20	.15
3	3.56	.93	.13
4	4.58	.78	.11
5	3.59	.96	.13
6	3.68	.98	.14
7	3.77	.90	.13
8	3.65	.90	.13
9	3.91	.78	.11
10	3.80	.94	.13
11	4.00	1.11	.16
12	3.40	1.08	.15
13	3.50	.97	.14
14	3.80	1.19	.17
15	3.86	1.06	.15
16	3.62	.97	.14
17	3.80	.80	.12
18	3.65	1.23	.18
19	3.11	.96	.14
20	3.59	1.41	.20
21	3.20	1.41	.20

Sixth Tenth—35 Cases

Period	Average	A. D.	P. E.
1	3.98	1.49	.21
2	4.44	1.11	.16
3	4.17	1.03	.15
4	3.89	1.01	.15
5	3.67	1.30	.19
6	4.11	1.14	.16
7	4.26	1.20	.17
8	4.80	1.12	.16
9	4.36	.94	.14
10	4.11	1.20	.17
11	4.62	.95	.14
12	4.00	.68	.10
13	4.50	1.18	.16
14	3.68	1.10	.16
15	4.23	1.21	.17
16	4.00	1.14	.16
17	4.14	.93	.13
18	4.26	1.22	.18
19	4.11	.77	.11
20	4.13	1.01	.15
21	4.12	1.30	.19

Seventh Tenth—34 Cases

Period	Average	A. D.	P. E.
1	4.66	1.55	.22
2	4.58	1.29	.19
3	4.89	1.17	.17
4	5.09	.96	.13
5	5.06	.84	.12

Eighth Tenth—34 Cases

Period	Average	A. D.	P. E.
1	5.29	1.12	.17
2	4.92	.76	.11
3	5.05	1.12	.16
4	5.44	1.05	.15
5	4.83	1.19	.17

TABLE XV (*Continued*)

Accurates of Seventh and Eighth Grades

Seventh Tenth—34 Cases				*Eighth Tenth—34 Cases*			
Period	*Average*	*A. D.*	*P. E.*	*Period*	*Average*	*A. D.*	*P. E.*
6	5.03	.57	.08	6	5.50	1.08	.16
7	4.41	1.04	.15	7	5.17	.93	.13
8	4.77	.63	.09	8	5.58	1.39	.19
9	4.48	1.00	.14	9	5.08	1.06	.15
10	4.38	1.25	.17	10	5.59	1.08	.16
11	4.76	1.31	.19	11	5.00	.94	.13
12	4.23	1.02	.15	12	4.71	1.12	.16
13	4.83	1.07	.15	13	5.20	.90	.13
14	4.29	1.24	.17	14	4.95	1.12	.16
15	4.74	1.20	.17	15	5.35	1.41	.16
16	4.29	1.08	.16	16	5.17	1.29	.18
17	4.74	1.02	.15	17	5.58	1.03	.15
18	4.86	1.04	.15	18	4.65	1.13	.16
19	4.26	1.28	.17	19	4.77	.92	.13
20	4.11	1.08	.16	20	5.11	1.18	.17
21	4.00	1.29	.18	21	5.00	1.11	.16

Ninth Tenth—34 Cases				*Tenth Tenth—34 Cases*			
Period	*Average*	*A. D.*	*P. E.*	*Period*	*Average*	*A. D.*	*P. E.*
1	5.95	1.37	.19	1	7.59	1.43	.20
2	5.45	1.09	.16	2	7.13	1.25	.17
3	5.95	1.00	.14	3	7.23	1.51	.16
4	5.92	.77	.11	4	7.80	1.36	.20
5	5.59	1.01	.15	5	7.14	1.08	.16
6	5.83	.98	.13	6	6.89	1.49	.20
7	5.86	.82	.12	7	7.08	1.03	.15
8	6.29	1.01	.15	8	6.92	1.27	.17
9	5.92	.87	.13	9	6.98	1.15	.16
10	5.89	1.06	.15	10	7.42	1.45	.20
11	5.95	1.12	.16	11	8.02	1.33	.20
12	6.08	.99	.13	12	7.35	1.35	.20
13	5.83	.98	.13	13	7.95	1.54	.16
14	5.80	1.11	.16	14	7.32	1.43	.20
15	5.24	1.22	.17	15	8.32	1.66	.24
16	5.56	1.16	.17	16	7.62	1.29	.17
17	6.23	.87	.13	17	7.38	1.74	.25
18	5.98	1.09	.16	18	7.32	1.83	.26
19	5.83	.95	.13	19	7.17	1.40	.20
20	5.55	1.08	.16	20	7.17	1.20	.16
21	5.77	1.07	.15	21	6.83	1.53	.16

TABLE XVI

Third and Fourth Grades.—Attempts

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
1	1	14	1.43	1.26	12% loss	1.37	1.31	1.27	1.00 .95 .92
2	14	5	1.93	1.99	3% gain	1.92	1.88	2.01	1.00 .97 1.04
3	6	12	2.40	2.10	12% loss	2.40	2.41	2.19	1.00 1.00 .91
4	3	1 and 7	2.84	2.72	4% loss	2.84	2.68	2.81	1.00 .94 .99
5	2	12	3.22	3.06	5% loss	3.22	3.11	3.13	1.00 .96 .97
6	4	12	3.59	3.45	4% loss	3.64	3.55	3.43	1.00 .96 .94
7	9	1	3.72	3.98	7% gain	3.75	3.94	4.03	1.00 1.05 1.07
8	8	12	4.22	4.16	1% loss	4.24	4.40	4.22	1.00 1.037 .99
9	10	1	4.47	4.62	3% gain	4.65	4.81	4.84	1.00 1.034 1.04
10	9	1	5.65	6.54	16% gain	5.80	6.36	6.55	1.00 1.09 1.13

Third and Fourth Grades—Accurates

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
1	9	1	.25	.21	16% loss	.25	.28	.21	1.00 1.12 .84
2	2 3	1	.66	.55	17% loss	.63	.64	.55	1.00 1.00 .87

TABLE XVI (*Continued*)

Third and Fourth Grades—Accurates

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
3	2	1	.97	.80	18% loss	.98	.95	.80	1.00 .96 .81
4	11	12	1.34	.94	30% loss	1.36	1.22	1.11	1.00 .89 .81
5	6	12	1.68	1.45	14% loss	1.71	1.61	1.49	1.00 .94 .87
6	6	14	2.01	1.66	17% loss	2.01	1.97	1.75	1.00 .98 .87
7	3 4	5	2.44	2.20	10% loss	2.51	2.21	2.29	1.00 .88 .91
8	3	12	2.81	2.46	12% loss	2.85	2.86	2.63	1.00 1.003 .92
9	8	1	3.21	3.25	1% gain	3.32	3.37	3.27	1.00 1.01 .98
10	9	1	4.37	4.21	4% loss	4.38	4.59	4.30	1.00 1.04 .98

TABLE XVII

Seventh and Eighth Grades—Attempts

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
1	1	12	3.50	3.53	5% loss	3.44	3.14	3.21	1.00 .91 .93
2	5	19	4.50	4.30	5% loss	4.38	4.07	4.32	1.00 .93 .98
3	13	21	4.82	4.29	11% loss	4.70	4.69	4.66	1.00 .99 .99
4	13	18	4.90	4.87	1% loss	4.97	5.13	4.93	1.00 1.03 .99

TABLE XVII (*Continued*)

Seventh and Eighth Grades—Attempts

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
5	15	19	5.34	4.99	7% loss	5.32	5.62	5.45	1.00 1.05 1.02
6	20	5	5.87	6.13	4% gain	5.80	5.95	6.07	1.00 1.02 1.04
7	15	19	6.25	6.10	2% loss	6.21	6.23	6.42	1.00 1.00 1.03
8	15	14	6.91	6.86	1% loss	6.74	6.73	7.01	1.00 .99 1.04
9	15	17	7.42	7.78	5% gain	7.37	7.53	7.53	1.00 1.02 1.02
10	21	2	8.92	10.35	16% gain	9.10	9.38	10.18	1.04 1.03 1.12

Seventh and Eighth Grades—Accurates

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
1	10	12	1.69	1.27	25% loss	1.70	1.62	1.36	1.00 .95 .80
2	10	14	2.26	2.09	8% loss	2.33	2.42	2.23	1.00 1.08 .95
3	4	19	3.18	2.64	17% loss	3.15	2.91	2.75	1.00 .92 .87
4	13	14	3.29	3.12	5% loss	3.36	3.43	3.31	1.00 1.02 .98
5	4	19	3.78	3.30	13% loss	3.85	3.23	3.55	1.00 .84 .92
6	8	5	4.20	4.12	2% loss	4.07	4.29	4.14	1.00 1.05 1.01
7	4	21	4.70	4.12	12% loss	4.82	4.63	4.42	1.00 .93 .91

TABLE XVII (*Continued*)

Seventh and Eighth Grades—Accurates

Sub-Groups	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
8	10	18	5.09	4.96	3% loss	5.17	5.16	5.09	1.00
									.99
									.98
9	8	2	5.78	5.72	1% loss	5.79	5.97	5.88	1.00
									1.03
									1.01
10	15	21	7.32	7.11	3% loss	7.27	7.42	7.42	1.00
									1.02
									1.02

EXPLANATION OF TABLES NOS. XVI AND XVII

In these tables we have under

I. The maximum period for each curve.
II. The minimum period for each curve.
III. Average columns for first three periods.
IV. Average columns for last three periods.
V. Per cent. of gain or loss when III and IV are compared with III as a basis.
VI. Average of first third of curve.
VII. Average of second third of curve.
VIII. Average of last third of curve.
IX. Relative value of VI, VII, and VIII, using VI as a basis.

The figures under V show that the group fatigued or did not fatigue—according as there was loss or gain. IX gives an idea of the general form of the curve.

To begin with the facts as summarized in Table XVI. The location of the maximum (Column I) is certainly later with the better workers, when attempts are considered, in either the Third and Fourth or the Seventh and Eighth grades; the minimum, on the other hand, tends to come earlier with the better groups than with the poorer. In the case of accurates, no clear relation appears between the location of the maximum or minimum and the total accomplishment.

Column V shows pretty clearly that the losses tend to be smaller towards the bottom of the column, which means that the better workers fatigued less than the poorer workers. In the case of attempts, the loss actually gives way to a gain, while in the case of accurates, the best workers lose very little at the end as compared with the beginning. Column IX, attentively considered, will show the same thing.

But, undoubtedly, the best presentation of the results is afforded by the curves.[1] The curves for attempts, in both the younger and the older group, show a clear general rise in the case of the upper tenth, indicating that the best workers increased their speed during the experiment; while the curve for the lowest tenth shows the

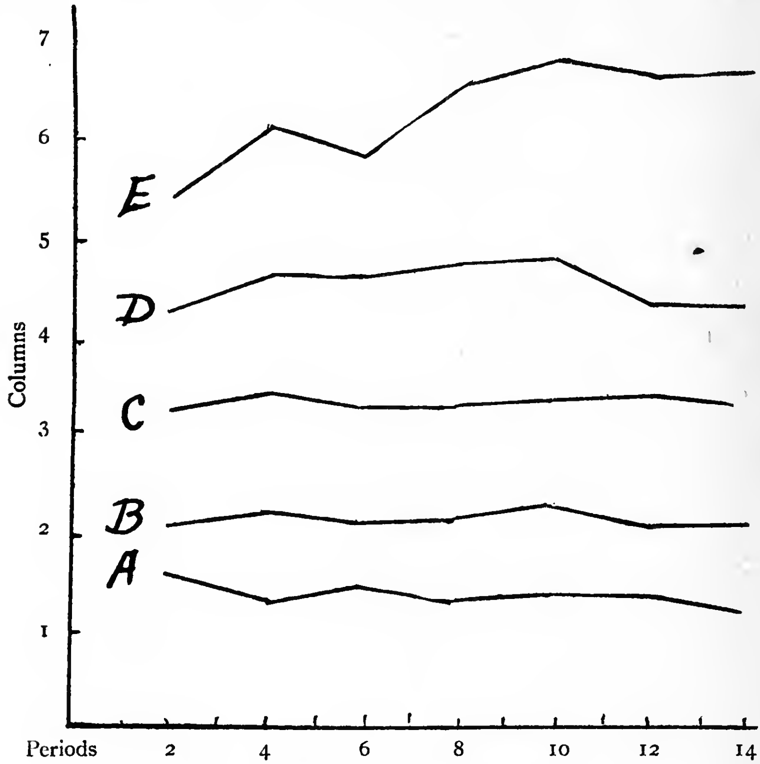


Fig. 9. Attempts—Third and Fourth Grades

EXPLANATION OF FIGURES 9–12

The points on these curves are for the average performances for every four minutes, the data being handled thus so as to smooth out the curves.

opposite course. The curves for the intervening tenths are rather confusing, but apparently have no definite tendency to deviate much from the horizontal—which would indicate that, on the whole, the medium workers neither increased nor decreased their speed.

[1] See Figures 5, 6, 7, and 8 on pages 31, 32, 33 and 34

The curves for accurates fail to reveal to the eye any clear difference between the sub-groups.

Evidently we have made too many sub-groups in dividing into tenths. While highest and lowest tenths should evidently be kept separate, as their curves stand so definitely apart from the rest, the second and third tenths may very well be combined, and likewise the eighth and ninth, leaving the four middle tenths to com-

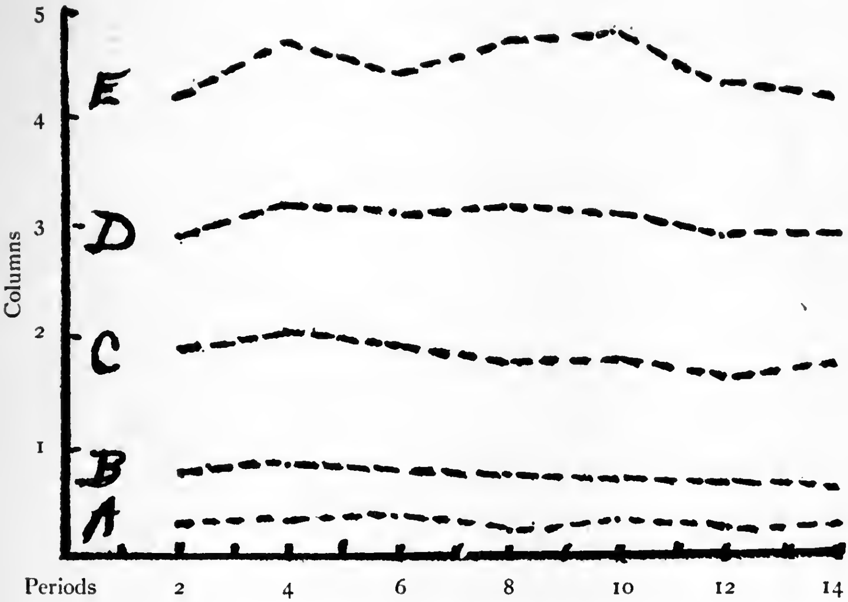


Fig. 10. Accurates—Third and Fourth Grades

bine into one central group. The resulting division into five unequal parts gives approximately equal spaces between the averages of the adjacent groups, as should be the case according to the theory of distribution, and as appears in fact to be the case in the following curves constructed as above indicated. In order further to eliminate irrelevant fluctuations, the points on the curves indicate the averages of two adjacent periods of two minutes. The Tables, XVIII to XXI, give the averages for each period. Group A, in these tables, consists of the lowest tenth, Group B of the second and third tenths, Group C of the fourth to seventh tenths, Group D of the eighth and ninth, and Group E of the uppermost tenth.

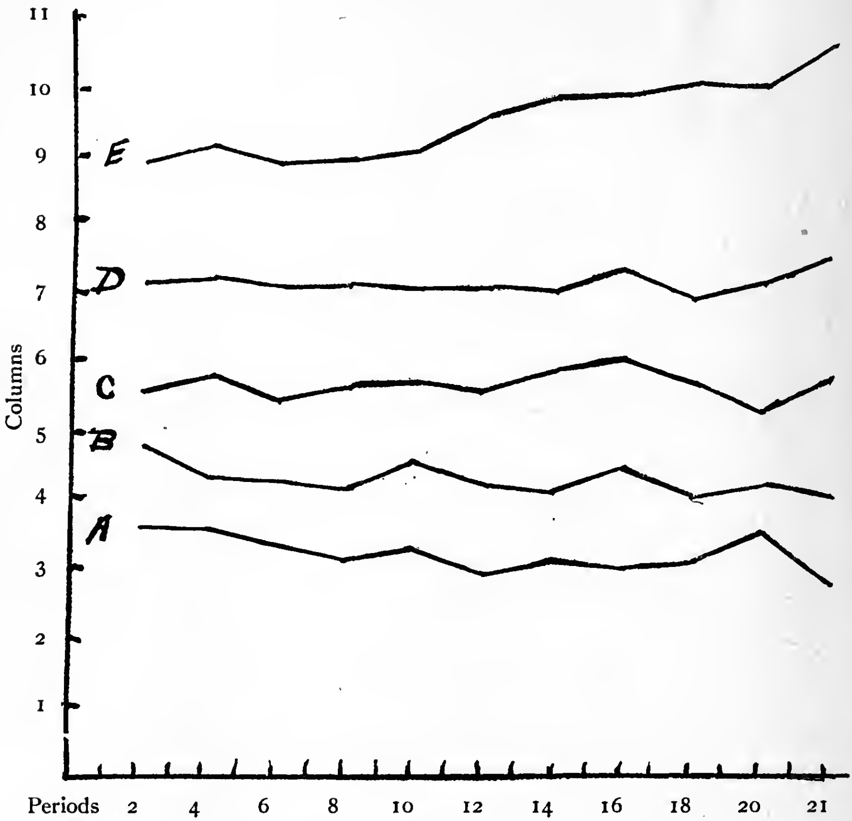


Fig. 11. Attempts—Seventh and Eighth Grades

The curves, thus reduced to their lowest terms, indicate pretty clearly that there is a relation in attempts between speed of work and liability to fatigue. The most rapid workers increase their speed as time goes by, while the slowest workers fall off, the middle groups remaining, on the average, almost perfectly steady for the time here considered. In accurate work, the general tendency of the whole group was shown in the last chapter to be slightly downward towards the end. But the decline comes later with the upper group, and is somewhat less pronounced than in the lowermost group, at least when the decline is considered in relation to the height of the curve above the base line. This is brought out by another way of examining the results.

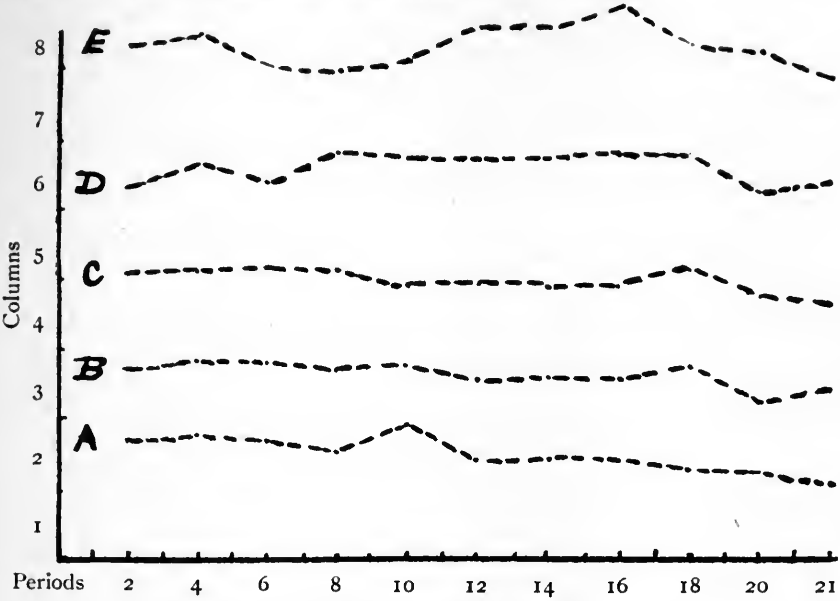


Fig. 12. Accurates—Seventh and Eighth Grades

As heretofore, the first three periods were compared with the last three periods. The results are given below:

Third and Fourth Grades

Attempts	*Accurates*
Group A, 12 per cent. loss	Group A, 16 per cent. loss
Group B, 5 per cent. loss	Group B, 18 per cent. loss
Group C, 1 per cent. loss	Group C, 18 per cent. loss
Group D, 1 per cent. gain	Group D, 6 per cent. loss
Group E, 16 per cent. gain	Group E, 4 per cent. loss

Seventh and Eighth Grades

Attempts	*Accurates*
Group A, 5 per cent. loss	Group A, 25 per cent. loss
Group B, 8 per cent. loss	Group B, 12 per cent. loss
Group C, 1 per cent. loss	Group C, 8 per cent. loss
Group D, 2 per cent. gain	Group D, 2 per cent. loss
Group E, 16 per cent. gain	Group E, 3 per cent. loss

In the case of attempts, the relation between total speed and fatigue comes out in this tabulation as clearly as before. The loss shown by the slowest groups gives way to a gain with the quickest workers. In the case of accurates, the change is certainly less clear, and there is no passing over to a gain; yet it may fairly be said that the better groups show a smaller loss than the poorer.

The facts of Column IX in Tables XVI and XVII are taken for the Groups A, B, C, D, and E of the previous tabulation. The facts for the lowest tenth are the facts for Group A, the second and third tenth are those for Group B, etc.

Third and Fourth Grades

Attempts Group	1	2	3	*Accurates Group*	1	2	3
A	1.00	.95	.92	A	1.00	1.12	.84
B	1.00	.985	.975	B	1.00	.98	.84
C	1.00	.99	.99	C	1.00	.92	.87
D	1.00	1.035	1.015	D	1.00	1.01	.95
E	1.00	1.09	1.13	E	1.00	1.04	.98

Seventh and Eighth Grades

Attempts Group	1	2	3	*Accurates Group*	1	2	3
A	1.00	.91	.93	A	1.00	.95	.80
B	1.00	.96	.985	B	1.00	1.00	.91
C	1.00	1.028	1.02	C	1.00	.96	.955
D	1.00	1.005	1.03	D	1.00	1.01	1.005
E	1.00	1.03	1.12	E	1.00	1.02	1.02

In the case of attempts, the relation between total speed and liability to fatigue is about as clear in this present tabulation as it was in the previous one. The quickest workers gain and the poorest workers lose. In the case of accuracy the change is clearer, for in the younger classes the better workers show a smaller loss than the poorer ones and in the older classes the best workers show a small gain. Consequently this way of examining the results supports the conclusion that on the whole the slower workers fatigue more quickly than the more rapid workers.

This, then, is put forward as one of the chief results of the experiment: that, on the whole, *the slower workers fatigue more quickly than the more rapid workers.*

TABLE XVIII

Attempts, Third and Fourth Grades

	Group I	Groups II, III	Groups IV, V, VI, VII,	Groups VIII, IX	Group X
	A	B	C	D	E
1	1.52	2.13	3.04	4.25	5.16
2	1.50	2.13	3.46	4.31	5.75
3	1.27	2.23	3.52	4.41	6.05
4	1.19	2.12	3.43	4.75	6.25
5	1.44	2.08	3.17	4.39	5.78
6	1.36	2.17	3.43	4.65	5.92
7	1.22	2.17	3.21	4.47	6.61
8	1.33	2.15	3.34	4.82	6.58
9	1.22	2.19	3.34	4.72	6.92
10	1.36	2.17	3.44	4.82	6.75
11	1.22	2.21	3.66	4.66	6.36
12	1.33	1.95	3.19	4.29	6.52
13	1.38	2.11	3.43	4.48	6.41
14	1.08	2.12	3.28	4.39	6.70

TABLE XIX

Accurates, Third and Fourth Grades

	Group I	Groups II, III	Groups IV, V, VI, VII	Groups VIII, IX	Group X
	A	B	C	D	E
1	.13	.51	1.65	2.61	3.84
2	.35	1.00	1.94	3.18	4.50
3	.27	.97	2.02	3.24	4.78
4	.24	.76	1.99	3.30	4.41
5	.29	.86	1.64	3.03	4.19
6	.32	.81	1.98	3.05	4.67
7	.21	.77	1.68	2.90	4.47
8	.16	.86	1.78	3.45	4.73
9	.40	.73	1.69	3.14	4.89
10	.21	.77	1.81	3.00	4.55
11	.21	.60	1.80	3.20	4.33
12	.21	.74	1.40	2.68	4.25
13	.24	.63	1.74	3.09	4.08
14	.18	.65	1.55	2.78	4.30

TABLE XX

Attempts—Seventh and Eighth Grades

	Group I	*Groups II, III*	*Groups IV, V, VI, VII*	*Groups VIII, IX*	*Group X*
	A	*B*	*C*	*D*	*E*
1	3.71	4.77	5.66	7.56	9.29
2	3.42	4.79	5.47	6.75	8.61
3	3.37	4.43	5.64	7.18	8.86
4	3.66	4.41	5.72	7.22	9.44
5	3.17	4.38	5.24	6.77	8.93
6	3.55	4.44	5.75	7.45	8.89
7	3.17	4.08	5.63	6.93	8.98
8	3.08	4.39	5.68	7.40	8.95
9	3.20	4.41	5.55	6.98	8.83
10	3.40	4.65	5.86	7.22	9.38
11	3.15	4.50	5.74	7.28	9.77
12	2.83	4.11	5.60	6.92	9.52
13	3.22	4.59	5.93	7.18	9.77
14	3.09	3.99	5.93	6.92	9.95
15	2.87	4.48	6.25	7.91	10.11
16	3.25	4.52	5.81	6.98	9.68
17	3.31	4.68	5.85	6.94	9.98
18	3.06	4.38	5.54	7.03	10.44
19	3.57	4.09	5.11	6.95	10.11
20	3.57	4.57	5.73	7.48	10.14
21	2.86	4.22	5.70	7.53	10.79

TABLE XXI

Accurates, Seventh and Eighth Grades

	Group I	*Groups II, III*	*Groups IV, V, VI, VII*	*Groups VIII, IX*	*Group X*
	A	*B*	*C*	*D*	*E*
1	1.43	2.72	4.02	5.62	7.59
2	1.86	2.66	4.01	5.19	7.13
3	1.77	2.73	3.89	5.50	7.23
4	1.80	2.90	4.26	5.68	7.80
5	1.49	2.82	4.19	5.21	7.14
6	1.89	2.77	4.08	5.66	6.89
7	1.67	2.54	3.92	5.56	7.08
8	1.52	2.79	4.24	5.94	6.92
9	1.80	2.57	3.99	5.50	6.98
10	2.17	2.96	3.95	5.74	7.42
11	1.83	2.80	4.26	5.47	8.02
12	1.02	2.37	3.70	5.39	7.35
13	1.54	2.79	4.14	5.52	7.95
14	1.45	2.34	3.68	5.37	7.32
15	1.40	2.59	4.05	5.79	8.32
16	1.58	2.48	3.85	5.36	7.62
17	1.48	2.68	4.07	5.90	7.38
18	1.22	2.58	4.02	5.32	7.32
19	1.37	2.23	3.65	5.30	7.32
20	1.31	2.38	3.77	5.33	7.17
21	1.14	2.49	3.58	5.39	6.83

CHAPTER IV

VARIABILITY OF THE INDIVIDUAL

Dependence of Variability upon Central Tendency

Everyone feels that variability is bound to increase as the measure of performance increases. There is evidently some dependence of variability upon the absolute measure of performance, such that the greater the absolute measure of performance the greater the variability.

Pearson has accepted the common-sense idea regarding this dependence, *i.e.*, that variability is normally or typically proportional to the absolute measure of performance. Fullerton and Cattell, and Thorndike, following the rule governing the variable error in the theory of measurement, have held that the variability normally increases as the square root of the absolute measure of performance, though probably with modifications in different kinds of performance.

Our data afford the opportunity of examining the variability of each individual from one period to another of the fourteen or twenty-one two-minute periods of the test, and then comparing the variability of individuals with smaller and greater total (or average) performance. This would be little to the purpose if a pronounced practice or fatigue effect occurred within the limits of the work; but, as these effects have been shown to be rather slight, the variability that appears in the individual's performance can be accepted as, in the main, simply the ordinary variability of performance.

TABLE XXII

Accurates of Third and Fourth Grades, showing average performance and variability of each individual for the twenty-eight minutes

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
0			
0			
1	.07	.13	
1	.07	.13	
1	.07	.13	
1	.07	.13	
1	.07	.13	.13
2	.14	.25	
2	.14	.25	
2	.14	.25	
2	.14	.25	
2	.14	.25	.25
3	.21	.34	
3	.21	.34	
3	.21	.34	
3	.21	.34	
3	.21	.34	.34
4	.29	.42	
4	.29	.42	
4	.29	.42	
4	.29	.46	
4	.29	.46	
4	.29	.46	.44
5	.36	.73	
5	.36	.73	
5	.36	.66	
5	.36	.66	
5	.36	.66	
5	.36	.66	
5	.36	.66	.68
6	.43	.63	
6	.43	.57	
6	.43	.57	
6	.43	.63	
6	.43	.57	
6	.43	.50	
6	.43	.63	
6	.43	.57	.58
7	.50	.64	
7	.50	.61	
7	.50	.61	

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
7	.50	.54	
7	.50	.54	
7	.50	.61	
7	.50	.61	
7	.50	.74	
7	.50	.54	.60
8	.57	.60	
8	.57	.52	
8	.57	.67	
8	.57	.60	
8	.57	.60	
8	.57	.60	.60
9	.64	.57	
9	.64	.73	
9	.64	.64	
9	.64	.42	
9	.64	.53	
9	.64	.53	
9	.64	.65	
9	.64	.53	.58
10	.71	.52	
10	.71	.46	
10	.71	.63	
10	.71	.75	
10	.71	.82	
10	.71	.75	
10	.71	.75	
10	.71	.68	
10	.71	.72	
10	.71	1.04	
10	.71	.72	
10	.71	.72	
10	.71	.81	.72
11	.79	.67	
11	.79	.69	
11	.79	.97	
11	.79	.81	
11	.79	.91	.81
12	.86	.48	
12	.86	.74	
12	.86	.34	
12	.86	.63	
12	.86	.74	
12	.86	.61	

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
12	.86	.52	
12	.86	.52	
12	.86	1.52	
12	.86	.73	
12	.86	.73	.69
13	.93	.80	
13	.93	.53	
13	.93	.68	
13	.93	.70	
13	.93	.80	
13	.93	.70	
13	.93	.57	
13	.93	.68	
13	.93	.52	
13	.93	1.05	
13	.93	.70	.70
14	1.00	.44	
14	1.00	.73	
14	1.00	.29	
14	1.00	.73	
14	1.00	.58	
14	1.00	.87	
14	1.00	.44	
14	1.00	.73	
14	1.00	.86	
14	1.00	.58	
14	1.00	.57	
14	1.00	.71	.63
15	1.07	.92	
15	1.07	.63	
15	1.07	.34	
15	1.07	.53	
15	1.07	.53	
15	1.07	.66	
15	1.07	.73	
15	1.07	.66	
15	1.07	.63	
15	1.07	.63	
15	1.07	.80	.64
16	1.14	.62	
16	1.14	.64	
16	1.14	.60	
16	1.14	.75	.65
17	1.21	.56	.56

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
18	1.29	.81	
18	1.29	.61	
18	1.29	.72	
18	1.29	.64	
18	1.29	.61	
18	1.29	.75	
18	1.29	.94	.72
19	1.36	.73	
19	1.36	.57	
19	1.36	.69	
19	1.36	.98	
19	1.36	1.11	
19	1.36	.73	
19	1.36	.54	
19	1.36	.74	
19	1.36	.83	
19	1.36	.78	.77
20	1.43	.64	
20	1.43	.69	
20	1.43	.72	
20	1.43	.77	
20	1.43	.63	
20	1.43	.63	
20	1.43	.72	
20	1.43	.63	.68
21	1.50	.86	
21	1.50	.71	
21	1.50	.57	
21	1.50	.50	
21	1.50	1.00	
21	1.50	.78	
21	1.50	.64	.72
22	1.57	.55	
22	1.57	.55	
22	1.57	.83	
22	1.57	.57	
22	1.57	.91	
22	1.57	.57	
22	1.57	.97	.70
23	1.64	.82	
23	1.64	.69	
23	1.64	1.21	
23	1.64	.59	
23	1.64	.44	

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
23	1.64	82	
23	1.64	1.08	.80
24	1.71	.75	
24	1.71	.75	
24	1.71	.75	
24	1.71	.75	
24	1.71	1.02	
24	1.71	.72	
24	1.71	.65	
24	1.71	.55	
24	1.71	.89	
24	1.71	1.02	
24	1.71	.75	
24	1.71	.90	.79
25	1.79	1.09	
25	1.79	.72	
25	1.79	1.07	
25	1.79	1.09	
25	1.79	1.50	
25	1.79	.72	
25	1.79	.72	
25	1.79	.93	
25	1.79	.44	.92
26	1.86	.87	
26	1.86	.63	
26	1.86	.87	
26	1.86	.89	
26	1.86	1.28	
26	1.86	.87	
26	1.86	.87	
26	1.86	1.00	.91
27	1.93	.66	
27	1.93	.93	
27	1.93	.66	
27	1.93	.95	
27	1.93	.66	.77
28	2.00	.84	
28	2.00	.58	
28	2.00	.70	
28	2.00	.70	
28	2.00	.70	
28	2.00	1.06	
28	2.00	.70	
29	2.07	.93	

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
29	2.07	.66	
29	2.07	1.08	
29	2.07	.79	
29	2.07	.66	.82
30	2.14	.86	
30	2.14	.62	
30	2.14	.98	
30	2.14	.86	
30	2.14	.62	
30	2.14	.86	.80
31	2.21	.80	
31	2.21	.73	
31	2.21	.84	
31	2.21	1.04	
31	2.21	.95	
31	2.21	.49	.81
32	2.28	1.03	
32	2.28	.59	
32	2.28	.93	
32	2.28	.50	
32	2.28	.75	
32	2.28	1.23	.84
33	2.36	1.16	
33	2.36	.78	
33	2.36	.68	
33	2.36	.78	
33	2.36	.83	
33	2.36	.68	
33	2.36	.78	
33	2.36	.83	
33	2.36	.68	
33	2.36	.87	
33	2.36	.78	
33	2.36	.68	.80
34	2.43	1.04	
34	2.43	.85	
34	2.43	.78	
34	2.43	.90	.89
35	2.50	.93	
35	2.50	.85	
35	2.50	1.13	
35	2.50	.63	
35	2.50	1.00	
35	2.50	1.05	.94

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
36	2.58	.70	
36	2.58	.90	
36	2.58	1.19	
36	2.58	1.29	
36	2.58	.70	.95
37	2.64	1.24	
37	2.64	1.06	
37	2.64	1.16	
37	2.64	.69	
37	2.64	.98	
37	2.64	1.00	
37	2.64	.82	.96
37	2.64	.69	
38	2.71	.76	
38	2.71	1.43	
38	2.71	1.17	
38	2.71	.95	
38	2.71	.79	1.02
39	2.79	.70	
39	2.79	.73	
39	2.79	.84	
39	2.79	.81	
39	2.79	1.26	.87
40	2.86	.60	
40	2.86	.92	
40	2.86	.74	
40	2.86	.78	.76
41	2.93	.82	
41	2.93	.64	
41	2.93	.94	
41	2.93	1.20	
41	2.93	.94	.91
42	3.00	1.20	
42	3.00	.94	
42	3.00	.57	
42	3.00	.66	
42	3.00	.54	
42	3.00	.66	
42	3.00	.81	
42	3.00	1.08	.73
43	3.07	1.14	
43	3.07	.92	
43	3.07	.77	
43	3.07	.80	

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
43	3.07	.52	
43	3.07	.37	.75
44	3.14	.66	
44	3.14	1.16	
44	3.14	1.16	.99
45	3.21	.83	
45	3.21	1.06	
45	3.21	.93	
45	3.21	1.09	.98
46	3.28	.98	
46	3.28	.79	
46	3.28	.89	
46	3.28	.98	
46	3.28	.89	.91
47	3.36	.91	
47	3.36	.74	
47	3.36	.81	
47	3.36	1.01	
47	3.36	.91	
47	3.36	.67	.84
48	3.43	1.20	
48	3.43	.93	
48	3.43	.99	
48	3.43	1.13	
48	3.43	.99	
48	3.43	1.13	
48	3.43	.99	1.05
49	3.50	.63	
49	3.50	.84	
49	3.50	.77	
49	3.50	.84	.77
50	3.57	.71	.71
51	3.64	.76	
51	3.64	.61	
51	3.64	.80	.72
52	3.71	1.04	
52	3.71	.74	.89
54	3.86	1.00	
54	3.86	1.15	
54	3.86	.73	.96
55	3.93	1.06	
55	3.93	.79	
55	3.93	.94	
55	3.93	.94	.94

TABLE XXII (*Continued*)

Total	*Average*	*A. D.*	*C. T. of A. D.'s.*
56	4.00	.56	.56
57	4.07	.77	
57	4.07	.77	.77
58	4.14	.75	
58	4.14	1.41	
58	4.14	1.16	1.11
59	4.21	.93	
59	4.21	1.18	1.06
60	4.28	.47	
60	4.28	.68	.57
61	4.36	1.09	1.09
62	4.43	1.51	
62	4.43	1.02	1.26
64	4.57	.96	
64	4.57	.96	.96
65	4.64	.74	.74
66	4.71	1.04	1.04
68	4.86	.87	.87
70	5.00	.70	.70
71	5.07	.66	.66
72	5.14	.87	.87
73	5.21	.83	.83
76	5.43	1.03	1.03
85	6.07	1.21	1.21
110	7.85	1.02	1.02

TABLE XXIII

Showing the Data for Third and Fourth Grades, Accurates with C. T.'s and Average A. D.'s Combined into Twelve Groups

C. T.	*Average A. D.*
.07 to .43	.40
.50 to .86	.67
.93 to 1.29	.65
1.36 to 1.71	.75
1.79 to 2.14	.13
2.21 to 2.58	.87
2.67 to 3.00	.88
3.07 to 3.43	.92
3.50 to 3.86	.81
3.93 to 4.28	.84
4.36 to 5.21	.90
5.43 to 7.85	1.09

The twelve groups given here were obtained by combining the data of Table XXII so that as nearly as possible the twelve groups would proceed by steps of .36. The data would not submit to this treatment in the last two groups of the twelve.

TABLE XXIV

Attempts, Third and Fourth Grades, Showing Average Performance and Average Variability of Certain Groups of Individuals for the Twenty-eight Minutes. Eighty-five Groups

Number of Columns		*Number of Columns*		*Number of Columns*	
C. T.	*Average of A. D.*	*C. T.*	*Average of A. D.*	*C. T.*	*Average of A. D.*
.28	.55	2.36	.62	4.07	.49
.43	.49	2.43	.52	4.14	.81
.64	.55	2.50	.67	4.21	.60
.78	.34	2.57	.53	4.28	.70
.86	.24	2.64	.41	4.35	.79
1.00	.10	2.71	.67	4.43	.67
1.07	.13	2.78	.54	4.50	.68
1.14	.29	2.86	.59	4.57	.62
1.21	.46	2.93	.41	4.64	.71
1.28	.41	3.00	.42	4.71	.89
1.35	.60	3.07	.68	4.78	1.18
1.43	.56	3.14	.53	4.85	.55
1.50	.58	3.21	.49	4.93	.42
1.57	.57	3.28	.67	5.00	.33
1.64	.47	3.35	.55	5.07	.95
1.71	.58	3.43	.66	5.14	1.08
1.78	.54	3.50	.66	5.21	.42
1.83	.47	3.57	.64	5.28	.60
1.93	.37	3.64	.56	5.36	.62
2.00	.46	3.71	.71	5.43	.65
2.07	.37	3.78	.60	5.50	.64
2.21	.55	3.92	.49	5.64	.53
2.28	.56	4.00	.23	5.78	.58
				5.85	.51
				5.93	.82
				6.00	.57
				6.07	.67
				6.28	.79
				6.43	.72
				6.50	2.14
				6.93	.53
				7.00	.43
				7.43	.92
				8.71	.78
				9.79	6.21
				13.64	3.72

TABLE XXV

Showing Data for the Foregoing Table Combined into Twelve Groups

C. T.	Average of A. D.
.28 to 1.71	.46
1.78 to 2.14	.45
2.25 to 2.57	.58
2.64 to 3.00	.51
3.07 to 3.43	.59
3.50 to 3.86	.63
3.92 to 4.28	.55
4.35 to 4.71	.73
4.78 to 5.28	.69
5.36 to 5.70	.67
5.78 to 6.93	.81
7.00 to 13.64	2.41

In this table we proceed as nearly by steps of .36 as possible. The first group and last four groups could not well be made to fulfill this condition. But as this does not affect the mathematical significance of the results, it matters little.

TABLE XXVI

Accuracy, Seventh and Eighth Grades, Showing Average Performance and Average Variability of Certain Groups of Individuals for the Forty-two Minutes of Working. There are 122 Groups

Number of Columns C. T.	Average of A. D.	Number of Columns C. T.	Average of A. D.	Number of Columns C. T.	Average of A. D.
.43	.53	2.38	1.10	3.57	1.05
.48	.59	2.43	1.21	3.62	1.00
.62	.65	2.48	1.07	3.67	.90
.71	.82	2.52	.83	3.71	1.24
.76	.58	2.57	1.30	3.76	1.08
.90	.61	2.62	1.01	3.81	1.15
1.04	.93	2.67	1.04	3.86	1.23
1.29	.92	2.71	.99	3.90	.65
1.33	.51	2.76	.82	3.95	1.11
1.38	.89	2.81	1.05	4.00	1.19
1.43	1.37	2.86	.84	4.05	1.00
1.62	.99	2.90	.90	4.10	1.09
1.81	.77	2.95	.60	4.14	1.09
1.86	1.62	3.00	.76	4.19	.98
1.90	1.03	3.05	1.00	4.24	1.27
1.95	.73	3.09	.97		

TABLE XXVI (*Continued*)

Number of Columns		*Number of Columns*		*Number of Columns*	
C. T.	*Average of A. D.*	*C. T.*	*Average of A. D.*	*C. T.*	*Average of A. D.*
		3.14	1.02	4.29	1.38
2.00	.48	3.19	.71	4.38	1.07
2.05	.78	3.24	.97	4.43	1.12
2.09	.92	3.29	1.04	4.48	.90
2.14	.89	3.33	.83	4.52	1.21
2.19	.95	3.38	1.11	4.57	1.08
2.24	.83	3.43	1.09	4.62	1.15
2.29	1.15	3.48	.89	4.67	1.11
2.33	1.01	3.52	1.21	4.71	1.49
4.76	1.16	5.33	1.02	6.09	1.03
4.81	1.11	5.43	1.81	6.14	1.39
4.86	1.84	5.48	1.26	6.19	.99
4.90	1.61	5.52	1.46	6.28	1.08
4.95	.97	5.57	1.54	6.43	1.30
5.00	.67	5.62	1.10	6.48	1.22
5.05	1.24	5.71	1.01	6.57	.95
5.09	1.14	5.76	1.14	6.62	1.14
5.14	1.14	5.86	.86	6.76	1.32
5.19	1.16	5.90	.87	6.90	1.28
5.24	1.34	5.95	.46	6.95	1.17
5.29	.98	6.00	.86	7.05	.78
				7.14	1.44
				7.19	1.38
				7.48	1.41
				7.57	1.31
				7.67	1.85
				7.71	1.35
				7.80	1.35
				8.00	1.33
				8.14	1.04
				8.52	.98
				8.71	1.59
				9.33	1.49
				11.52	1.65

TABLE XXVII

Showing Data of Foregoing Table Combined into Twelve Groups

C. T.	*Average of A. D.*
.43 to 1.90	.85
1.95 to 2.38	.88
2.43 to 2.86	1.02
2.90 to 3.33	.88
3.38 to 3.81	1.07
3.86 to 4.29	1.10
4.38 to 4.76	1.14
4.81 to 5.24	1.15
5.29 to 5.71	1.28
5.76 to 6.19	.95
6.28 to 6.62	1.14
6.76 to 11.52	1.34

The data of Table XXVI could not be easily handled in a graphic representation, consequently it was combined into twelve groups as in Table XXIII. In this particular table we proceed as nearly by steps of .43 of a column as possible. The first and last groups of the series could not be well made to go by the step.

TABLE XXVIII

Showing Central Tendencies and Corresponding Average A. D. of Columns Attempted by Groups of Individuals of Seventh and Eighth Grades. There are 127 Groups

Number of Columns		*Number of Columns*		*Number of Columns*	
C. T.	*Average of A. D.*	*C. T.*	*Average of A. D.*	*C. T.*	*Average of A. D.*
2.14	.70	5.29	.65	8.00	2.22
2.43	.68	5.33	.62	8.05	1.13
2.76	.58	5.38	.94	8.09	.79
2.81	.61	5.43	.79	8.14	.41
2.90	.61	5.48	1.04	8.24	.67
2.95	1.17	5.52	.77	8.62	1.13
3.00	.29	5.57	.82	8.71	.74
3.05	.54	5.62	.73	8.81	.92
3.09	.53	5.67	.65	8.86	1.12
3.19	1.15	5.71	.92	9.09	.88
3.24	.67	5.76	1.05	9.29	.64
3.29	.62	5.81	.57	9.33	1.27
3.33	.51	5.86	.81	9.38	.93
3.38	.69	5.90	.79	9.52	1.02
3.43	.53	5.93	.66	9.57	.88

TABLE XXVIII (*Continued*)

Number of Columns C. T.	*Average of A. D.*	*Number of Columns* C. T.	*Average of A. D.*	*Number of Columns* C. T.	*Average of A. D.*
3.52	.69	6.00	.72	9.67	.92
3.57	.79	6.05	.83	9.71	.84
3.67	.64	6.09	.64	10.24	1.03
3.71	.90	6.14	.75	10.57	.89
3.76	1.20	6.19	.72	10.62	.91
3.81	.62	6.24	1.14	10.71	.99
3.90	.88	6.29	.91	10.76	.89
3.95	1.14	6.33	1.24	11.19	.81
4.00	.76	6.38	.81	11.43	1.29
4.05	.76	6.43	.83	11.62	2.77
4.09	.67	6.52	.83	11.86	5.14
4.14	.92	6.57	.69	13.09	1.63
4.19	.65	6.62	.95		
4.24	.65	6.67	.70		
4.29	.72	6.71	.72		
4.33	.68	6.76	.65		
4.38	.72	6.86	.61		
4.43	.20	6.90	.63		
4.48	.51	6.95	.71		
4.52	1.10	7.05	.89		
4.57	.59	7.09	.98		
4.62	.69	7.14	.30		
4.67	.74	7.19	.61		
4.71	.57	7.24	.56		
4.76	.55	7.29	.92		
4.81	.99	7.33	.96		
4.86	.68	7.38	.64		
4.90	.45	7.43	.90		
4.95	.78	7.48	1.69		
5.00	.38	7.52	.71		
5.05	.70	7.57	.68		
5.09	.68	7.62	.87		
5.19	.74	7.71	.97		
5.24	1.26	7.76	.88		
		7.82	.83		
		7.95	.88		

The data of Table XXVIII could not be readily handled graphically, and were consequently combined into twelve groups as in Table XXV. In this particular table we proceed as nearly by steps of .7 of a column as possible, but it is difficult to do this because at the beginning and ending of the series—and in the middle—the numbers were so very much scattered. The best grouping was made that could be made. This does not impair the mathematical value of the graph.

TABLE XXIX

Showing Data of Foregoing Table, Combined into Twelve Groups

C. T.	Average of A. D.
2.14 to 2.90	.64
2.95 to 4.05	.75
4.09 to 4.76	.66
4.81 to 5.48	.76
5.48 to 6.19	.77
6.24 to 6.90	.82
6.95 to 7.62	.81
7.71 to 8.24	.98
8.62 to 8.81	.93
8.86 to 9.71	.94
10.24 to 10.76	.94
11.19 to 13.09	2.34

The actual facts for the present test are seen on comparing the average variability with the absolute measure of performance. (See the accompanying tables—XXII–XXVIII—which give, for the individuals having the same central tendency, the average of their A. D.'s.) Each of these tables is followed by a regrouping which brings out more fully the tendency of the A. D.'s to increase with the C. T.

In the absence of any universally accepted law, we can take these figures (Tables XXII, XXIV, XXVI, XXVIII) as our basis and see what law of dependence would fit them best. We have assumed that this law would have the form $\frac{Var.}{C.\,T.^{\frac{1}{n}}} = C.$ and find what value of the unknown n will give best agreement with this law. That is, we are to find such a root of the C. T. as will give the most constant quotient when divided into the corresponding variability.

Thus two methods were used for determining what value of n gave the closest approximation to a constant ratio:

1. Find for what value of n the variability of the quotient is least.

2. Observe for what value of n the curve approaches most closely to the horizontal line.

Tables XXX, XXXI, XXXII, XXXIII give the Average, Average Deviation, and Probable Error for each value of $\frac{x}{y^{\frac{1}{n}}}$ in attempts and accurate performance.

TABLE XXX

Showing for Accurates of Third and Fourth Grades the Central Tendency and Average Deviation of Different Values of $\frac{x}{y^{\frac{1}{n}}}$, *where* $x = A.\,D.$, $y = C.\,T.$

Values of $\frac{x}{y}$	Average	.489	A. D.	.302	P. E.	.254
Values of $\frac{x}{y^{\frac{1}{1.9}}}$	Average	.549	A. D.	.1227	P. E.	.104
Values of $\frac{x}{y^{\frac{1}{2}}}$	Average	.559	A. D.	.119	P. E.	.100
Values of $\frac{x}{y^{\frac{1}{2.2}}}$	Average	.569	A. D.	.1065	P. E.	.090
Values of $\frac{x}{y^{\frac{}{2.3}}}$	Average	.577	A. D.	.1011	P. E.	.085
Values of $\frac{x}{y^{\frac{1}{2.5}}}$	Average	.593	A. D.	.10069	P. E.	.084
Values of $\frac{x}{y^{\frac{1}{3}}}$	Average	.613	A. D.	.088	P. E.	.075
Values of $\frac{x}{y^{\frac{1}{4}}}$	Average	.646	A. D.	.090	P. E.	.076
Values of $\frac{x}{y^{\frac{1}{6}}}$	Average	.691	A. D.	.096	P. E.	.081
Values of $\frac{x}{y^{\frac{1}{8}}}$	Average	.7089	A. D.	.108	P. E.	.091
Values of $\frac{x}{y^{\frac{1}{10}}}$	Average	.7283	A. D.	.111	P. E.	.094
Values of x	Average	.793.	A. D.	.150	P. E.	.127

TABLE XXXI

Showing for Attempts of Third and Fourth Grades the Central Tendency and Average Deviation of Different Values of $\frac{x}{y^{\frac{1}{n}}}$

Values of $\frac{x}{y}$	Average	.241	A. D.	.130	P. E.	.012
Values of $\frac{x}{y^{\frac{1}{2}}}$	Average	.376	A. D.	.126	P. E.	.011
Values of $\frac{x}{y^{\frac{1}{3}}}$	Average	.453	A. D.	.130	P. E.	.011
Values of $\frac{x}{y^{\frac{1}{4}}}$	Average	.501	A. D.	.167	P. E.	.016
Values of $\frac{x}{y^{\frac{1}{5}}}$	Average	.560	A. D.	.192	P. E.	.016
Values of $\frac{x}{y^{\frac{1}{6}}}$	Average	.596	A. D.	.212	P. E.	.019
Values of $\frac{x}{y_{10}}$	Average	.615	A. D.	.225	P. E.	.017
Values of x	Average	.700	A. D.	.290	P. E.	.026

TABLE XXXII

Showing for Accurates of Seventh and Eighth Grades the Central Tendency and Average Deviation of Different Values of $\frac{x}{y^{\frac{1}{n}}}$

Values of $\frac{x}{y}$	Average	.329	A. D.	.1431	P. E.	.011
Values of $\frac{x}{y^{\frac{1}{2}}}$	Average	.558	A. D.	.1133	P. E.	.009
Values of $\frac{x}{y^{\frac{1}{3}}}$	Average	.684	A. D.	.1057	P. E.	.008
Values of $\frac{x}{y^{\frac{1}{4}}}$	Average	.766	A. D.	.1196	P. E.	.091
Values of x	Average	1.069	A. D.	.2079	P. E.	.054

TABLE XXXIII

Showing for Attempts of Seventh and Eighth Grades the Central Tendency and Average Deviation of Different Values of $\frac{x}{y^{\frac{1}{n}}}$

Values of $\frac{x}{y}$	Average	.148	A. D.	.048	P. E.	.004
Values of $\frac{x}{y^{\frac{1}{2}}}$	Average	.346	A. D.	.085	P. E.	.006
Values of $\frac{x}{y^{\frac{1}{3}}}$	Average	.467	A. D.	.107	P. E.	.008
Values of $\frac{x}{y^{\frac{1}{4}}}$	Average	.542	A. D.	.116	P. E.	.009
Values of x	Average	.86	A. D.	.304	P. E.	.0228

Explanation

Tables XXXIV, XXXV, XXXVI, and XXXVII give the different values of $\frac{x}{y^{\frac{1}{n}}}$ combined into twelve groups. (See Figures 13, 14, 15, and 16.)

Figures 13 to 16 show graphically the position of each value of $\frac{x}{y^{\frac{1}{n}}}$ on a vertical scale for increasing total performance on a horizontal scale.

TABLE XXXIV

Accuracy, Third and Fourth Grades, Values $\frac{x}{y^{\frac{1}{n}}}$

Total	Average x	Average $\frac{x}{y}$	Average $\frac{x}{y^{\frac{1}{1.9}}}$	Average $\frac{x}{y^{\frac{1}{2}}}$	Average $\frac{x}{y^{\frac{1}{2.2}}}$	Average $\frac{x}{y^{\frac{1}{2.3}}}$	Average $\frac{x}{y^{\frac{1}{2.5}}}$
1 to 6	.40	1.67	.82	.79	.74	.69	.72
7 to 12	.67	.99	.82	.85	.80	.81	.78
13 to 18	.65	.59	.62	.62	.62	.63	.61
19 to 24	.75	.48	.60	.60	.62	.63	.64
25 to 30	.83	.42	.59	.59	.61	.62	.64
31 to 36	.87	.36	.55	.56	.59	.60	.62
37 to 42	.88	.31	.51	.52	.55	.56	.60
43 to 48	.92	.28	.49	.51	.53	.55	.57
49 to 54	.81	.21	.34	.42	.45	.46	.49
55 to 60	.84	.20	.40	.41	.44	.46	.48
61 to 73	.90	.19	.40	.41	.44	.46	.48
76 to 110	1.09	.16	.41	.43	.47	.49	.52

TABLE XXXIV (*Continued*)

Total	*Average* $\frac{x}{y^{\frac{1}{3}}}$	*Average* $\frac{x}{y^{\frac{1}{4}}}$	*Average* $\frac{x}{y^{\frac{1}{6}}}$	*Average* $\frac{x}{y^{\frac{1}{8}}}$	*Average* $\frac{x}{y^{\frac{1}{10}}}$
1 to 6	.62	.67	.50	.47	.46
7 to 12	.76	.74	.71	.70	.69
13 to 18	.63	.64	.64	.64	.64
19 to 24	.65	.67	.70	.71	.72
25 to 30	.66	.70	.74	.75	.78
31 to 36	.65	.70	.76	.78	.80
37 to 42	.62	.66	.74	.77	.79
43 to 48	.62	.69	.75	.79	.81
49 to 54	.53	.59	.64	.69	.71
55 to 60	.51	.59	.66	.65	.72
61 to 73	.54	.61	.69	.74	.77
76 to 110	.59	.69	.79	.86	.90

TABLE XXXV

Attempts, Third and Fourth Grades, Values

Total Number of Columns	*Average* x	*Average* $\frac{x}{y}$	*Average* $\frac{x}{y^{\frac{1}{2}}}$	*Average* $\frac{x}{y^{\frac{1}{3}}}$	*Average* $\frac{x}{y^{\frac{1}{4}}}$	*Average* $\frac{x}{y^{\frac{1}{6}}}$	*Average* $\frac{x}{y^{\frac{1}{8}}}$	*Average* $\frac{x}{y^{\frac{1}{10}}}$
4 to 24	.46	.54	.48	.46	.46	.46	.46	.46
25 to 30	.45	.28	.32	.38	.38	.40	.41	.42
31 to 36	.58	.36	.43	.46	.49	.49	.49	.53
37 to 42	.51	.18	.30	.35	.39	.42	.45	.46
43 to 48	.59	.18	.33	.36	.41	.48	.51	.53
49 to 54	.63	.17	.32	.41	.45	.49	.54	.55
55 to 60	.55	.13	.27	.35	.39	.44	.48	.48
61 to 66	.73	.17	.32	.46	.49	.56	.60	.63
67 to 74	.69	.14	.31	.40	.46	.52	.57	.59
75 to 80	.67	.12	.28	.39	.44	.50	.54	.57
81 to 97	.81	.13	.32	.44	.51	.60	.62	.57
98 to 191	2.41	.23	.54	1.11	1.34	1.63	1.78	1.91

TABLE XXXVI

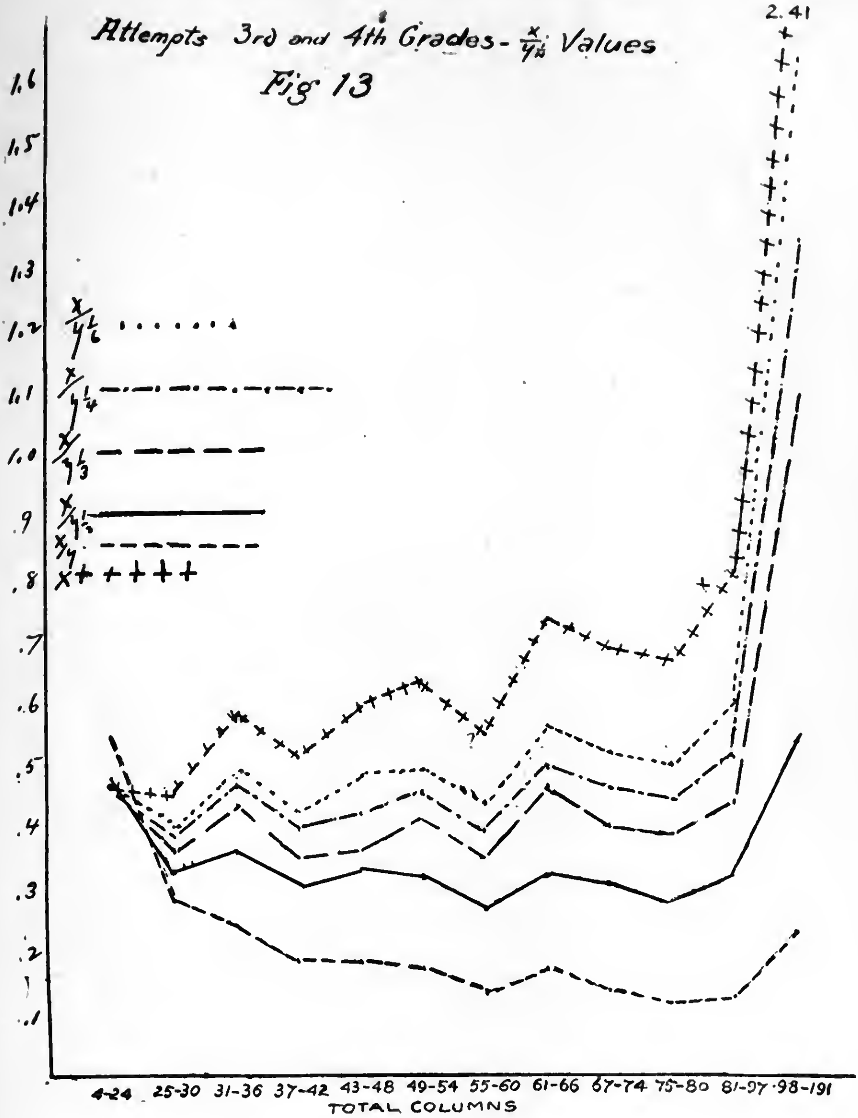
Accuracy, Seventh and Eighth Grades, Values of $\frac{x}{y^{\frac{1}{n}}}$

Total Columns	*Average* x	*Average* $\frac{x}{y}$	*Average* $\frac{x}{y^{\frac{1}{2}}}$	*Average* $\frac{x}{y^{\frac{1}{3}}}$	*Average* $\frac{x}{y^{\frac{1}{4}}}$
9–40	.85	.81	.80	.82	.83
41–50	.88	.41	.60	.67	.75
51–60	1.02	.39	.63	.74	.80
61–70	.88	.21	.52	.60	.64
71–80	1.07	.30	.58	.68	.79
81–90	1.10	.27	.54	.69	.77
91–100	1.14	.25	.53	.69	.78
101–110	1.15	.23	.52	.67	.77
111–120	1.28	.23	.55	.73	.86
121–130	.95	.17	.35	.52	.61
131–140	1.14	.18	.45	.61	.71
141–242	1.34	.17	.48	.67	.80

TABLE XXXVII

Attempts, Seventh and Eighth Grades, Values of $\frac{x}{y^{\frac{1}{n}}}$

Total Columns	*Average* x	*Average* $\frac{x}{y}$	*Average* $\frac{x}{y^{\frac{1}{2}}}$	*Average* $\frac{x}{y^{\frac{1}{3}}}$	*Average* $\frac{x}{y^{\frac{1}{4}}}$
45–61	.64	.249	.397	.464	.502
62–85	.75	.215	.403	.495	.542
86–100	.66	.153	.296	.405	.458
101–115	.76	.148	.335	.429	.507
111–130	.77	.131	.316	.424	.524
131–145	.82	.126	.329	.441	.515
146–160	.84	.143	.301	.415	.495
162–173	.98	.122	.346	.488	.581
181–185	.93	.107	.315	.452	.507
186–214	.94	.101	.312	.451	.542
215–221	.94	.089	.289	.437	.527
226–275	2.34	.196	.672	1.011	1.254

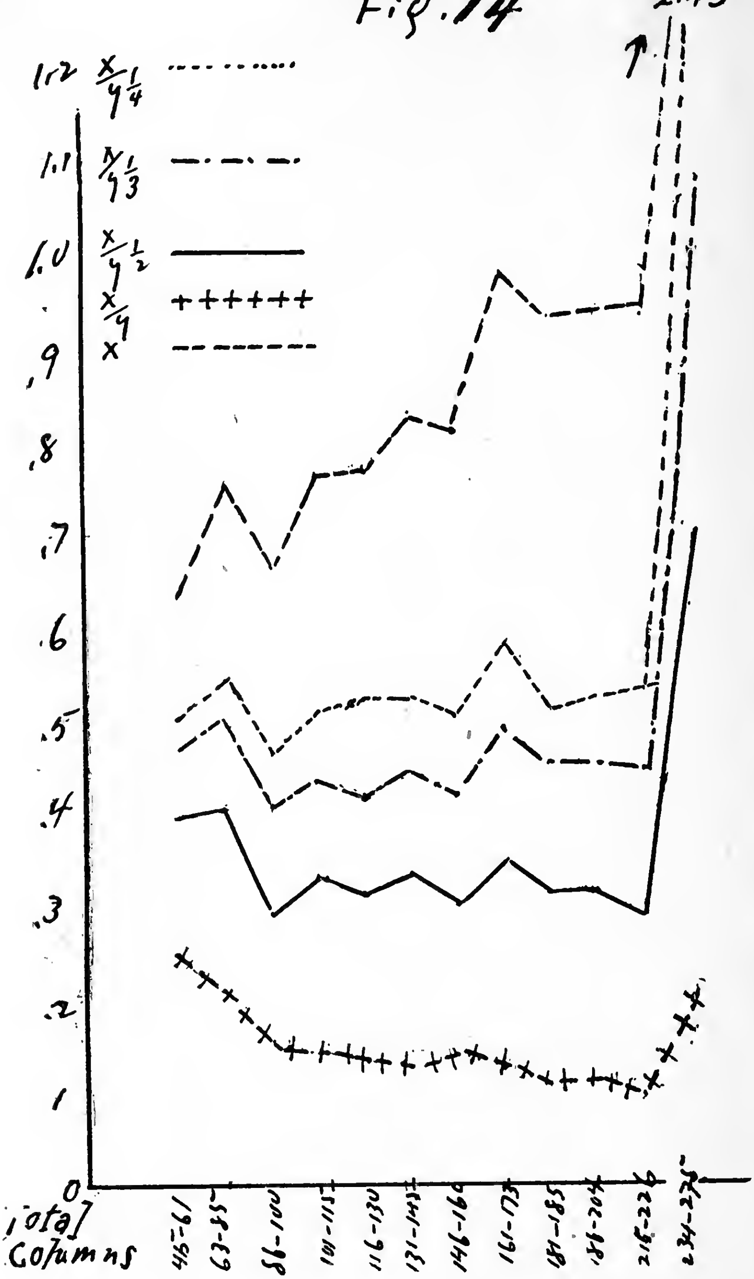


Corrected Variability in Attempts

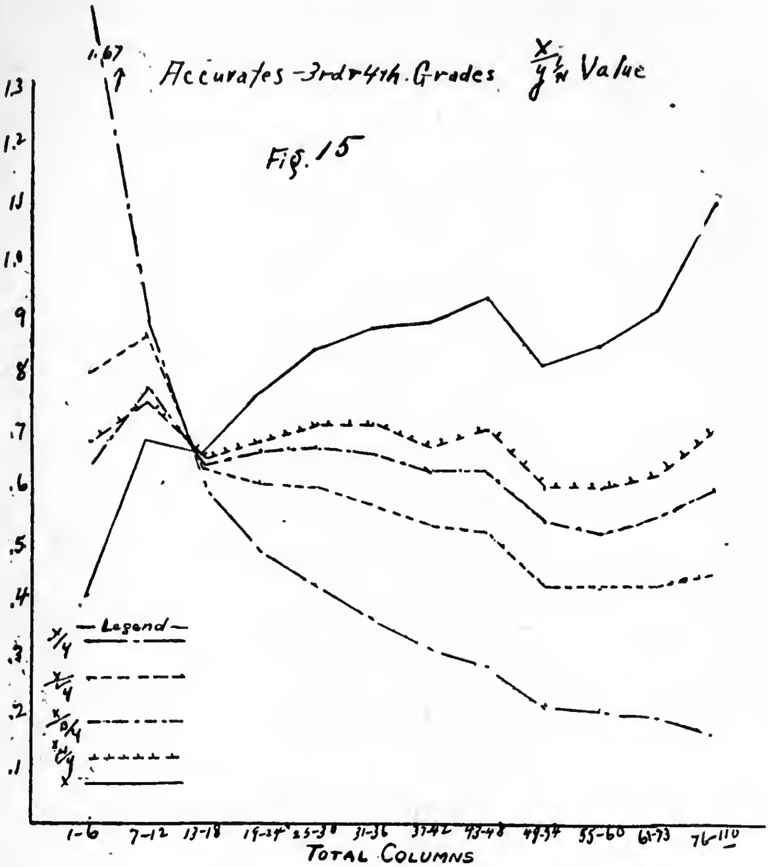
Figures 13 and 14 give the graphical representation for the attempts data of corrected variability for the younger and older groups respectively.

Attempts 7th. and 8th. Grades $\frac{x}{y^k}$ Values

Fig. 14

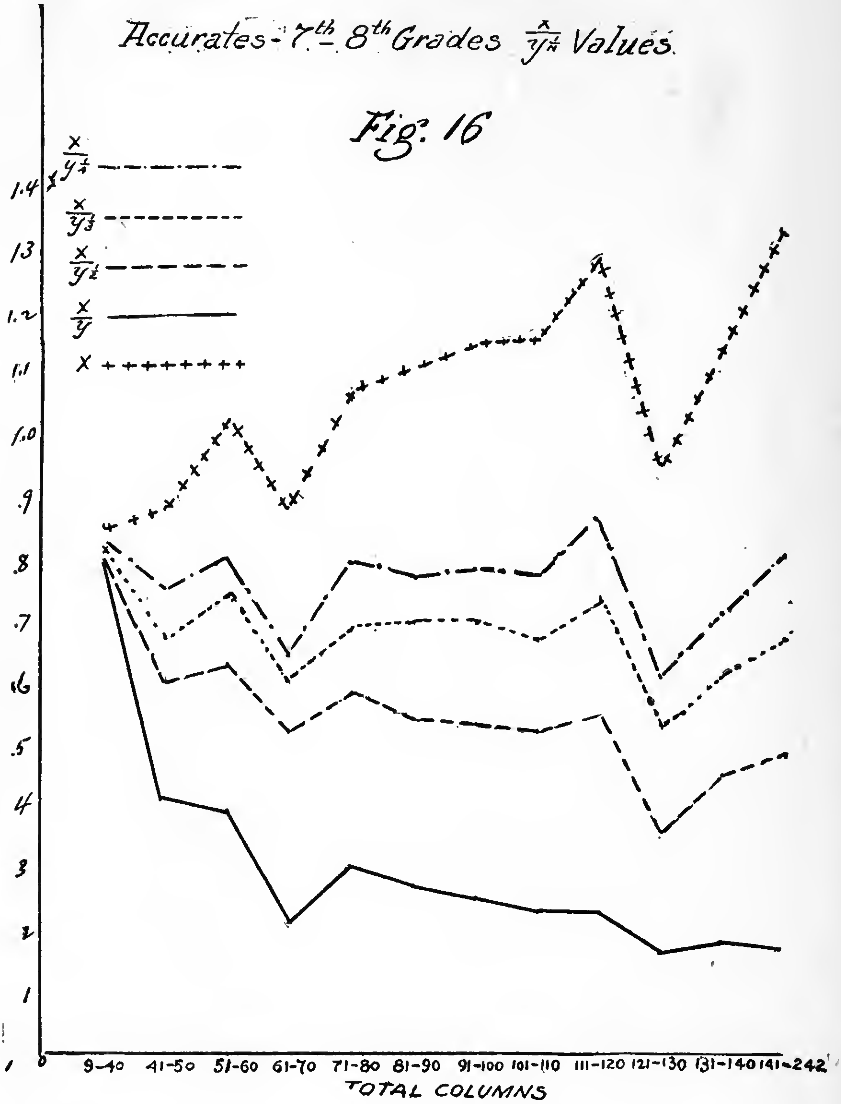


For the Third and Fourth grades attempts, Figure 13 shows the $\frac{x}{y^{1/2}}$, $\frac{x}{y^{1/3}}$ and $\frac{x}{y^{1/4}}$ lines to be nearly horizontal, but the preference goes to the $\frac{x}{y^{1/3}}$ value as it is the most nearly horizontal line of those lines



and besides the $\frac{x}{y^{1/3}}$ series has the smallest average deviation of all the different series of the $\frac{x}{y^{1/n}}$ values.

As to the attempts data of corrected variability for the Seventh and Eighth grades, the $\frac{x}{y}$ values and the $\frac{x}{y^{1/3}}$ values compete most



strongly for the preference. We have chosen here, however, the $\frac{x}{y}$ line, since it appears to be the most nearly horizontal one and the judgment seems justified by the fact that the average deviation of the $\frac{x}{y}$ series is smaller than that of the other series. But the experimenter thinks that the cause of this inconsistency is due to the fact that some of the subjects at the last attempted too recklessly and thus impaired the data to a small extent. If this had not happened the $y^{\frac{1}{3}}$ value would have been, the writer believes, the chosen value of $y^{\frac{1}{n}}$. (Figure 14.)

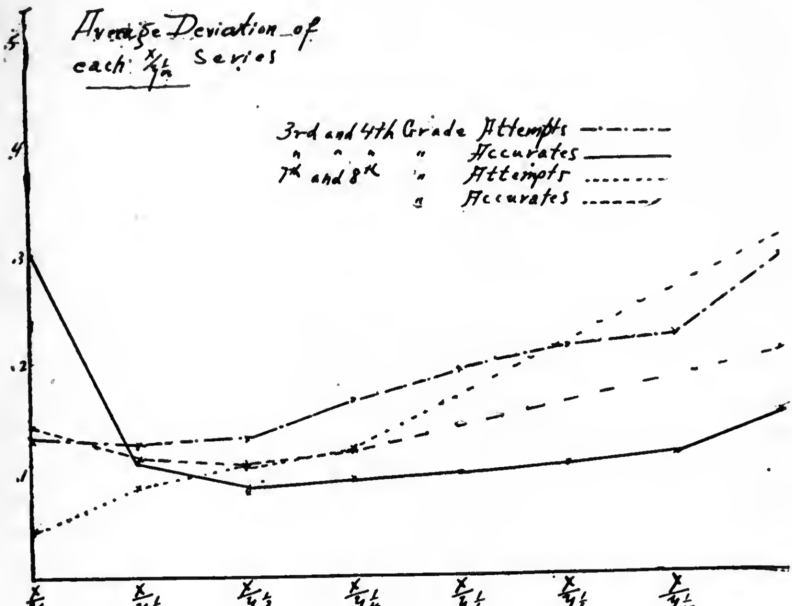


Fig. 17

Figure 17 gives the different positions on a vertical scale for the A. D. of each $\frac{x}{y^{\frac{1}{n}}}$ series for both groups in both attempts and accurates. In the case of the Third and Fourth grades, in attempts, the curve drops lowest for the $\frac{x}{y^{\frac{1}{3}}}$ series; but in accurates it is lowest in the $\frac{x}{y^{\frac{1}{3}}}$ series of values. In the case of the Seventh and Eighth grades, in attempts, the curve is lowest in the $\frac{x}{y}$ series; but in the accurates it goes lowest in $\frac{x}{y^{\frac{1}{3}}}$ as it did in the accurates for the younger children.

Corrected Variability in Accurate Performance

For corrected variability in accurate performance of Third and Fourth grades, Figure 15 shows the lines for values of $\frac{x}{y^{\frac{1}{2}}}$, $\frac{x}{y^{\frac{1}{3}}}$ and $\frac{x}{y^{\frac{1}{4}}}$ (when these are taken respectively as in the expression $\frac{x}{y^{\frac{1}{n}}}$) to be all of them nearly horizontal with the preference for a value between $y^{\frac{1}{3}}$ and $y^{\frac{1}{4}}$ values. We have taken, however, the line representing the $y^{\frac{1}{3}}$ value as the one most nearly horizontal of those in the figure. Also this value has the smallest average deviation from the average. And since these two criteria are those which shall decide in the case, $\frac{Var.}{(C.\ T.)^{\frac{1}{3}}} = C.$ is the formula taken here.

What has been said of the Third and Fourth grades group may also be said of the Seventh and Eighth grades group, Figure 16. While the $\frac{x}{y^{\frac{1}{2}}}$, $\frac{x}{y^{\frac{1}{3}}}$ and $\frac{x}{y^{\frac{1}{4}}}$ lines are all nearly horizontal, the line representing the $y^{\frac{1}{3}}$ values is probably the one most nearly horizontal; and since the $y^{\frac{1}{3}}$ series of values has the smallest average deviation, we give the preference to this value of $\frac{x}{y^{\frac{1}{n}}}$ and so $\frac{Var.}{(C.\ T.)^{\frac{1}{3}}} = C.$ is the formula taken here for this data. However, it is possible here as above that a value lying somewhere between $y^{\frac{1}{3}}$ and $y^{\frac{1}{4}}$ is the correct one for the data of accurate performance.

I think these results agree for the total work and for the accurate work. Total work gives greatest constancy at $\frac{x}{y^{\frac{1}{2}}}$ for Third and Fourth grades and $\frac{x}{y}$ for Seventh and Eighth grades. Accurate work, in the handling of these data gives, or very close to, $\frac{x}{y^{\frac{1}{3}}}$ in both the younger and older groups. It is likely that the data do not lend themselves in their handling to bearing out the theoretical facts. Theoretically this should be $\frac{x}{y^{\frac{1}{2}}}$.

Variability and Performance

Now, if we accept the formula in one case $\frac{Var.}{C.T.^{\frac{1}{2}}} = C.$, or $\frac{Var.}{C.T.^{\frac{1}{2}}} = C.$, as fitting the general run of our facts, having based our correction upon the general run of our own data, of course we cannot find whether those who do less or do more are more variable since what we have done shows them equally variable.

But we can from this graph of the accepted law of dependencies see if there are any eccentricities not according to the normal. We may say that the group has variability in excess of the general run of the data, or that it is close to the general run of the data.

In all the graphs illustrating the accepted law of dependencies we find the last one or two ratios departing, or so tending, from the norm and likewise the first one or two.

In the case of the accurate performances for the younger group, this departure from the norm is only slight (see Figure 15—$y^{\frac{1}{2}}$ line). With the older group the departure at beginning and end is a little more pronounced (Figure 16). But in the case of the attempts data the departure is quite pronounced, especially at the end of the $\frac{x}{y^{\frac{1}{2}}}$ and $\frac{x}{y}$ lines, for both groups. We may then say that this shows a tendency for the subjects who attempt most, and also for those who attempt least, to exceed the normal variability to some extent. This surplus of variability in the slowest and quickest workers is probably due to the fact that they show, more than the medium workers, a progressive change throughout the course of the work, the quickest workers increasing their speed as the work proceeds, and the slowest workers gradually slackening (See pp. 49–54).

A PSYCHOLOGICAL STUDY OF TRADE-MARK INFRINGEMENT

BY

RICHARD H. PAYNTER, JR.

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. WOODWORTH

No. 42, January, 1920

Columbia University Contributions to Philosophy and Psychology
Volume XXVI, No. 3

NEW YORK

THE SCIENCE PRESS

Agents: G. E. STECHERT & CO.; London (2 Star Yard, Carey St., W. C.); Leipzig (Koenigstr., 37
Paris (16, rue de Condé).

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

TABLE OF CONTENTS

INTRODUCTION

CHAPTER I

CHAPTER II

MATERIAL EMPLOYED AND EXPERIMENTAL PROCEDURE

CHAPTER III

RESULTS OF THE RECOGNITION EXPERIMENTS

CHAPTER IV

RESULTS OF THE RELATIVE POSITION EXPERIMENT

A PSYCHOLOGICAL STUDY OF TRADE-MARK INFRINGEMENT

INTRODUCTION

1. Aim of the Study

This study is concerned with the application of psychology to an important division of law, the infringement of trade-marks,—and its chief purpose is to help the courts and the Patent Office to decide more accurately the questions of trade-mark infringement and unfair competition. Our experiments have led us to the discovery of serious theoretical and practical objections to the present judicial procedure. Under the judicial procedure it is practically impossible to judge questions of infringement or non-infringement of similar trade-marks consistently. Psychology can place at the disposal of the courts established facts, bearing on questions of confusion, such as have never been arrived at before in any reported case. Straightforward psychological methods can determine whether the mark complained of does actually cause confusion, what the exact amount of the confusion is, and how it compares with the extent of confusion between other litigated trade-marks. The present study proceeds from the conviction that in giving definite answers to these questions psychology can be of material assistance to the courts.

2. Various Phases of the Subject of Trade-Marks

The subject of trade-marks is indeed a broad and complicated one. Names, words, marks, emblems, designs, symbols, or devices, alone or in combinations, are used as trade-marks. We have restricted our investigation to trade-mark names and words. The whole subject may be viewed from at least six different angles, *i. e.,* (*a*) the *history and development of the use* of trade-marks, (*b*) the *psychological selection* of trade-marks, (*c*) the *economics* of trade-marks, (*d*) trade-mark *laws and decisions,* (*e*) the *ethics of infringement,* and (*f*) the *psychology of trade-mark* infringement. The present discussion is primarily concerned both with the psychology of trade-mark infringement, and with trade-mark laws and decisions. The scope of these six divisions is about as follows:

(*a*) *The History and Development of the Use of Trade-Marks.*—Recent excavations reveal that trade-marks have existed as far back as 6,000 B.C. From almost the dawn of history, they were used for the same purpose as to-day, namely: to identify or indicate the origin of the commodity so marked. They were applied to a great variety of goods. Then as now they differed in kind, some being simple in design and others rather complex. When trade developed and goods came to be exchanged between neighboring towns and countries there followed a rapid increase in the adoption and use of trade-marks. There was a time in the Middle Ages when the craft guilds demanded that the merchant mark his goods as a duty to the public. Today manufacturers and dealers of all nations employ trade-marks, but the greatest increase in the number, quality, and commercial importance of trade-marks is a matter of the last two decades.

(*b*) *The Psychological Selection of Trade-Marks.*—Known contributions falling under this suggestive heading are rare and date back only a few years. Just now there is, however, a tendency on the part of some users of trade-marks to discard the customary "happy inspiration" way of adopting trade-marks and to adopt psychological methods of selection. After a large number of possible marks have been devised, experiment and statistics are employed to select the best trade-mark for advertising the particular article. Moreover, psychological study both of good and bad trade-marks and of advertisements has already led to the establishment of certain fundamental principles for determining the aptness and effectiveness of trade-marks, as a means of popularizing the products to which they are applied.

(*c*) *The Economics of Trade-Marks.*—Trade-marks play a very important part in commerce. Often the success or failure of a business is dependent on that of the trade-mark. There appear to be two methods of determining the relative financial value of trade-marks, namely, by psychological investigation and by business experience. The results of such study or experience yield information of the efficiency of the trade-mark itself, of the effectiveness of the advertising behind it, and of the good will or reputation attaching to the mark. Thru the investigations of Strong, Hollingworth, Adams, Cheney, and Geissler a foundation in methods has been laid, and initial results achieved. The vast sums of money that are spent in advertising campaigns popularize not only the merchandise but also the trade-marks. While the trade-mark is but a means of selling the article it often represents the owner's greatest asset, the cumulative good will or reputation of his business.

(*d*) *Trade-mark Laws and Decisions.*—These are measures for the protection of the consuming public and of the owner of the trade-mark. In recent years the expansion of business has made the field of trade-mark law increasingly extensive and complicated. What constitutes an illegally deceptive or infringing trade-mark is defined in the United States Trade-Mark Law of February 20, 1905, as follows: "Trade-marks which are identical . . . or which so nearly resemble a registered or known trade-mark . . . as to be likely to cause confusion or mistake in the mind of the public, or to deceive purchasers shall not be registered . . . " (Sec. 5, b). In section 16 of this act there is another statement concerning unlawful simulation of trade-marks. It says: "Any person who shall, . . . reproduce, counterfeit, copy, or colorably imitate any such trade-mark . . . shall be liable to an action for damages therefor . . . ." In order that the imitation may be actionable, the commodity bearing it must compete in use and sale with that on which the original trade-mark is used. The penalties of the law extend to restraining by injunction the use of the imitation, assessing damages, or imposing a fine or imprisonment on the infringer. The Trade-Mark Division of the United States Patent Office rejects a trade-mark offered for registration, if found likely to be confused with another trade-mark previously registered for like goods.

(*e*) *The Ethics of Infringement.*—Too little attention has been given to this aspect of the subject, to require much notice, or to be of any practical significance. The little that exists may be found in the rhetorical denunciations of the plaintiff's counsel, or in the court's reprimand to the infringer.

(*f*) *The Psychology of Trade-mark Infringement.*—In our treatment of the subject attention will be directed mainly to this topic and its application to (*d*) the laws and court decisions relating to infringement.

3. Legal Outlook

In view both of the great commercial importance of trade-marks and the desirability of an exact administration of justice in any case, the legal procedure which determines the measure of protection to be extended to a trade-mark should be trustworthy. That it is not so in fact is evident from the lack of any scientific tests for determining questions of infringement, from the blind manner in which questions of this character are handled by the courts, and from the diversity of the results attained. A number of patent lawyers and psychologists have challenged the ability of a court to determine with any certainty or exactitude the question of confusion. The reversal of so many of the reported decisions points to the same conclusion. The reason for this uncertainty and blundering by the courts is mainly, (1) defective laws, and (2) the non-possession by the court of the real facts of the case. No attempt has ever been made, so far as reported cases show, to treat the question of confusion as one of exact fact to be determined by evidence and not by inference from the marks themselves. The extremely vague statements in the law and in decisions on questions of infringement give the courts a variable standard for their guidance, and the whole matter has belonged to the realm of guess rather than proven facts. After the court has rendered its decision as to the probability of confusion no check is ever applied to determine if it was right.

There has been no exactness of knowledge as to the mental mechanism of trade-mark confusion. Our knowledge of confusion is confessedly deficient. Not a single scientific principle has as yet been recognized or evolved by the courts to aid in determining the reality of an imitation. Neither have the courts shed any light on the mental processes of the customer while buying. An introspective study as to how the courts arrive at their conclusions would be instructive, but none has been forthcoming. Such conditions are not favorable to accuracy or consistency in the decisions.

However, a number of external and significant factors that contribute to confusing the imitation with the original have been noticed and emphasized by the legal profession. Thus, it is maintained that the intelligence of the purchaser, the degree of his attention to the trade-mark while buying, the similarity of the labels, packages and goods, and the honesty of the salesman, determine, in part whether there will be confusion of trade-marks. It is often important to know whether the salesman presents the trade-mark to the customer visually of vocally; whether the trade-mark is usually seen by the customer at the time of purchaser; and whether the goods are sold in the trade-marked package, or not. An expert in dyes for example may be expected to be less easily deceived by imitative trade-marks on dyes than a person unskilled in that line. The keen interest of the expert makes him discriminate more readily between the different brands of dyes. The courts have recognized all these considerations as important, but they have differed in their interpretation and application of them, and there is, consequently, no harmony in the decisions on questions of confusion.

Unless legal procedure is satisfied to adopt psychological methods, an innovation which seems more radical than it really is, the solution of the problem before the courts will not become easier as time goes on. Rather it will become increasingly difficult and probably less accurate. New trade-marks are introduced to the public by thousands annually. Up to December 31, 1915, there were over one hundred thousand trade-marks registered in the United States Patent Office, while each year shows annual additions to this number of about seven thousand trade-marks. But the number of unregistered trademarks in use is several times the number of those registered. Thus, with each year's increase in the number of trade-marks in use, conflict of marks becomes more prevalent and litigation increases. It has also been observed that the subtlety of infringers, in devising more cunning methods of imitations, tends to outstrip the slowly advancing wheels of justice. To suppress the infringements on their trade-marks, some firms spend vast sums of money in litigation; some even maintain special legal departments, to fight trade-mark infringement and unfair competition. As many as four or five hundred suits have been instituted by one well known company against infringers of its trade-marks, labels and packages. A casual examinations of court records, or the perusal of any text-book on trade-marks will show that the most aggressive prosecutors of infringers are, in many instances, among the most successful business houses in the country. The owners of these valuable trade-marks are not, however, the only ones who suffer from the inaccuracy of court decisions. Rich and powerful houses are able to demand the suppression even of very remote imitations and to win their contentions in the courts, simply because they are able to throw a greater weight of authority and prestige into their prosecution of the case, than can be mustered by an obscure antagonist. Thus the uncetrainty of present judicial methods of determining infringement assists the larger and more resourceful business, often probably in good faith, to oppress the smaller, and to stifle legitimate competition.

4. Scope of the Work

Not only because of their commercial and legal significance have we dwelt at length upon these infirmities of our judicial procedure, but because it is the task of psychology to provide their remedies. Our study of the problem does not pretend to furnish a final solution. We realize that this is only the beginning, and that much more remains "behind." In doing our work the first step was to obtain, for measuring confusion between trade-marks, experimental methods that were capable of practical application. Two different psychological methods were employed: the recognition or identification method, and the order of merit or relative position method.

In the recognition experiment the observer is first shown a number of original trade-marks, and later a series of imitations. He is then asked to say whether the last series of marks, or any of them, were shown in him in the first series. This method furnishes exact numerical measurements of confusion, and determines the percentage of individuals mistaking the imitation for the original. In this way the effect of the imitation is accurately defined. This recognition experiment studies confusion from the standpoint of the customer. Whereas in the relative position experiment, confusion is studied from the standpoint of the court. In the latter method, the observer is required to arrange a number of pairs of litigated, and presumably conflicting trade-marks in order according to the relative amounts of confusion which he detects between the members of the respective pairs. Both methods were employed to secure a better insight into the problem of confusion and as checks upon each other. They are simple, straightforward methods requiring neither elaborate apparatus, nor an educated subject, and consume little time. The technique of the methods and the theory upon which they are based are fully described later.

A number of judicial decisions on trade-mark infringement were tested in these experiments with striking results. It was found that some imitations, which the courts had declared not to infringe, actually deceived more individuals than other imitations, which the courts found to be infringements. Thus it was made evident that the decisions in question were inexact and inconsistent, the use of some trade-marks being restrained improperly, and the use of others improperly permitted. Another series of experiments demonstrated the inaccuracy of the courts in treating all imitations, as falling into one of two distinct groups, those likely to deceive and those not likely to deceive. On the contrary these imitations exhibit varying amounts, from very little to almost absolute confusion, thereby forming a continuum or uninterrupted sequence.

The greatest assistance that psychology will be able to render to the law will be found in the construction of a scale, consisting of a number of litigated pairs of trade-marks, showing varying degrees of confusion. To help in accomplishing this, we have attempted herein to illuminate the mental processes of the observer, purchaser, and court. An analysis,—and incomplete one, it is true—has also been made of some of the psychological factors entering into confusion. By examining the methods employed in devising imitations, several principles have been formulated as a general guide to aid in the detection of dangerous imitations. Other points of interest, especially to the psychologist, include the positive correlation found between the results of the recognition and relative position methods, and between the results of the recognition method, when used both with and without knowledge on the part of the subjects experimented upon, of the presence of imitations in the experiment.

Psychology hopes to render a service to law and business in this field by placing their problems upon scientific foundations. By thus simplifying the task of the courts much labor may be saved. The expenditure of time and money by the state, by courts, and by litigants will doubtless be greatly lessened. By the application of scientific methods more consistent and harmonious decisions will be rendered, and a system of trade-mark law more practically useful and more scientifically symmetrical will be developed.

In Chapter I. the literature relating to our problem is reviewed. In Chapter II. the technique and material employed are described. Chapter III. gives actual measurements of the amount of confusion between deceptively similar trade-marks. Chapter IV. deals with confusion as measured relatively. In those chapters two illustrative scales of confusion between trade-marks are shown, and their construction is demonstrated. Chapter V. presents the results of two psychological tests on the accuracy of the judicial decisions, together with some of the principals to be applied in detecting dangerous imitations.

5. Acknowledgments

I very gratefully acknowledge the help that I have received from my subjects and teachers. Professors J. McK. Cattell, R. S. Woodworth, and H. L. Hollingworth have rendered much valuable aid. Mr. Arthur Wm. Barber, Secretary of the United States Trade-Mark Association, has also given me assistance. To my friend Dr. S. A. Tannenbaum, who reworded and clarified many statements in the Ms., I am deeply grateful. To Captain A. T. Poffenberger, Jr., Chief Psychological Examiner, U. S. A., at Camp Wheeler, Georgia, and of Columbia, I am also deeply grateful for his kind help and advice in numerous instances. I am most indebted to Lieut.-Colonel Edward K. Strong, Jr., of the Committee on Classification of Personnel, Adjutant General's Department, who suggested the problem to me, and planned much of the experimental technique, and whose assistance has been invaluable throughout.

CHAPTER I

Historical Account

The literature of psychology as applied to trade-mark infringement is drawn principally from two independent fields of thought, law and psychology. It begins with the appearance of an article on "The Market and Psychology" by Professor Münsterberg.[1] This article is of importance in that it calls attention for the first time to the application of psychology to this fertile branch of law. In criticizing the existing legal procedure, Münsterberg points out that there is no "definite standard" by which to judge an imitation. The amount of attention which the customer ordinarily shows in buying, his intelligence, and the degree of deceptive similarity that is illegal are subject to very different interpretations with the judges. The plan of solution suggested by him consists in the use of a scale, with different degrees of attention, and varying according to the difficulty of the recognition of certain impressions. The scale is to be constructed under laboratory conditions, including the "mental principles involved" in everyday purchases.

In his "Psychology and Industrial Efficiency," Münsterberg has a chapter on "Experiments with Reference to Illegal Imitation,"[2] but gives no data. He declares that the inability of the law to determine by means of general conceptions the exact point at which infringement begins constitutes a source of economic disturbance that cannot be removed until the psychological background has been systematically studied. Here, he thinks, the only aim of the psychologist should be to construct a scale of various similarity values, by which decisions may be made comparable and by which standards may be obtained. A little further on he suggests a somewhat different procedure, namely, to find an exact formula which may be adjustable to any marketable material. Then it will be possible to measure the deceptive similarity of an imitation, independently of individual arbitrariness, from the percentage of subjects that discovered the substitution under certain experimental conditions. Thus, provided the methods were accepted and the

[1] "American Problems," 1912, Chap. 7, 151–173; this paper first appeared in *McClures Magazine*, 1909–1910, 34, 87–93, under the title "Psychology and the Market."

[2] Pp. 282–293, 1913.

degree of similarity necessary to constitute infringement were agreed upon, all uncertainty would disappear. This work was taken up by Dr. G. A. Feingold in a monograph that is reviewed later on in this chapter.

The year following the first appearance of Münsterberg's article brought forth an article[3] by Mr. Edward S. Rogers, a patent lawyer of the Chicago bar. Chronologically Rogers' article should have appeared first, for it was he who suggested to Münsterberg the application of psychology to problems of trade-mark imitation. Rogers gives a critical analysis of the judicial procedure from the legal point of view. He believes that cases of trade-mark infringement and other similar forms of unfair trading show an "irreconcilable conflict," due to the courts' neglect of the psychology of the unwary purchaser. In the following paragraph Rogers illustrates how the unwary purchaser is merely a judicial myth:

> " . . . when the court thinks the exhibit sufficiently alike that a judicially ideal unwary purchaser ought reasonably to be deceived and one steps forward and testifies that it has happened, it is hailed as a confirmation of the court's judgment and much is made of the testimony, but if in such a case no testimony of actual deception is adduced the "unwary purchaser" is pressed again into service on the pretext that he is in the course of his imaginary purchasing to be imposed upon by the imitation,[4] . . ."

The unwary purchaser is usually assumed to represent the average person, but sometimes he is identified with any type of individual. This is indicated in two statements by Rogers, who strangely enough seems not to have been aware of the contradiction. In the one, the unwary purchaser is distinguished from the intelligent, expert, and careful person, and identified with the ultimate, ordinary, normal, everyday, unpracticed, inattentive and ignorant purchaser.[5] In the other, it is explicitly stated that the unwary purchaser may appear in almost any station in life.[6] With such different interpretations of terms it seems impossible to attain consistent decisions.

In a very interesting book,[7] Rogers takes up in a critical manner an historical survey of the development of trade-marks and of judicial procedure relating to trade-marks, and gives a number of useful suggestions, on the economic and advertising aspects of trade-marks. In criticism of the present judicial procedure, he writes:

[3] "The Unwary Purchaser: A Study in the Psychology of Trade-Mark Infringement," *Michigan Law Rev.*, 1910, 8, 613–622.

[4] *Ibid.*, p. 621.

[5] *Ibid.*, pp. 613–614.

[6] *Ibid.*, pp. 615–616.

[7] "Good Will Trade-Marks, and Unfair Trading," 1914.

"After an hour or so of theoretical disputation, the court takes the case under advisement and after a while writes a nice theoretical opinion and decides one way or the other, depending very largely on the eyesight of the judge and his capacity or incapacity for being fooled himself, and from the record before him no one can demonstrate that he is wrong, whichever side he takes."[8]

Rogers maintains that, in the absence of psychological methods, the remedy for this uncertain state of affairs exists in proven cases of actual deception of normal individuals. It seems to me that the opposing counsel might counter-balance those instances by an equal number of proven cases showing no actual deception, and thus leave the question of deception as unsettled as before. Rogers gives a long list of word trade-marks that have been held to infringe.[9] It should be noted, however, that he errs in stating that they "have been held to infringe as words simply and independently of their surroundings and accomplishments." Tho many of these decisions on these trade-marks are simple, careful reading of the records shows that many others are undoubtedly complicated, involving various features of the labels, packages, and other legal technicalities.

Two papers by Professor Cattell that appeared about twenty-five years before those of Münsterberg and Rogers may also be considered as constituting the starting point of our problem. In those studies he obtained a number of definite results on the psychology of reading, the great scientific and practical importance of which he made clear. Their relation to the problem of our thesis will readily be noticed. Cattell found[10] that, as the time of perception is only slightly longer for a word than for a single letter, words are read as a whole, the letters composing it not being perceived separately. He also found that the time is longer for long or rare words, and for words in a foreign language; that different letters in the same alphabet are not equally legible; that certain letters like capitals S, C, and E are hard to recognize in themselves; that some letters are mistaken for similar letters, as in the case of capitals O, Q, G, and C, and in the case of the small letters i, j, l, f and t. One of the many interesing results in the second article[11] by Cattell, is that it took a slightly longer time to distinguish words that are very similar in visual appearance, as "hand" from "band." The time of discrimination, therefore, furnishes us with an additional method for studying confusion.

[8] *Ibid.*, p. 131.

[9] *Ibid.*, pp. 142–143.

[10] "The Inertia of the Eye and Brain," *Brain*, 1885, 31, 13, 16–17.

[11] "The Time Taken up by the Cerebral Operations," *Mind*, 1886, 11, 387.

In 1911 Mr. Arthur Wm. Barber, Secretary of the United States Trade-Mark Association, studied the problems of trade-mark infringement and unfair competition from the legal and psychological standpoints.[12] He believes that every trade-mark problem is a psychological problem, and that the legal rules governing their use will gain immeasurably in certainty and definiteness from the application of psychology thereto. He says that the greatest service that the psychologist can render is in solving questions of infringement of trade-marks and imitations of packages. How primitive the legal procedure in such cases is he shows plainly in the following remarks:

"Usually in our helplessness, we bring the trade-marks, labels or packages into the court, show them to the judge, and according as we are on one side of the case or the other, point out their likenesses or analyze their differences. Then the judge, in the exercise of his judicial common-sense, settles the rights of the parties as they look to him."[13]

"What we get from the court in a case thus presented is never the determination of a fact upon conflicting evidence, but the mere opinion upon the probability of a future event, the deception of the purchaser."[14]

This author thinks that experiment should determine the degree of ease, or difficulty with which an average person recalls the impression made upon him by an article which he is used to buy; as well as the clearness or haziness of that impression, and the extent to which the clearness of that impression varies with the character of the article, its uses, and its price. The vividness of the recollection of the original article depends on the character of the trade-mark, the intelligence of the purchaser, the importance and the frequency of the transaction, *e. g.*, whether the article purchased is a cigar or a piano. Furthermore, experiment should be able to determine upon what particular features of a given trade-mark, label, or style of packing recognition mainly depends; and what degree of similarity to the familiar form of the original article, its style of marking or packing, is confusing. Barber concludes with a pleasing picture of the improvements in the methods and results of judicial inquiry, when the testimony of psychology shall replace the conjectures of the court.

The only case that I have been able to find containing anything like a psychological test of confusion is one that came up in the Patent Office in 1912.[15] The word "King" had been refused regis-

[12] "The Psychology of Trade-Marks," *Bulletin of the U. S. Trade-Mark Assn.*, 1911, 7, 152–165.

[13] *Ibid.*, p. 162.

[14] *Ibid.*, p. 163.

[15] *Ex parte*, J. C. Blair & Co., 2 Trade-Mark Rep., 483, 104 Ms., Dec., 136, Jan. 13, 1913, Manuscript Decision of the Patent Office.

tration by the Examiner of Trade-Marks, because of its deceptive similarity to the previously registered trade-mark "Shen-King" used in connection with similar goods. This decision was, on appeal, over-ruled by the Assistant Commissioner, and the word "King" was admitted to registration. On the argument, applicant's counsel stated that, in a test made by him, the adult members of his family said that they would not purchase paper bearing the trade-mark "Shen-King" when they desired to purchase the paper marked "King." The Assistant Commissioner states that he put the test to several employees in the Patent Office and a majority immediately responded that they would not be confused by the concurrent use of the two marks.

Both counsel and the Assistant Commissioner committed the error of testing for the confusion of "Shen-King" with "King." Their questions sought information as to whether goods of the original or first mark "Shen-King" would be mistaken for those of the second or imitative mark "King." What in fact usually happens in the use of trade-marks is that acquaintance is first formed with the original or earlier mark, and that confusion is more likely to happen in the direction of mistaking the imitation which appears later for the original than of mistaking the original for the imitation.

The language of the law and of the judicial decisions is clear that the confusion to be considered is that of the imitation with the original. That there may be an equal amount of confusion of "King" with "Shen-King" as of "Shen-King" with "King" would be no reason for using the couplet of trade-marks in the latter order. Tho there would in some cases be little or no difference in the amount of confusion found by means of either couplet, in other cases there may be a great deal of difference. Neither the Assistant Commissioner nor the applicant's counsel give any figures from their tests. The counsel's family were unanimous that in their opinion the goods under the two trade-marks would not be confused. But a test for the deceptive similarity of trade-marks is not necessarily determined by the existence of deceptive similarity between their goods. It is not clear from the record whether the majority of the Patent Office employees would not have been confused by either the goods or by the names. Furthermore, the opinions of those tested do not give information as to the actual existence or non-existence of confusion. The reliability of the counsel's statement must be decidedly lowered in view of the probable prejudice resulting from the wish of his family to see him win the case. The Patent Office employees are also unsuitable subjects for tests of this character, because they are a specially skilled class of individuals, who are more than ordinarily familiar with the subject of imitations.

An imitation trade-mark and its article may not only be mistaken for the original trade-mark and its article but may, as Professor Hollingworth remarks, simply remind us of the latter or the general atmosphere surrounding it.[16] A trade-mark should popularize the article on the market and then keep it popular, by virtue of its being easily recalled, recognized, and asked for. To effect the choice of such a trade-mark the memorability of different kinds of facts must be considered. As Hollingworth says:

> "In selecting a trade-mark by which goods, designed for popular consumption, are to be known, it is of real value, for instance, to know that *persons*, and *faces* are more easily remembered than *objects*, and *objects* more easily than *actions;* that *form* is more easily remembered and recognized than numbers. More numbers can be remembered than colors, but they are likely to be wrongly remembered or remembered as existing in a false order or position."[17]

He also found that the relative accuracy in reporting different kinds of facts when subjects are directly questioned about them, runs from 97 to less than 8 per cent. The order of accuracy is: presence of things, number of people, space relations, conditions of objects, order of events, color, size and quantity, sound, time, and actions. From this we may infer that, as parts of a trade-mark, these features are likely to play varying rôles in producing confusion.

In a second experiment, perhaps the first psychological study of trade-marks, Hollingworth found large variations in the relative attention and memory value of geometrical forms representing many common trade-marks symbols.[18] The correct recognitions ranged from 28 to 92 per cent. "The general principle suggested by this experiment is that those forms are best remembered to which specific names can be given, as 'star,' 'crescent,' 'crown,' etc."

In this connection, Miss Edith Mulhall working with pictures, forms, words, and nonsense syllables suggests also that trade-marks having a wealth of associations will be recalled and recognized more easily.[19]

There is no doubt that upon the recognitive value of a trade-mark may depend its chances of being displaced by an imitation. In his chapter on "The Psychology of Trade-marks and Trade-Names," Hollingworth emphasizes the principle that trade-marks to be valu-

[16] "Advertising and Selling," 1913, 198.

[17] *Ibid.*, pp. 208–210.

[18] *Ibid.*, pp. 212–213.

[19] "Experimental Studies in Recognition and Recall," *Amer. J. of Psychol.*, 1915, 26, 218, 226.

able must make vivid and permanent impressions, inasmuch as recollection and recognition are their primary functions.[20]

In the principles underlying the invention of trade-marks may lie one reason why they so often conflict. A large number of trade-marks have been grouped by Professor Louise Pound according to the mode of their formation. It is seen that those that fall under the same heading often display a striking similarity.[21] There are all kinds of word or syllable combinations, shortenings, and extensions, diminutions, arbitrary new formations, fancy or phonetic spellings, striking hypenations, novel capitalizings, and blendings. It is evident that when it becomes popular to add the same kind of prefix, suffix, or diminutive to trade-marks, a greater likelihood of confusion results.

To the recent increase in our knowledge of recognition no one has contributed more than Professor Edward K. Strong, Jr. His careful investigations of a variety of problems have yielded results of great theoretical and practical importance.[22] They include experiments on the nature of recognition, its relation to certainty, localization, association, free association, the effect of the size of advertisements and the frequency of their presentation on recognition, of the length of the series, and of the time interval on recognition and its relation to Ebbinghaus' curve of forgetting. The recognition procedure used in many of these experiments has been the model after which that in the present investigation has been fashioned. It is the main purpose of advertising, Strong maintains, to develop a very strong associative bond or connection between the need for a commodity and a trade-mark, and the development of a very favorable attitude toward the latter.[23]

Dr. H. F. Adams makes an important contribution in his experiments measuring the strength of the bonds between commodities and trade-marks.[24] On the average the association time was 1.90 seconds when the stimulus word was the trade-mark and 2.36 seconds when

[20] "Advertising, Its Principles and Practise," Tipper, Hollingworth, Hotchkiss, Parsons, 1915, 132–133.

[21] "Word-Coinage and Modern Trade-Names," *Dialect Notes,* January, 1914, 29–41.

[22] "The Effect of Length of Series upon Recognition Memory," *Psychol. Rev.,* 1912, 19, 447–462; "The Effect of the Time-Interval upon Recognition Memory," *ibid.,* 1913, 20, 339–372; M. H. Strong and E. K. Strong, Jr., "The Nature of Recognition Memory and the Localizations of Recognitions," *Amer. J. of Psychol.,* 1916, 27, 341–362; and others.

[23] "The Effect of Size of Advertisements and Frequency of their Presentation," *Psychol. Rev.,* 1914, 21, 139.

[24] "Advertising and Its Mental Laws," 1916, 180–195.

the stimulus word was the name of the commodity. When the trade-mark was the stimulus the subject was asked to respond by naming the first commodity suggested, and vice versa. Dr. Adams believes that this experiment is "a measurement of the advertisements which have been most effective with any given individual." His tables clearly show which trade-marks are most often thought of in connection with their commodities; and which commodities bear trade-marks that are well known, or little known. The commodities most widely advertised were most frequently mentioned in the experiments. Another experiment of Adams showed a certain amount of confusion between trade-marks and slogans, and between the commodities to which they were attached, and pointed to inefficiency in advertising.[25] He cites a number of tables from a study by Cheney.[26] They indicate the strength of associations between the following couplets: firm names and their products, trade-marks and their products, trade-marks and their firm names, firm names and their trade-marks, slogans and their firm names.[27]

A very interesting and instructive study has been made by Dr. L. R. Geissler in connection with trade-marks and commodities.[28] He planned to see how one could be best persuaded to buy a certain brand of goods and to obtain a "general insight into the mind" of a prospective purchaser of a given article. To do this experiments were conducted to see which brands of articles are thought of first, most frequently, and why.

In an earlier study by the writer[29] of the accuracy of court decisions, on questions of confusion of trade-marks, the figures representing the percentage of confusion produced by infringing and non-infringing imitations are incorrect. More careful examination of the litigated cases from which these imitations were selected indicated that some were not appropriate for use in testing the accuracy of the decisions. The authors of the textbooks from which some of the imitations were selected had committed numerous errors, as was later discovered. This necessitated the omission of a number of imitations that had been included among the infringements and non-infringements, and accordingly changed the average amounts of confusion and the average percentage of overlapping, which gave the

[25] *Ibid.*, pp. 203–204.

[26] *Ibid.*, pp. 175–179.

[27] *Printer's Ink*, August, 1914, 61–62.

[28] "Association-Reactions Applied to Ideas of Commercial Brands of Familiar Articles," *Jour. of Applied Psychol.*, 1917, 1, 275–290.

[29] "Experiment versus Court Decision," an abstract of a paper read before the New York Branch of the Am. Psychol. Ass., *J. of Philos., Psychol., Etc.*, 1915, 12, 45–47.

measure of the accuracy of the decisions. Nevertheless, the general conclusion of the experiment,—that the judicial decisions were unreliable,—remains unimpaired.

Dr. Gustave A. Feingold, working under the direction of Münsterberg, has made a very interesting study that is more related to our own than is any other that has been so far made.[30] The purpose of his investigation was to supply "a scientific guide to courts of law whereby they could settle disputes arising from Trade-Mark Infringement more equitably."[31] His study is also suggestive and valuable to psychological theory because he made several important discoveries about recognition.

Before considering his guide or Correction Formula for the courts we shall briefly review his methods and some of his findings. In one set of his experiments Feingold used words as his material and in another picture post-cards. Two methods were used by him in getting the objective similarity of the various pairs of words, the "mathematical" and the "psychological," as he calls them. He alone measured the mathematical similarity by rating each pair of words, on a percentage basis, according to the number of symmetrically or correspondingly arranged identical letters that they contained. The psychological similarity was obtained by having 35 individuals estimate the similarity of each pair of words in terms of percentage. The next step was to see how well each pair of words could be recognized. Eight words were typewritten in capital letters in two vertical columns on 3- by 5-inch plain filing cards. These cards came in pairs; the first was the original card shown to the subject; and the second the variable which, with the exception of the change, was a reproduction of the first. In all the experiments the interval between the first and second exposure was twenty seconds. The original and variable cards were each exposed for four seconds. The cards were seen thru a window opened and closed by a drop shutter. Before the experiment began the subject was told that he would be required to name on the second card all the words that were changed and the original ones that had been displaced. On the variable card all the words might be changed, all but one, only one, or none, *i. e.*, 8, 7, 1, or 0. After the recognitions, introspections were asked for.

In connection with the experiments with words, three methods of recognizing an object were noted:[32] (1) by memory images that

[30] "Recognition and Discrimination," *Psychol. Rev., Monogr.*, 1915, 18, No. 78.

[31] *Ibid.*, p. 111.

[32] *Ibid.*, pp. 44–45.

persist from the presentation to the test; (2) by the revival of faded memory images thru the re-perception of the original object or merely by the perception of the position originally occupied by the missing object; or (3) it may be a matter of feeling, kinæsthetic attitude, adjustment, or of the revival of a mood. In connection with the post-card experiments two methods of recognition were observed.[33] The variable card was recognized because it lacked something that appeared on the old card, or because a new element was noticed. Among the objective factors[34] that determined recognition, a change of position, as right and left, up and down, was the most easily recognizable factor; change of direction was next most easily recognized; human beings were more distinguishable in various attitudes than animals; and animate things were in general more distinguishable than inanimate.

From a comparison of the introspective and objective data, Feingold brings out the following points, which are instructive in a psychological analysis of errors of recognition:[35] Failure to remember the original; incomplete perception of the original; incomplete perception of the variable, mistaking it for the original, owing to the superiority of the ideo-motor force over the sensory-motor force. "With high similarity the majority of errors are due to the obliteration from memory of the original N [normal] impression by the perception of the V [variable] stimulus, owing to the superiority of the sensory-motor force over the ideo-motor force." "Doubt arises from a conflict of these two factors—the perception or the memory triumphing according as objective similarity is high or low." Errors also occur with the rise of a new association in connection with the original which makes it seem new; and with the failure of the original to revive the same associations as when first presented.

The complexity of the picture-post cards made it difficult to analyze the errors of recognition,[36] but by far the majority of errors were due to the obliteration from memory of the original by the perception of the variable. The feeling familiarity aroused by the identity of the other cards was another source of error. "In general, the mistakes were directly due to the degree of similarity, . . ."

His experiments took up the following variants:[37] (1) The effect on recognition of the distribution of attention; (2) of the time of perception; (3) of similarity and whether there is a "mathematical"

[33] *Ibid.*, p. 92.
[34] *Ibid.*, pp. 98–99.
[35] *Ibid.*, p. 49.
[36] *Ibid.*, pp. 99–100.
[37] *Ibid.*, p. 8.

relation between the two; (4) the difference between the effects on recognition of structural and meaningful similarity; (5) the difference between substitution and interchange of position; and (6) the influence of old and new environments.

It is found[38] that recognitions varies "inversely as the number of objects perceived—time being constant"; "inversely as the number of objects exposed—time being proportionate and not more than one second per object"; and "directly as the temporal length of perception."[39] But, for reasons which cannot be discussed here, we do not agree with Feingold's conclusions on the influence of the old and new environments.[40] The ratios of their difficulty are far too high, and the arithmetical reasoning behind them is inconsistent and wrong.

Similarity in meaning and in structure are found to produce confusion in recognition.[41] It is also stated that similarity in meaning "has no fine gradations, the highest attainable degree being that which would fall midway on the scale of structural similarity." We cannot, however, accept this statement because the very data on which the two points in it are based show them to be false. Fine gradations are shown not only in similarity but also in recognition, as in columns 4 and 5 of Table 5.[42] Feingold's use of the phrase "structural similarity" seems ambiguous here, but perhaps means the percentage of correct recognitions to which it is inversely related. The pair of words " obtain—acquire" in Table 5 with 10 per cent. correct recognitions or 90 per cent. confusion, lies far above the midway point on the scale, in fact it is very near the upper limit.

Next we will pass to one of Feingold's principal conclusions. "There is a simple inverse relation between degree of similarity and recognitive ability, the one being to the other in terms of percentage, as X: $(100 - X)$."[43] The accuracy of Feingold's statement, as to the relation of similarity and recognition, is of vital importance to the validity of the guide or Correction Formula. At first sight the statement seems plausible enough, but let us look at the data. This law, as Feingold calls it, is based on a comparison of the degree of mathematical similarity of forty pairs of words with the percentage of time that they are correctly recognized. When the averages that support this law are consulted in Table 8[44] it is found that they

[38] *Ibid.*, pp. 113–114.

[39] See the remarks on these three conclusions in the paper by M. H. Strong and E. K. Strong, Jr., *op. cit.*, pp. 349–350.

[40] *Ibid.*, pp. 61, 113.

[41] *Ibid.*, p. 60.

[42] *Ibid.*, p. 27.

[43] *Ibid.*, p. 61.

[44] *Ibid.*, p. 36.

hold only approximately, the variations running from 2 per cent. to 12 per cent. in the uncorrected results, and from 4 per cent. to 17 per cent. in the corrected results. The duplicate and new scores are not taken into account here. When the two scores of the individual pairs of words are consulted, the variations run from 0 per cent. or exact agreement to 62 per cent. or more than six-tenths of the distance to complete disagreement; these are uncorrected results and no corrected ones are given. When the variations from a formula (which does not indicate that there are any divergencies) are as large as 62 per cent. at a single step, the extremes of the scale differing by only 100 per cent., it is objectionable to state that the inverse relation is a simple or exact one. Furthermore, no figures are given to show the limits within which the relation between degree of similarity and recognition is inverse. There is no doubt that among many words an inverse relation holds, but Feingold's results do not justify the ratio X: $(100-X)$.[45] Neither do the tables containing the results of the picture-post cards experiments reveal evidence for this simple inverse relation.

Any guide for practical application in the courts should be good psychology, and should be examined carefully by psychologists. Feingold offers a Correction Formula as a guide. It is stated in his preface that to furnish such a guide it is necessary "to construct a scale of graded similarity among meaningful objects of the same category, and then to find out what relation there exists between each unit of the scale and ability to recognize under such conditions of attention, perception and judgment as most prevail in actual life." But the determination of the relation between similarity and recognition is not necessary for a scientific guide nor does Feingold show that it is. This determination is a very complicated affair, and will therefore have to be considered in detail.

Feingold first proceeds to develop the Correction Formula on the basis that between similarity and recognition there is a simple inverse relation in terms of percentage as X: $(100-X)$,—a relation which as shown in the second preceding paragraph does not hold. As the inverse relation is the very thing the formula is to effect, besides being a guide for the courts, it evidently cannot obtain support from it. The formula requires in application much calculation, but changes the score generally only a little. It is not possible to analyze the errors in it as he tries to do.

There is still another fault to be found with Feingold's formula, and it is a most important one as it may have a serious effect on

[45] See Tables 16, 20, 25, 29, *ibid.*, pp. 71, 73, 78, 83; and Plate II., p. 84.

practical affairs. His formula does not only so change his recognition results as to produce the simple inverse relation between similarity and recognition, and so make them ideal, but it makes many of them still less nearly ideal than if they had been uncorrected. In his Table 8, out of 6 degrees of similarity, 4 are made less ideal and 2 are unaltered. In other tables, many scores are also unaltered by the formula, and some are altered only very slightly in the ideal direction. His Correction Formula is therefore inaccurate and impractical.

In order to establish a relation between degree of similarity and recognition, Feingold thinks[46] it is necessary to find those conditions which, with the greatest number of items exposed for the shortest possible time would give 100 per cent. of discriminations of a new item of 0 per cent. similarity, and 100 per cent. of correct identifications of a duplicate item of 100 per cent. similarity. These are called ideal conditions and their attainment would give ideal results. The errors of duplicates and new items seem to be due either to the number of words exposed, or to the length of exposure, or to both.[47] Tho these errors are said not to distort the general results, they obscure the exact or ideal relation between the recognition of the various degrees of similarity.[48]

It was his object to find that combination of length of exposure and number of objects exposed which would give such ideal conditions and results.[49] Throughout the entire investigation these conditions were never obtained exactly, but he devised a Correction Formula[50] which he says changes the results to what they would have been, if the conditions had been ideal. By eliminating the chance correct and chance incorrect recognitions from all the recognitions and thus separating those recognitions (correct and incorrect) due to similarity, the formula is intended to find the exact relation of degree of similarity to recognition. Tho the formula is based on experiments, it is not strictly empirical nor scientific, because when examined carefully the experimental results represented in the formula are found to rest on an unequal basis and to have unwarranted interpretations. This is the Correction Formula:[51]

[46] *Ibid.*, p. 35.

[47] *Ibid.*, p. 64 he says a "definite mathematical relation."

[48] *Ibid.*, p. 24.

[49] *Ibid.*, pp. 35–36.

[50] *Ibid.*, pp. 38–40.

[51] So as to make it more easily grasped here I have substituted for the alphabetical symbols the things Feingold means them to signify.

$$\left\{\begin{matrix}\%\text{ Difference or }100\,\%-\\ \%\text{ Similarity}\end{matrix}\right\}\times\left\{\begin{matrix}\%\text{ Incorrect Identifi-}\\ \text{cations of a New}\\ \text{Item}\end{matrix}\right\}-\%\text{ Similarity}$$

$$\times\left\{\begin{matrix}\text{Incorrect Discrimina-}\\ \text{tions of all Dupli-}\\ \text{cate Items}\end{matrix}\right\}\times\frac{\text{No. of Items Changed}}{\text{No. of Items Exposed}}=\text{K.}$$

To obtain the correct result this value (K) is added algebraically (with a + or — sign) to the percentage of correct recognitions of the variable item, the percentage of whose similarity is given in the formula.

The formula involves a number of assumptions that seem to be untenable and otherwise objectionable. In the first place the errors or chance recognitions of the duplicate and new items are carried over to the variables.[52] The errors of the variables should undoubtedly be calculated from their own recognitions, and not be made to depend upon the errors of the duplicate and new items. The psychological bases of the errors or chance recognitions of the duplicate and new items are not entirely the same as those of the variables. The chance incorrect discriminations of the duplicates are transmuted into chance correct discriminations of the variables. The resort to guessing in the case of the chance incorrect discriminations of the duplicates is generally due to the forgetting of the first of the two duplicates and in the case of the chance correct discriminations of the variables to the forgetting of the dissimilar original. The chance incorrect identifications of the new items are also transmuted into chance incorrect identifications of the variables. The resort to guessing in the case of the chance incorrect identifications of the new items is due generally to the failure to note that the new did not appear in the presentation, and in the case of the variables to the failure to discriminate the variable from its original. Feingold, of course, considers that there is a difference between chance correct discriminations and correct discriminations of a variable and between chance incorrect identifications and incorrect identifications; the former in each case being the result of a guess, and the latter of a conscious reaction to objective dissimilarity. Furthermore, the errors of the duplicates and of the new items are not calculated on the same basis. The per cent. of chance incorrect identifications of the new items is computed from the errors made in connection with a single new item exposed in a test among 7 duplicates or in an old environment.[53] The per cent. of chance incorrect discriminations

[52] *Ibid.*, pp. 23–24, 37.

[53] *Ibid.*, pp. 36, 38, the results of a number of such exposures are used, but in the illustration of the practical application on p. 124 only one is used.

of the duplicates, on the other hand, is computed from the average of the errors of all the 8 duplicates exposed in the test or of the entire old environment.[54] The discrimination of a single new item, being in a different environment, is made under a less favorable condition than the recognition of the duplicate which appeared in the same environment. To be comparable with the errors of the duplicates those of the new items should come from an entire group of 8 new items or from a new environment.

In the second place, owing to the fact that Feingold proceeds on the theory that "there is a simple inverse relation between degree of similarity and recognition,"[55] the errors of the duplicate and the new items enter in definite proportion in each degree of similarity of the variable. He is therefore arguing in a circle, for it is just the exact relation between similarity and recognition that the formula is to make, and therefore, cannot rest on that assumption. The relation is really not simple, tho it is, in general, inverse. Feingold's position includes the two additional assumptions that (1) chance correct discriminations of a variable item vary in direct proportion and that (2) chance incorrect identifications of a variable item vary in inverse proportion to its similarity.

Concerning chance correct discriminations of variables Feingold states:

> "But we cannot say that the same amount of error [incorrect discriminations of duplicates] would enter into the judgments of change rendered for objects of less than 100 per cent. S. For such objects would possess a certain amount of dissimilarity which would have a tendency of its own to elicit the judgment 'change.' And in proportion as such objects were removed from 100 per cent. S, in that proportion would that inherent tendency increase, while the tendency to say 'change' on the basis of illusion or mere chance would diminish."[56]

Reasoning in like fashion concerning chance incorrect identifications of variables, he says:

> "For we have no right to assume that 9 per cent. [belonging to words of 0 per cent. S] of the substitutions of 87 per cent. S also escaped recognition owing to the conditions of the experiment, because such words contained a certain amount of identity—87 per cent.—which would have a tendency of its own to elicit the judgment 'the same.' Likewise with words of every value of S."[57]

On the surface, Feingold's analysis is likely to seem true, but this is due to the apparently logical character of his explanation. With the higher degrees of similarity and the more incorrect identifica-

[54] *Ibid.*, pp. 37–38, 124.
[55] *Ibid.*, p. 37, and see pp. 23–24.
[56] *Ibid.*, p. 37.
[57] *Ibid.*, p. 38.

tions, there are fewer chance incorrect identifications, and vice versa, and with the higher degrees of similarity and the fewer correct discriminations, there are more chance correct discriminations, and vice versa. In other words, there is a proportionate decrease of chance incorrect identifications with an increase of incorrect identifications, and there is a proportionate decrease of chance correct discriminations with an increase of correct discriminations. Moreover, according to Feingold's position there can be no chance correct discriminations of new items of 0 per cent. similarity, nor can there be any chance correct identifications of a duplicate item of 100 per cent. similarity; and this is implied, whether or not there are any errors of the duplicate and new items. The assumptions are evidently false. If, in the experiment, there were no chance incorrect discriminations of the duplicates, the formula would not serve to correct chance correct discriminations of the variables; and if there were no chance incorrect identifications of the new items, the formula would not correct for chance incorrect identifications of the variables.

The first step in the formula takes it for granted that similarity and difference are inversely related precisely, *i. e.,* when the similarity is 0 per cent. the difference is 100 per cent., and vice versa; and when the similarity is 25 per cent. the difference is 75 per cent., and vice versa. Table 8[58] of Feingold's monograph does not show this to be true; it would hold perhaps if there were a simple inverse relation between similarity and recognition. The inverse relation between similarity and difference, and similarity and recognition is also disturbed by the fact that Feingold himself finds no confusion below 20 per cent. or 25 per cent. similarity.

The criticism, in brief, of the Correction Formula is that it is based on false psychology, that it is unscientific, inaccurate, and impractical. Its practical outcome is quite negligible. In the closing chapter on The Application of the Laws of Recognition to Jurisprudence, Feingold illustrates the application of the recognition method and the formula in the court room. Tho the experiment is adjusted for ideal conditions, a few errors among the duplicate and new items cause Feingold to say, "Now apparently either the conditions of the experiments were not ideal or some of the observers were not absolutely normal, . . ."[59] The Correction Formula is then applied to obtain the ideal results. The uncorrected score is 42 per cent. of correct discriminations of the imitation; the formula shows that the amount of correction should be $+0.5$; and according to

[58] *Ibid.,* p. 36.
[59] *Ibid.,* p. 125.

Feingold's computation 42.5 per cent. of correct discriminations is the ideal result. He has spent a vast amount of labor to develop the formula which yields the insignificant change of $+0.5$. The erroneous assumptions on which the formula rests and the results obtained by it prove it to be useless.

It would certainly be unfair to judge all of Feingold's work by the Correction Formula. His experimental technique presents a good study in method. Moreover, some of the conclusions that we first mentioned will be found to be of decided psychological interest and others to have a direct practical bearing on various trade-mark problems.

To devise a scientific guide for the courts it is not necessary to establish a relation between recognition and similarity. Nor is it possible, as Feingold has attempted to do, to analyze completely the intricate errors of memory. A scale of confusion including a sufficient number of individuals to overcome such errors would be all that is necessary. In other words, it is not necessary nor is it possible at the present time, to arrive at a final conclusion regarding confusion, thru a synthesis of the many factors which comprise confusion. But it is possible to measure directly the amount of confusion involved. And this is just what we have done here as distinguished from what Feingold attempted to do.

CHAPTER II

MATERIAL EMPLOYED AND EXPERIMENTAL PROCEDURE

1. Material Employed

This chapter describes the trade-marks used and the manner of conducting the experiments. Sixty pairs of litigated trade-marks, all word trade-marks, were picked at random from a large number of court decisions. The 120 were all employed in the experiments; no trade-marks consisting of or including emblems, designs, symbols, or devices were used. Sixty of these trade-marks were originals, and the other 60 their respective imitations, which had become the subject of litigation. The collection represented many varying degrees of deceptive similarity, from a slight resemblance to approximate identity. The judicial decisions involving these 60 pairs of trade-marks were not wholly confined to questions of deceptive similarity between the word trade-marks, and hence are not homogeneous. Some of the judicial decisions were confined simply to questions of the deceptive similarity in law of mere word trade-marks, while others were complicated by the presence of various other questions. About one half of the decisions are adjudications of infringement, restraining the use of the imitative trade-marks, while the other half are adjudications of non-infringement, permitting the use of the imitation. The results of the experiments with these 60 pairs of trade-marks are given in Chapters III. and IV.

A second selection of trade-marks was made from other than litigated cases. In this selection there are 180 trade-marks. They were chosen at random from advertisements in current magazines, trade-journals, newspapers, etc. For experimental purposes 60 of the trade-marks were selected from the 180 and duplicates of them were made. We shall refer to both members of each pair as duplicates or duplicate trade-marks. The remaining 120 trade-marks are called new trade-marks.

All the trade-marks used in the experiment were shown with the commercial names of the articles or commodities to which they are applied in commerce, thus Green River Whiskey. The articles represented many classes of goods of various descriptive properties, for example, soap, shoes, flour, incandescent lights, tobacco, and oil. In some of the court decisions, however, the litigated pairs of trade-marks are applied to different articles, whose commercial names are not identical, tho belonging to the same class of goods. The name of the article of the original trade-mark is identical with that of its respective imitation. In a few of the decisions with which Chapters III. and IV. are concerned, the names of the articles were not identical, but were made so in the experiment. Likewise, the name of the article in each pair of duplicate trade-marks is identical. In each set[1] the names of the articles of the original and imitative trade-marks are all different from the names of the articles of any of the duplicate or new trade-marks; and the names of the articles of the new trade-marks are all different from those of the duplicates.

2. Experimental Procedure

(*a*) *General Description of the Recognition Experiments.*—Two different psychological methods were used to measure confusion, the recognition method[2] and the order of merit or relative position method. In Chapter II. there was three variations of the recognition method, a different group of observers being used for each variation. In the Uninformed group, the observers recognized the trade-marks without knowing that the imitations were to appear; in the Informed group, the observers were aware of this; and in the Control group, the recognitive value of the originals themselves was studied. In each group, there were 20 observers, 10 of each sex, all university students. Thus these results are based on the mental processes of well educated persons. In Chapter V., where only the Uninformed group was used, 40 observers were included in the experiment. Thirty-four were men and 6 were women, representing 28 different occupations and professions. Their ages ran from 16 years to about 65.

The 60 originals and the 60 imitations, the 60 pairs of duplicates, and the remaining 120 trade-marks were distributed in six sets. Each set consisted of two parts, the presentation which included the trade-marks first shown the observer, and the test which included the trade-marks shown later, to be recognized by the observer. The presentation contained 20 trade-marks, 10 originals, and 10 old or duplicate trade-marks. The test, on the other hand, contained 40 trade-marks, 10 imitations of the 10 originals shown in the presentation, 10 duplicate trade-marks of the remaining 10 shown in the presentation, and 20 new trade-marks, *i. e.*, trade-marks that had not been shown in the presentation. There were then four classes

[1] The sets are described on pp. 26–27.

[2] For a brief account of the form of this test and of others see H. L. Hollingworth, "Vocational Psychology," 1916, 109–121.

of trade-marks employed,—the originals, the imitations, the duplicates, and the new. In all the six sets there were no other trade-marks that were identical with the 60 originals, the 60 imitations, the 60 pairs of duplicates, or the 120 new trade-marks. This was so only in the Uninformed and Informed experiments, but not in the Control experiment where the imitations were omitted and duplicates of the originals replaced them. Unless otherwise specified in this study the words trade-mark, original, imitation, duplicate, and new, alone or in connection with the word trade-mark, signify the trade-mark itself plus the name of its article.

The six sets of trade-marks were tried in one sitting, the presentation of each successive set following immediately after the test of the preceding one. The task was not laborious, and the entire experiment did not usually require over twenty-five minutes. Very few complaints were made in the introspections to indicate that the observer thought himself confused by having gone thru the previous sets. The order in which the six sets were given varied so that the possible effects of practice and fatigue might be equally distributed among them. At the close of the experiment, the question was asked "How did you react when the words were shown to you in the presentation and in the test?" It was found that in the presentation that the subject had as a rule been able only to read the words. In Chapter V. only one set of trade-marks was studied.

The trade-marks and the names of the articles were all typewritten in black ink on small slips of white paper, two and three quarters by four and one quarter inches in size. The name of the article appeared on the line directly beneath the trade-mark. Care was taken to keep the blackness of the ink the same for all the letters in the words. All the words appeared in the second horizontal quarter from the top of the slip of paper. The arrangement of the words on an original and its respective imitation slip is shown below.

Original	Imitation
Holeproof Hosiery	Knotair Hosiery

The following directions were given to each observer in the Uninformed and Informed groups:

"You are going to be shown, one at a time, on slips of paper, a number of ordinary word trade-marks and the names of their articles, as 'Uneeda Biscuit,' 'Garford Automobile,' and 'Standard Oil.' You are to read all the words on each slip of paper. Read them naturally as though you were reading an advertisement in a magazine or in a street car. Immediately after you have been shown the last slip, you will be given a second list (the test) and asked to pick out those slips that you have just seen in the presentation and those you have not seen. You will be asked further to sort the slips into seven piles, according to the degree of your confidence or certainty of recognition of the slips. There are three degrees of certainty for the slips that are recognized as seen, and three similar degrees for those that are recognized as not seen. The three degrees are 'absolutely certain,' 'reasonably certain,' and 'faint idea.' In the seventh are put the guesses or doubtful recognitions, those slips that cannot be recognized as seen or not seen.''

There were not many doubtful recognitions, as the observer was asked in each case to force a decision one way or the other.

The experimenter placed one slip on top of the other on the table before the observer. The slips in the presentation were shown at the uniform rate of one a second. The exposure gave the observer just time to read all the words on the slip. In the test he took his own time in recognizing the slips. As the slips in both the presentation and the test were thoroughly shuffled after each observer did the experiment, no slip was given undue prominence by its position in the list. The general technique of the recognition method, as described above, is similar to that employed by Strong, Hollingworth,[3] and others.

The confusion caused by the imitations could have been studied with the duplicate and new trade-marks altogether omitted from the experiment. They were included, however, to make the experimental conditions more comparable to situations in everyday life. The duplicates were included in the test to make part of its setting identical with that of the presentation; whereas the new trade-marks were included to introduce a difference. In everyday life, if an imitation is met with in a store, its setting, surroundings, or environment (composed partly of the same and other kinds of goods, and of store fixtures), are in part new, in part identical with, or similar to that in the store in which the original was first seen. Furthermore, the various circumstances under which the original is met with may contribute to magnify the difference between its setting and that in which the imitation appears at a purchase. It may have been that the original was seen in a magazine or newspaper advertisement, or on a card in a street, elevated, or subway car, or on some poster, or that it was orally recommended by a friend. No

[3] ''Characteristic Differences between Recall and Recognition,'' *Amer. J. of Psychol.,* 1913, 24, 535.

attempt, however, is made to reproduce in the experiments the multitudinous variations that occur in everyday life.

(*b*) *Uninformed Experiment.*—The observers in the Uninformed group were not told of the purpose of the experiment; nor that there would be any imitations in the test replacing some trade-marks shown in the presentation. In daily life the circumstances under which a trade-mark is first met with and later recognized often resemble conditions similar to those of incidental memory.[4] If the observer inquired whether all the trade-marks shows in the presentation were in the test, or whether there were imitations in the test replacing some trade-marks in the presentation, he was told that these questions could not be answered until after the experiment was over. Nothing was said or done by the experimenter to lead the observer to suspect that there were imitations in the test and that their originals were omitted. Yet, if the observer said "here is a slip on which the trade-mark is different, but the name of the article is the same as that which was seen the first time in the presentation," and that he did not know whether it was correct to say he had seen the slip, he was told that if all the words on the slip were exactly the same as those seen before he should say that he had seen it, if all the words on the slip were not exactly the same he should say that he had not seen it. Even those observers who showed by their remarks or behavior that they noticed a catch in the experiment and were consequently on the alert against being deceived, were nevertheless later often deceived.

(*c*) *Informed Experiment.*—In addition to the directions given in the Uninformed experiment the Informed observers were told of the purpose of the experiment. They were informed that they were going to be shown 20 trade-marks and the names of their articles in the presentation, and that in the test there would be 40, 10 identical with 10 shown in the presentation, 10 imitations, and 20 new. Besides picking out the slips that were the same as those in the presentation and those that were new, they were asked also to pick out those that were changed. Their knowledge and caution against imitations were of course much greater than in the case of the Uninformed group.

(*d*) *Control Experiment.*—The directions in the Control experiment were the same as in the Uninformed, but the formation of the test was different. The purpose of this experiment was to study the recognitive value of the originals and to determine how well the

[4] In this connection G. C. Myer's monograph gives some interesting results of the inaccuracy of knowledge of familiar objects and events, "A Study in Incidental Memory," *Arch. of Psychol.*, 1913, No. 26.

duplicates and new trade-marks could be recognized when no imitations appeared. In the Control there were no imitations in the test, but the 10 originals of the presentation reappeared in the test. The test then contained duplicates of 10 of the originals, the usual old or duplicate trade-marks, and the 20 new. All the duplicates and new trade-marks were the same as those in the Uninformed and Informed experiments.

(*e*) *Relative Position Experiment.*—This experiment measures confusion differently from the recognition experiments. It does not state "how many" individuals are confused by the imitations, but it does give a measure of their relative differences. The degree of confusion between any pair of trade-marks is then measured by its position in the list. There are two forms of relative position experiment, the *serial,* and the *group.*[5] Both are used in this study; the group form in Chapter IV., and the serial in Chapter V. The directions for the group form required the observer to arrange 50 pairs of trade-marks in an order according to their likelihood of confusion. This he was to do by considering the deceptive similarity in visual appearance, sound, and meaning of the trade-mark. There were 11 grades into which each pair could be sorted, the limiting ones being 0 representing absolute non-confusion and 10 absolute confusion.

The observer may assign any pair of trade-marks to any one of these grades. If he thinks for example that there is no possibility of confusion between the imitation "Black Diamond" and the original "Syphon" when applied to refrigerators he grades it 0; and if there seems to be absolute likelihood of confusion between another pair he grades it 10.[6] If the pair of trade-marks "Shipmate—Messmate" seems to be midway between absolute non-confusion and absolute confusion the observer grades it 5. Again, if a pair of trade-marks seems to have just the least likelihood of confusion it is put in pile 1. It is not believed that all the observers judged the trade-marks in the above manner.

[5] The group form is a modification of the serial, and either may be quickly transformed into the other. Hollingworth, who introduced the group form, finds that its results correlate closely with the serial. It has the advantages of being quicker, less fatiguing and monotonous, and permitting the observation of any changes in value of all the items, advantages which cannot be obtained by the serial. "Judgments of the Comic," *Psychol. Rev.,* 1911, 18, 135–136.

[6] There appears to be a discrepancy here in the fact that on the basis of pure chance alone one can distinguish an imitation that is identical with its original in 50 per cent. of the cases. It should also be said that both these limits cannot be held to be strictly absolute.

Each observer was given an envelope and a sheet of directions. Enclosed in the envelope were 50 slips of white paper, containing the 50 pairs of trade-marks and their articles. The directions were:

> "Please arrange the enclosed slips in an order of merit according to the following directions.
>
> "Each slip contains two trade-marks of a common article. The upper is the original trade-mark and the lower is the imitation of it.
>
> "The different pairs of trade-marks on the different slips vary in their likelihood of confusion, or in their deceptive similarity, or in the likelihood that the imitation trade-mark will be mistaken for the original. This confusion may be due to the appearance of the trade-marks, their sound, their meaning or significance, or any combination of these three factors.[7]
>
> "Arrange the slips according to their likelihood of confusion as follows: Put in the grade marked 0 those slips whose trade-marks would cause *absolute non-confusion.* Put in grade 1 those which would cause the least amount of confusion, in grade 2 those which would cause a little more confusion, in grade 3 those which would cause still more confusion, and so on up to grade 10 which represents *absolute confusion.* The intervals between the grades are all equal.
>
> "It is not required that an equal number of slips be put in each grade, nor that every grade be used.
>
> "In arranging the slips consider each trade-mark as if you have had no previous experience with it.
>
> "Finally, it must be remembered that the name of the article forms no part of the trade-mark."

In addition the observer was told that there was no time limit, and that he could take as long as he wished to make the arrangement. He was allowed to rearrange the order until it satisfied him. When the observer indicated that he was satisfied with his arrangement, he was asked to state how he did it.

Fifty college students, 25 men and 25 women, acted as judges. They were students of psychology at Columbia University during the fall of 1914. No observer in this experiment took part in the others.

By stating in the directions that the intervals between the grades are all equal the intention was to answer in advance any question in this regard, and to give the impression that they were not to be considered unequal. Yet it is not believed that the observer in grading always considered the differences between the grades equal, nor that he could always be so exact in making his judgments. Thus it cannot be maintained that the same amount of difference between the average grades of two pairs of trade-marks in two different regions of the scale stand for equal objective differences in deceptive similaritiy. In spite of all this, the inclusion of the above state-

[7] Another factor that might have been included is similarity in linguistic formation.

ment in the directions is perhaps justified because it may have acted as a mild persuasion to the observer to try to do this. It would tend to make the results more reliable.

Thirty-nine of the 50 pairs of trade-marks were litigated and were employed also in the recognition experiments. The remaining 11 pairs of trade-marks were uncontested and artificial, being coupled by the experimenter. Three of the 11 pairs were purposely made very dissimilar, 5 representing random ratings, and 3 identical matings. None of the 11 pairs were studied in the recognition experiments. The artificial matings were included among the litigated trade-marks for several reasons. We wanted to know what the grades of the dissimilar, random, and identical matings would be, and how they would compare with the grades of the litigated trade-marks. It was thought that the artificial matings would give a more clear and definite meaning to the limits of the series. In some cases they would tend to prevent a piling up at the limits, and in others to fill out the range between the limits.

All the trade-marks appeared with the names of the articles, the name of the article of the original trade-mark being identical with that of its imitation. With the exception of the dissimilar and random matings, the names of the articles were the same as those to which they were applied in commerce. In these cases the name of the article applied to the trade-mark in each pair was in fact used in connection with one of them in commerce.

In preparing the material the same precautions and care were taken as in the recognition experiments. On each of the 50 slips of white paper, two and three-quarters by four and one quarter inches in size, appeared the imitation and original trade-marks, each with the name of its article. The original trade-mark and the name of its article directly beneath appeared in the second horizontal quarter from the top of the slip; the imitation and the name of its article in the third horizontal quarter. The words were arranged on the slips as shown below.

Royal Irish Linen
Writing Paper

Royal Vellum
Writing Paper

The serial form of the relative position method used in Chapter V. differs from the group form in that it permits the trade-marks to be graded without any gaps between them in an unbroken continuum. In this experiments there were 9 pairs of litigated trade-marks picked at random from simple decisions. They were studied also by the recognition method. Their results test, as has been mentioned before, the accuracy of the judicial decisions. The departure from the instructions in the group form will be observed.

"Please arrange as well as you can the enclosed slips in an order according to the following instructions.

"Each slip contains two trade-marks of a common article. The upper is the original trade-mark and the lower is the imitation of it.

"The different pairs of trade-marks on the different slips vary in their likelihood of confusion, or in their deceptive similarity, or in the probability that the imitation trade-mark will be mistaken for the original. *This confusion may be due to the appearance of the trade-marks, their sound, their linguistic formation, their meaning or significance, or any combination of these four.*

"Arrange the slips serially according to their likelihood of confusion as follows: Put on the top of the pile the slip whose trade-marks show the greatest likelihood of confusion and on the bottom the slip whose trade-marks show the least likelihood of confusion. Between the top slip and the bottom slip put those slips whose trade-marks show intervening degrees of likelihood of confusion.

"In arranging the slips consider the trade-mark as if you have had no previous experience with it."

Forty college students, 20 men and 20 women, acted as judges. They were students of psychology at Columbia University in the spring of 1915. Their ages were fairly uniform, varying only within a few years of each other.

At the hands of Fechner, Galton, Cattell, Thorndike, Hollingworth, Strong, and others the relative position method has yielded various scientific and practical investigations of great importance.[8]

[8] For a more complete account of the technique of measurements by relative position and of mental scales see E. L. Thorndike, "Mental and Social Measurements," 1913, 7–26; R. S. Woodworth, mimeographed lecture notes on *Judgment*, 1917, 3–6, 16–20; and H. L. Hollingworth, The Method of Relative Position, "Psychological Researches of J. McK. Cattell," *Archives of Psychol.*, 1914, No. 30, 75–91.

CHAPTER III

RESULTS OF THE RECOGNITION EXPERIMENTS

1. Average Results

The first section of this chapter deals mainly with the psychological aspect of recognition, and the following sections with the practical side. We shall begin by examining the extent to which the imitations were confused when the observers were not told about them. It will be remembered that in each test there appeared 10 imitations, the originals having previously appeared in the presentation. Tho the imitative trade-mark itself was in every instance different from the original trade-mark, the name of the commodity was identical in both connections. In the Uninformed experiment a recognition of an imitation is scored correct if it is discriminated and placed in one of the "not seen" piles; in the Informed experiment it is scored correct if placed in one of the "changed" piles. In the Uninformed experiment an imitation is scored incorrect if identified and placed in one of the "seen" piles.[1] This score represents the percentage of confusion (or mistaken recognitions) between the originals and the imitations. In the Informed experiment an imitation is scored incorrect if it is discriminated and placed in one of the "new" piles, or if identified and placed in one of the "identical" piles. The identifications represent the percentage of confusion between the originals and the imitations. Then the "seen" pile in the Uninformed and the "identical" in the Informed are comparable in respect to confusion.

TABLE I.

The Average Per Cent. and Probable Error of Incorrect Identifications of the *Imitative* Trade-Marks

Uninformed Subjects

Trade-Marks Number	Kind	Av. Per Cent. Confused	P. E.
60	Imitative	44	*1.5*

[1] The discriminations in the Uninformed group represents three kinds of recognitions, (1) the imitations recognized as somewhat altered or changed, (2) a recognition of no similarity or difference in the imitation, and (3) the imitation recognized as entirely new, the original having been forgotten. In the Informed the discriminations represent only the two latter kinds of recognitions. In the results, however, these are all massed together.

(*a*) *Uninformed Experiment.*—Table I. presents the results of the imitative trade-marks in the case of the Uninformed group of subjects. The first column in the table states the number of trade-marks studied, and the second the kind of trade-mark, *i. e.*, whether imitative, duplicate, or new. The third column shows the average per cent. of incorrect identifications or the average per cent. of times the imitative trade-marks were confused with their respective originals. The fourth column gives the probable error of the average.

Forty-four per cent. of the imitative trade-marks were incorrectly identified, *i. e.*, were indicated as having been seen before in the presentation. The probable error of this average, 44 per cent., is 1.5, the chances being even that the true average (obtained from an infinite number of subjects) lies between 42.5 and 45.5, or outside it. The average of 44 per cent. has then a high degree of validity. If all the imitative trade-marks had been discriminated by every subject, or indicated as not having been seen before, all the recognitions would have been scored correct and the average per cent. of incorrect identifications would have been 0. But if, on the other hand, all the imitations had been confused with their respective originals, the average per cent. of incorrect identifications would have been 100. The difference between 44 per cent. and the greatest possible score of confusion 100 per cent. gives 56 per cent., the average per cent. of correct discriminations. The average per cent. 56 is 12 per cent. or about 1.3 times greater than the average per cent. 44 of incorrect identifications. In other words there are about one fourth more imitations correctly discriminated than confused.

Pure chance, involving no question of memory at all, would give about 50 per cent. correct and 50 per cent. incorrect recognitions. Altho the percentages of correct and incorrect recognitions of Table I. are 6 per cent. away from what pure chance would give, they should not be considered as influenced by chance to this large extent. That this is discredited is apparent from an examination of the individual scores of the imitative trade-marks;[2] those possessing greater similarity in form and meaning receiving higher scores of confusion, and vice versa. If chance were so greatly active in producing the average 44 per cent., the imitative trade-marks of high similarity would have about an equal number of high and low scores of confusion. The data from the duplicate and new trade-marks would likewise be affected so as to result in scores close to 50 per cent. correct and 50 per cent. incorrect recognitions. On the con-

[2] See Table IV., pp. 39–41.

trary, they both possess much higher scores.[3] Thus in the case of the imitations chance would not necessarily be the cause of a result of 50 per cent. correct and 50 per cent. incorrect recognitions. It is not denied that some of the individual scores were influenced by guessing and chance, but it is denied that the average 44 per cent. has been influenced to such an extent that it is only 6 per cent. different from what pure chance would give.

TABLE II

THE AVERAGE PER CENTS. AND PROBABLE ERRORS OF THE IMITATIVE TRADE-MARKS RECOGNIZED AS IDENTICAL, CHANGED AND NEW

Informed Subjects

Trade-Marks		Av. Per Cent. Recognized as					
Kumber	Kind	Identical	(P. E.)	Changed	(P. E.)	New	(P. E.)
60	Imitative	23	*1.4*	51	*1.4*	26	*1.0*

(*b*) *Informed Experiment.*—These observers were aware of the presence of the imitations, and had to recognize in addition to those trade-marks they had seen and had not seen, those that were changed. Table II. presents the recognitions of the imitative trade-marks in the Informed groups. The first column in the table states the number of trade-marks studied, and the second the kind of trade-mark. The third column shows the average per cent. of incorrect recognitions, identifying the imitations with the original trade-marks. The fifth column shows the average per cent. of correct recognitions, noting a change of the imitation from the original. The seventh column shows the average per cent. of incorrect discriminations of the imitations or those imitations recognized as new. Columns 4, 6, and 8 give respectively the probable errors of the averages in columns 3, 5, and 7.

Column 3 shows that on an average 23 per cent. were confused or incorrectly identified as having been seen exactly the same in the presentation. Column 5 shows that on an average 51 per cent. of the imitations were correctly recognized as changed from the form in which they appeared in the presentation.[4] Finally, column 7

[3] The duplicates have 73 per cent. correct identifications and the new trade-marks 93 per cent. correct discriminations.

[4] The phrase a "correct recognition of change" is used in connection with the imitations and originals to denote not a single absolutely correct kind of recognition, but rather a variety in recognitions of change as well as degrees of accuracy. In some cases a recognition of change may indicate that the original is completely recalled with the correct noting of all things changed in the imitation. In other cases it may represent just a bare feeling that the imitation is different from some trade-mark seen before in a presentation. In still other

gives the average per cent. (26) of incorrect discriminations, *i. e.*, of the failure to recognize the imitations as resembling some trade-marks seen in the previous presentation. The average per cent. (51) of correct recognitions of change is about 2.2 times greater than the average per cent. (23) of incorrect identifications, and about 2.0 times greater than the average per cent. (26) of incorrect discriminations, which is about 1.1 times greater than the average per cent. of incorrect identifications. The last difference is not large.

It might well be asked now which is the more typical method in everyday life and which is the better psychologically, the Uninformed or the Informed. In everyday life we find each of these methods and a combination of the two; the corresponding kinds of recognition described under our experimental conditions also occurring. The Informed method is perhaps from the psychological point of view the better, as it gives us more information concerning the subject's performance than the Uninformed. A court of equity would most likely favor an experiment conducted under the Uninformed conditions, as it generally considers that the ordinary purchaser in daily life to be similarly unwary.

TABLE III

THE AVERAGE PER CENT. AND PROBABLE ERROR OF CORRECT IDENTIFICATIONS OF THE *Original* TRADE-MARKS

Control Subjects

Trade-Marks Number	Kind	Av. Per Cent. of Correct Identifications	P. E.
60	Original	84	.9

(*c*) *Control Experiment.*—Whether there is or is not confusion in the preceding recognition experiments depends upon two factors, the degree of deceptive similarity of the imitation and the memory of the original. At first it was thought that by determining the recognitive value of the originals it would be possible to assign an exact value to each of these factors. It was seen later, however, that such an analysis could not be made. The Control experiment was not in fact a control, for we did not keep the deceptive similarity constant while measuring at the same time the memory of the

cases it may represent a noting of incorrect changes in the imitation, as a result of an incorrect recalling of the original, an incorrect comparison of the imitation and the original, or both. The recognitions of change described in the above cases with the numerous possible variations give one an idea of the different degrees of accuracy. The data at hand do not enable one, however, to distinguish these degrees. The recognitions are still more complicated by the rôle played by the name of the article in aiding or thwarting recognition.

original. Under the conditions observed with the experiments with the imitations both these factors usually cooperated. Tho our primary object failed, a few definite results were obtained. The Control did determine the degree to which the original could be identified under conditions comparable with those in the Uninformed and Informed experiments. In the two latter experiments the 10 originals appeared in the presentation, but their imitation replaced them in the test. No imitations appeared in the Control; the 10 originals appeared in the presentation and in the test.

Table III. presents the results of the original trade-marks obtained in the Control group. The first column states the number of trade-marks obtained in the Control group. The first column states the number of trade-marks studied, and the second the kind of trade-mark. The third column shows that the average per cent. of correct identifications is 84, and the fourth column that its probable error is 0.9. The difference (16) between per cents. 84 and 100 gives the average per cent. of incorrect discriminations, or the average per cent. of originals that were recognized as not seen in the presentation, where in fact they did appear. The average per cent. (84) of correct identifications is 68 per cent. or about 5.3 times greater than the average per cent. (16) of incorrect discriminations.

2. Results for the Separate Trade-Marks

This section deals with the individual scores of the originals and imitations, whose averages were reported in the preceding section. About half of the imitations were declared to be illegal and the other half legal. But as was mentioned not all the decisions represented by the trade-marks in this chapter and Chapter IV. were rendered merely on the question of deceptive similarity of the word trade-marks; in many of them other legal questions were taken into account. For this reason all the scores cannot be taken as a measure of the accuracy of the decisions. But we have mentioned them because they seem to throw light on other legal and psychological matters.

Table IV. presents the separate scores of the 60 original and imitative trade-marks for the Uninformed, Informed, and Control groups. The first column in the table shows whether the imitation in the decision was held to be an infringement (I) or a non-infringement (N). When the letters I and N are printed in ordinary type the decisions are generally of simple cases. But when the letters *I* and *N* are printed in italics the decisions are of complicated or questionable cases, involving other things besides the point of de-

TABLE IV

THE ORIGINAL TRADE-MARK, THE IMITATIVE, THE NAME OF THE ARTICLE, THE ORDER, AND THE RECOGNITION SCORES IN THE UNINFORMED, THE INFORMED, AND THE CONTROL GROUPS

Decision	Trade-Marks[1] Original	Imitative	Name of Article	Order	Uniformed: Per Cent. Confused	Informed: Per Cent. Recognized As: Identical	Changed	New	Control: Per Cent. of Correct Identification of Originals
I	Welcome	Welcome A. Smith	Soap	1	5		80	20	85
I	Our Little Samson	Samsoncalf	Shoes	2	10		50	50	75
N	Golden Charm	Charm	Flour	3	10	5	50	45	80
N	Walkeasy	Waulkwell	Shoes	4	10		100		80
N	Holeproof	Knotair	Hoisery	5	10	10	80	10	95
I	Rubberset	Rubber-vulc	Shaving Brush	6	20		75	25	80
N	Keepclean	Sta-Kleen	Toilet Brushes	7	20	5	70	25	90
I	Yusea	U-C-A	Incandescent Lights	8	20	10	70	20	80
N	Every Day	Everybody's	Soap	9	20	20	70	10	85
N	Union Leader	Union World	Tobacco	10	30	5	60	35	70
I	Kalamazoo Wagon	Kalamazoo Buggy	Company	11	30	10	60	30	95
N	No-To-Bac	Baco-Curo	Medicine	12	.30	15	55	30	100
N	Don Carlos	Don Caesar	Olives	13	30	15	60	25	75
N	Royal Irish Linen	Royal Vellum	Writing Paper	14	30	15	65	20	85
I	Uno	Ino	Medicine	15	30	20	60	20	100
I	Liveraid[1]	Liverine	Medicine	16	30	30	50	20	85
I	Beats-All	Knoxall	Lead Pencil	17	35		65	35	95
N	S. B.	B. & S.	Cough Drops	18	35	30	70		80

[1] Through error these words were spelled, and used in the experiments, slightly different from the way in which they occur in the court records: Liveraid should be Liveroid, Seafoam should be Sea Foam, and Mormaja should be Momaja.

TABLE IV (*Continued*)

	Trade-Marks			Groups: Uniformed		Groups: Informed — Per Cent. Recognized As:			Control
Decision	Original	Imitative	Name of Article	Order	Per Cent. Confused	Identical	Changed	New	Per Cent. of Correct Identification of Originals
N	Maraschino	Marceno	Candy	**19**	35	45	35	20	90
I	Pep-Kola	Pepko	Tonic	**20**	35	40	45	15	80
N	Ruberoid	RubberO	Roofing	**21**	40	10	60	30	90
I	Shipmate	Messmate	Galley Stove	**22**	40	15	50	35	65
N	Pratt's Astral	Standard White Astral	Oil	**23**	40	20	50	30	70
I	Worth	Our Worth	Edge Tools	**24.5**	40	20	55	25	85
N	Besteyette	Veribest	Raincoat	**24.5**	40	20	55	25	95
I	Sorosis	Sartoris	Shoes	**26**	45	5	70	25	90
N	Sozodont	Kalodont	Tooth Paste	**27**	45	5	80	15	85
I	Cyco	Cyco Prize	Carpet Sweeper	**28**	45	10	55	35	75
N	Ma-Le-Na	Man-a-lin	Medicine	**29**	45	15	60	25	95
N	Old Country	Our Country	Soap	**30**	45	20	40	40	75
I	Six Little	Six Big	Tailors	**31**	45	30	55	25	90
N	Dermacura	Dermakola	Skin Ointment	**32**	45	30	25	45	90
I	Maizena	Maizharina	Corn Flour	**33.5**	45	30	40	30	80
N	Bear Lithia Springs	Great Bear Springs	Company	**33.5**	45	30	40	30	85
I	Mellwood	Mill Wood	Whiskey	**35**	45	30	45	25	85
I	Seafoam[1]	Sodafoam	Baking Flour	**36**	45	35	50	15	60
N	Magic	Magico	Cleanser	**37**	45	40	30	30	60
I	Amber Bead	Amber	Beer	**38**	45	40	40	20	75
N	Victor	Victoria	Millinery	**39**	45	45	35	20	70
I	Capital	Capitol	Coffee	**40**	45	45	40	15	90

TABLE IV (*Concluded*)

Decision	Trade-Marks: Original	Trade-Marks: Imitative	Name of Article	Order	Uniformed: Per Cent. Confused	Informed: Per Cent. Recognized As: Identical	Informed: Changed	Informed: New	Control: Per Cent. of Correct Identification of Originals
N	Electric	Electric Light	Flour	**41**	50		70	30	75
N	Eagle	Gold Eagle	White Lead	**42**	50	10	45	45	70
I	Green River	Green Ribbon	Whiskey	**43**	50	15	50	35	95
I	Carbolineum	Creo-Carbolin	Preserving Paint	**44**	50	20	30	50	80
N	Social Register, Newport	Newport Social Index	Directory	**45**	50	20	40	40	80
I	German Sweet	Sweet German	Chocolate	**46**	50	25	50	25	95
N	Henderson	Anderson	Whiskey	**47**	55	15	55	30	85
I	Mormaja[1]	Mojava	Coffee	**48**	55	20	60	20	95
I	Nitro	Nitro-Hunter	Firearms	**49**	60	10	35	55	100
I	Grenadine	Grenade	Syrup	**50**	60	25	25	50	100
N	Muresco	Murafresco	Wall Covering	**51**	65	25	60	15	90
I	Trenton	Trenton Style	Pork Roll	**52**	65	55	10	35	85
I	Johnston's	Johnson's	Chocolates	**53**	65	75	15	10	75
I	Willoughby Lake	Willoughby Ridge	Scythe-Stones	**54**	70		50	50	70
N	West End	East Ridge	Distilling Co.	**55**	70	20	55	25	95
I	Cottolene	Cottoleo	Substitute for Lard	**56**	70	40	35	25	80
I	Dyspepticure	Dyspepticide	Medicine	**57**	75	35	60	5	95
I	Ceresota	Cressota	Flour	**58**	80	25	65	10	85
N	Siphon	Siphon System	Refrigerator	**59**	80	50	30	20	85
I	Nubia	Nubias	Cigarettes	**60**	85	40	30	30	90

ceptive similarity of the word trade-marks, and hence are not suitable for testing the accuracy of the decisions. The second column in the table shows the original trade-mark, the third the imitative, and the fourth the name of the article which appeared with both the original and imitative trade-marks. Column 5 presents the trade-marks in the order of confusion, from least to most. Column 6 gives the percentage of individuals in the Uninformed group that confused the imitative trade-mark with the original. The seventh column gives the percentage of individuals in the Informed that confused the imitative trade-mark with the original; the eighth that recognized the imitation as changed; and the ninth that discriminated it. The tenth gives the percentage of correct identifications of the original trade-marks in the Control group.

The order of confusion in the fifth column is determined, first, by the amount of confusion shown in the Uninformed, and secondly, in case two or more pairs of trade-marks have equal scores, by the amount of confusion shown in the Informed. To obtain in the Informed a complete estimate of the amount of deceptive similarity it is necessary to take into account the three kinds of recognition. This was done by taking as the most confusing the pair of trade-marks that had most incorrect identifications; but if the incorrect identifications were equal here also, the pair which had most correct recognitions of change or the least discriminations was then taken as the most confusing. As the equating of these three values of deceptive similarity is not absolutely exact the second arrangement is a little rough and arbitrary. Nevertheless, the method in the main is sound. There are just two cases in which two pairs of trade-marks have identical scores in the Uninformed and in the Informed. The pairs of trade-marks in which the original trade-marks had a higher per cent. of correct identifications were put above the others.

One of the first things to be noticed about the table is that all imitations, non-infringements as well as infringements, cause some confusion in the Uninformed group. Furthermore, both in the Uninformed and Informed they do not divide into two groups, but extend over about three fourths of the entire length of the scale, forming a continuum. These two findings do not support the legal treatment of imitations. Indeed, they present clear proof against the statements made in trade-mark laws, judicial decisions, and legal textbooks which consider imitations as falling into only two discrete groups, those likely to deceive and those not likely to deceive. For legislative bodies and courts to think and work on false principles and assumptions in regard to imitations has surely a detrimental effect not only on their own proceedings but also on business and commerce. Confusion, tho it be a subjective fact, is also a quantitative one. To handle it correctly it needs to be measured, not merely defined. In ordinary conversation we are usually satisfied in remarking that the weather is hot, warm, cool, etc. But in scientific and industrial laboratories these adjectives become exceedingly vague and are replaced by degrees on the thermometer. By standardizing our notions about degrees of heat, the thermometer permits of greater accuracy in working with them. Clear and quantitative meanings attached to the legal and illegal categories of deception would undoubtedly favor greater accuracy in handling them too.

Let us now see what the highest and lowest scores are for each kind of recognition. In the Uninformed "Nubia-Nubias" is the most confusing pair of trade-marks, deceiving 17 out of 20 indiviudals or 85 per cent. of the group. The least confusing pair is "Welcome—Welcome A. Smith" with a score of 5 per cent., deceiving only 1 out of 20 individuals. Column 7 shows that in the Informed group seven pairs of trade-marks have 0 per cent. scores of confusion. In the Uninformed one of these, "Welcome—Welcome A. Smith," has 5 per cent. of confusion and another, "Willoughby Lake—Willoughby Ridge," has 70 per cent. For the 0 per cent. scores in the Informed this is the largest difference between the confusion of any two in the Uninformed. "Johnston's—Johnson's" with 75 per cent. of confusion is the most confusing imitation in the Informed, it being eighth from the top in the Uninformed where its score is 65 per cent. This pair of trade-marks is one of the three which received a higher score of confusion in the Informed than in the Uninformed; these being the only cases where the natural results of more confusion in the Uninformed than in the Informed does not obtain. As the difference is not large in any of the three cases it is quite possible that further experimentation would reverse the advantage. It will be noticed that the scores of confusion in the Informed tend in general to increase with their corresponding ones in the Uninformed. In 52 cases out of 60 the confusion in the Uninformed is higher than in the Informed; and in 5 cases they are equal.

In the column for correct recognitions of change "Walkeasy—Waulkwell" has a 100 per cent. or perfect score, and "Trenton—Trenton Style" with 10 per cent. stands at the lower limit. The former in the Uninformed confused 10 per cent., and in the Informed 0 per cent.; the latter in the Uninformed confused 65 per cent., and in the Informed 55 per cent. An imitation that is easily detected is apt to cause little confusion, and vice versa. In the column for incorrect discriminations or those imitations recognized as new the highest score is 55 per cent., and the two lowest are each 0 per cent. Two scores at 60 per cent. are the lowest for correct identifications of the originals, and 4 perfect ones are the highest. The identifications of the originals are higher in correctness than any other kind of recognition. A comparison of the correct identifications of the originals with the incorrect identifications of the imitations shows that in all but two instances the originals were more often identified as originals than imitations are identified as originals. ''Johnston's—Johnson's'' in the Informed with 75 per cent. and ''Willoughby Lake—Willoughby Ridge'' in the Uninformed with 70 per cent. are the exceptions. These scores indicate that the imitations look just like the originals and under the experimental conditions are not distinguishable. Each of the 3 most confusing pairs of trade-marks in the Uninformed differ in percentage of identifications from that of their respective originals by only 5 per cent. We may see from these cases the extent to which an imitation may displace its original.

The scoring permitted a difference of no less than 5 per cent. The range in percentage of the various kinds of recognitions vary from 40 per cent. of the entire length of the scale (from 60 per cent. to 100 per cent. of correct identifications of the originals) to 90 per cent. of the entire length of the scale (from 10 per to 100 per cent. of correct recognitions of change of the imitations); or from four tenths of the entire range to nine tenths. In the different ranges there are only a few gaps between any two pairs of trade-marks, and those not very wide. If we had experimented further, or had used more trade-marks of other degrees of deceptive similarity, there is no doubt that all the missing steps would have been filled in. The frequencies of the different percentages of the various recognitions in the 3 groups tend, when plotted, to resemble the normal probability curve, excepting the incorrect identifications in the Informed which are skewed toward the high end.

3. Scale of Confusion in Recognition

Mental measurements and scales have been the subject of very fruitful inquiry in psychology, constituting also one of its most practical branches. In recent years mental scales or standard tests have shown a rapid development and a wide range of practical application. They have been treated from various aspects by many psychologists. Galton, Cattell, Binet and Simon, Terman, Thorndike, Yerkes, and others have done much to place the subject on a scientific basis. There are now at least 29 mental scales in use, measuring such things as English, reading, spelling, handwriting, drawing, mathematics, teachers' efficiency, intelligence, and eminence. The chief value of these scales lies in their possessing greater accuracy in measuring mental traits than the more usual and more subjective methods.[5] A psychological scale for the measurement of the amount of deceptive similarity between two trade-marks would also be a more accurate method for the determination of lawful and unlawful imitations than the present judicial procedure, but such a scale has not yet been made. Even if it were, it is likely that slowly progressive judicial opinion would not be quite ready to accept it. Nevertheless, it might be worth while to show by an example the manner of construction and the mode of application of such a scale.

SCALE I

Scale of Visual Recognitive Confusion of Trade-Marks

Originals	Imitations	Confusion Order	Confusion P. C.
Welcome	Welcome A. Smith	1	5
Golden Charm	Charm	2	10
Yusea	U-C-A	3	20
Royal Irish Linen	Royal Vellum	4	30
Beats-All	Knoxall	5	35
Shipmate	Messmate	6	40
Six Little	Six Big	7	45
Carbolineum	Creo-Carbolin	8	50
Mormaja	Mojava	9	55
Grenadine	Grenade	10	60
Muresco	Murafresco	11	65
Cottolene	Cottoleo	12	70
Dyspepticure	Dyspepticide	13	75
Siphon	Siphon System	14	80
Nubia	Nubias	15	85

Professor Woodworth has said that "an order series assumes to some degree the character of a graduated scale."[6] The sample scale may then be derived from Table IV. Scale I. contains 15 pairs of

[5] For further information concerning the aims and scope of mental tests consult, "Report of the Committee on the Academic Status of Psychology; A Survey of Psychological Investigations with Reference to Differentiations between Psychological Experiments and Mental Tests," *Amer. Psychol. Assn.*, December, 1916.

[6] *Op. cit.*, p. 4.

trade-marks, representing every amount of confusion in the Uninformed group. Column I. gives the originals, 2 the imitations, 3 the order of confusion from least to most, and 4 the per cent. of confusion. With two exceptions the distances between the steps are equal. For a number of reasons, some of which have been mentioned, the scale is not one that can be adopted by the courts; but for the purpose of illustration, let us assume that it is ready for use. Later on, the characterisics of a suitable scale will be considered. Let us assume that the trade-marks "Welcome—Welcome A. Smith" to "Royal Irish Linen—Royal Vellum" inclusive are non-infringements, and that those from "Beats-All—Knoxall" to "Nubia—Nubias" are infringements. In order to set the limits of legal and illegal amounts of confusion it may be imagined that the legislature or the courts decided that under the conditions of these experiments an imitation causing over 30 per cent. confusion is illegal and one causing 30 per cent. or less is legal. To investigate the question of infringement or non-infringement in a new case in court, it would be necessary by psychological experiment to find the amount of confusion caused by the imitation in respect to its original, and then to compare this amount with the limit of infringement on the scale. This may be done in two ways.

One way would be to test the confusion between the trade-marks in a new case under the same conditions as were those of the scale. Let us suppose that "Walkeasy" is the original trade-mark in the new case and "Waulkwell" the imitation; and that 10 per cent. of the individuals confused the latter with the former, as they really do when used in connection with the name of the article.[7] As 10 per cent. of confusion falls within the bounds of the legally allowable amount of confusion, "Waulkwell" is a non-infringement and is permitted to exist. If, on the other hand, "Ceresota—Cressota," the conflicting trade-marks in another new case, give 80 per cent. confusion, as they do,[7] the confusion caused by the imitation "Cressota" falls within the bounds of the illegal amount of confusion. It is declared an infringement and its use is restrained. Thus is rendered a psychological decision following an objective and scientific method, and omitting none of the requirements and technicalities of the law.

In the second way of using the scale a number of individuals are asked to match independently as closely as possible the new pair to some pair on the scale, and to assign to it the corresponding per cent. of confusion. From the average per cent. of confusion of the

[7] See Table IV., pp. 39–41.

gradings thus obtained, it may be determined whether the imitation is an infringement or a non-infringement, according as the amount of confusion lies above or below the limit of infringement. The second method may require less time than the first, but as the rating is less objective than the recognition procedure and different from the actual conditions giving the scale, it may not be as accurate. Both methods are, however, much superior to the judicial procedure.

A psychological scale of confusion for measuring the deceptive similarity between trade-marks is more complicated than would at first be supposed. In the first place the scale should be made up of pairs of word trade-marks that were litigated in the highest courts. Some should be infringements and some non-infringements. It is important that they should be selected from only simple cases in which the decision of infringement or non-infringement clearly appears to be colored by nothing else than the likelihood of confusion of just the two word trade-marks. More complicated cases in which other points aided in deciding this question should not be included. The pairs of trade-marks from the simple cases represent then decisions that are both legally and psychologically homogeneous.

The recognition method with the subjects uninformed would fulfill best the legal and psychological requirements. It is better to experiment with the trade-marks alone, not in connection with the names of their articles. The addition of the latter in the experiment changes not only the absolute but also the relative amounts of confusion of the trade-marks. Many more degrees of confusion should be included in the scale and should spread over the whole range, *e. g.*, from 0 per cent. to 100 per cent. of confusion. The difference between the degrees should be equal or very nearly so. The subjects should come from many walks of life. Several hundred subjects should at least be tested. The questions of infringement and non-infringement must be fixed in terms of amounts of confusion. This may be done in two ways. According to the one the legislature or the courts may decide these amounts under certain experimental conditions. Either a point or a space on the scale may divide the legal from the illegal amounts of confusion. According to the other the standards of infringement and non-infringement may be fixed by two averages with their probable errors, one representing the average amount of confusion of a number of infringements, and the other of a number of non-infringements. Those trade-marks composing the scale should be so selected after experimentation that only infringements should lie in the illegal and non-infringements in the legal limits.

The proposal to give a precise meaning to the term infringement would be likely to meet with opposition. In his chapter on Infringement Hopkins writes:[8]

> "In conclusion it is important to bear in mind that courts of equity have always avoided laying down any hard and fast rules by which to determine what constitutes fraud. The reason for this absence of set rules has been well stated as follows: 'Were courts of equity to once declare rules prescribing limitations with their power of dealing with it, the jurisdiction would be perpetually cramped and eluded by new schemes which the fertility of man's invention would contrive.'"[9]

It would certainly be imprudent to ignore this legal maxim, but it would be even more so not to see how the psychological methods do actually work.

The scale we have constructed on the basis of confusion is not an absolute scale, *i. e.*, it does not start from zero or "just not any of" confusion. In this respect it is inferior to the scales in the physical sciences. The advantage in having an absolute scale lies in being able to make a "time as ——" comparison. Four pounds is twice as heavy as 2 pounds, and 3 feet is one third as long as 9 feet. But we cannot say that 50 per cent. of confusion is twice as confusing as 25 per cent. of confusion, because we do not know that the 50 per cent. is twice as far from 0 per cent. as 25 per cent. is. Even if the scale contained a pair of trade-marks which gave 0 per cent. of confusion among 1,000 individuals, testing a few more thousand individuals might show some evidence of confusion. However, just a little confusion might be due to chance. Other things being properly accounted for, an absolute scale would be preferable, as it would be more reliable. The larger the number of individuals that the percentage scale of confusion represented the more it would tend to be absolute.

4. Comparison of the Results of the Originals and Imitations

This section treats of the relations to each other of the amounts of the various kinds of recognition of the originals and imitations in the Uninformed, Informed, and Control groups. The first column in Table V. gives the number of pairs of the originals and imitative trade-marks that entered into the calculations of the succeeding columns. Column 2 indicates the limits of the amounts of confusion of the trade-marks in the Uninformed, and column 3 the

[8] "The Law of Trademarks, Tradenames, and Unfair Competition," 1905, 303.

[9] Weinstock, Lubin & Co. v. Marks, 109 Cal. 529–539.

difference between these limits. Column 4 shows the average per cent. of confusion in the Uninformed, and column 5 the probable error of the average. Column 6 shows the average per cent. of incorrect identifications or confusions of these trade-marks in the Informed. The remaining columns up to 12 are self-explanatory. Column 12, however, gives the correct identifications of the originals in the Control, the imitations not figuring in this group. All the percentages of the preceding Table IV. are condensed in Table V.

TABLE V

THE RECOGNITION SCORES AND THEIR PROBABLE ERRORS IN THE UNINFORMED, THE INFORMED AND THE CONTROL GROUPS FOR DIFFERENT TRADE-MARKS OF VARIOUS AMOUNTS OF CONFUSION

	Uniformed Group				Informed Group						Control Group	
					Av. Per Cent. Recognized As:							
No. of Pairs of Trade-Marks	Range of Per Cent. of Confusion	Difference between Extremes of Range[10]	Av. Per Cent. Confused	P. E.	Identical	P. E.	Changed	P. E.	New	P. E.	Av. Per Cent. of Correct Iden. of Originals	P. E.
9	5–20	15	**14**	*1.5*	**6**	*1.5*	**72**	*3.0*	**23**	*3.4*	**83**	*1.3*
16	30–40	10	**34**	*.8*	**19**	*1.8*	**56**	*1.4*	**25**	*1.4*	**74**	*2.8*
21	45–50	5	**46**	*.3*	**23**	*2.0*	**47**	*2.0*	**30**	*1.5*	**81**	*1.6*
7	55–65	10	**61**	*1.2*	**32**	*5.9*	**37**	*5.7*	**31**	*4.4*	**90**	*2.3*
7	70–85	15	**76**	*1.6*	**30**	*4.1*	**46**	*4.0*	**24**	*3.2*	**86**	*2.1*

The first row of figures reads as follows: The first 9 pairs of trade-marks or the lowest 9 in the percentage of confusion in the Uninformed range in percentage of confusion from 5 per cent. to 20 per cent., a difference of 15 per cent.; the average percentage of confusion of these trade-marks is 14 with a probable error of 1.5. In the Informed group the average percentage of confusion for these same 9 pairs of trade-marks is 6 with a probable error of 1.5 etc. The second row gives the average results of the next 16 pairs of trade-marks of higher confusion in Table IV.

Column 4 and 6 show that an increase of confusion in the Uninformed is accomplished with an increase of confusion in the Informed, with the exception of the two highest degrees in the Informed; along the entire range confusion in the Informed is on an average about one half that in the Uninformed. The widest departure from this is with the trade-marks of the highest degree of

[10] The differences in this column could have been made equal instead of unequal had there been a sufficient number of trade-marks to give reliable averages at every step.

confusion, the average per cent. of confusion (30) in the Informed being about 39 per cent. of the average per cent. of confusion (76) in the Uninformed. The difference between 30 per cent. and 32 per cent. is not large enough to be reliable as it is covered by the probable errors or is within the limits of chance.

Column 8 indicates in general also the natural result. In other words the lower the degree of deceptive similarity or confusion between two trade-marks the greater is the ability to recognize the change. This statement is true with the exception of the two lowest scores (which correspond to the same trade-marks as in the exception in the previous paragraph). The difference between 37 per cent. and 46 per cent. is covered by the probable errors. With trade-marks of the lowest degrees of deceptive similarity the average per cent. of correct recognitions of change (72) is about 5.1 times greater than the score of confusion in the Uninformed and 12 times it in the Informed. The other degrees show less difference; when the average per cent. of confusion in the Uninformed is 46, the average per cent. of correct recognitions of change in the Informed is 47. With the highest degrees of deceptive similarity the average per cent. of confusion in the Uninformed (76) is 30 per cent. more than the score of correct recognitions of change (46).

Column 10 shows the average per cent. of trade-marks recognized as new. It is seen that the scores do not fall into serial order, but follow a haphazard arrangement. The significance of this is that these scores are, in general, no indication of confusion. They are therefore based on recognitions that are determined either by a failure to note any resemblance or difference between the original and imitation, or by the forgetting of the original. It will be noticed that the corresponding scores of the 2 groups of trade-marks that were exceptions above are not this time reversed; but their amounts are not in their proper relative positions; the difference between them is covered by the probable errors.

Column 12 shows that the amounts of correct identifications of the originals in the Control do not fall into serial order. They are about the same whether the imitations cause much, moderate, or little confusion. In other words those originals that were most often displaced by their imitations were just as well recognized as were those that were least often displaced. Therefore, these differences in amounts of confusion are chiefly due to the degree of deceptive similarity between the originals and imitations and not to a difference of memory of the original. In the last row the average percentage of confusion of the imitations in the Uninformed is 2 per cent. higher than the average per cent. (74) of correct identifications of the originals in row 2. Tho the difference is covered by the probable errors it is seen that some imitations on the average resemble their originals as much as some originals are identified with themselves.

5. Introspective Notes

At the close of the experiment the following question was put to each observer. "How did you react when the words were shown to you in the presentation and in the test?" The observers stated generally that in the presentation they had time to do no more than read the words. Some, however, articulated the words and others made associations with them. In the Informed experiment one observer made associations with the initial letters, another noted particularly the ends of the words, and in several cases imagery was noted as being present in the presentation and test. A few observers articulated also in the test. The observers in the Uninformed experiment often made casual remarks while making their recognitions in the test, and so gave voluntary and uncalled for introspections about the effects of the imitations on them. Many said that there were not the same number of slips in the test as in the presentation. One said, "I notice slight changes;" a second, "Some names are mixed up;" a third, "Do you have the same product with different trademarks;" and a fourth, "Do you mean exactly the same?" Still another said, "I saw 'Six Little Tailors' the first time, but this is 'Six Big Tailors,' are they meant to be the same or did you make a mistake?" Sometimes before recognizing an imitation it would be set aside to see whether its supposed original would turn up later. In a few cases when the observer became convinced of the presence of the imitations, he stated that in the beginning of the experiment an imitation was incorrectly identified.

While some were apparently in doubt as to the presence of the imitations and before they had decided upon it, I often caught them trying to read the answer from my facial expression or looking there for a clue. Furthermore, some of the observers gave clear indication by their curt expressions and behavior that they had noticed a trick in the imitations. Altho some became aware of the imitations after the first one met, others were not conscious that there had been imitations present even at the end of the experiment.

6. Summary

1. When the individuals are *unaware* of the presence of imitations they confuse on the average 44 per cent. of the imitations with the originals, and correctly discriminate 56 per cent.

2. When the individuals are *aware* of the presence of imitations they confuse on the average 23 per cent. of the imitations with the originals, correctly discriminate 51 per cent., and fail to recognize 26 per cent. of the imitations as resembling the originals.

3. The correct identifications of the duplicates of the originals are about 5.3 times as numerous as the incorrect discriminations.

4. Every imitation, non-infringements as well as infringements, cause some confusion among the individuals *unaware* of their presence.

5. The scores of confusion of the imitations do not divide into two distinct groups, those likely to deceive and those not likely to deceive, but their distribution forms a continuum. The findings in the last two statements discredit the present legal treatment of imitations.

6. The construction and application of a psychological scale for the measurement of deceptive similarity of two trade-marks would be the most scientific method of determining the question of infringement.

7. The lower the degree of deceptive similarity of two trademarks the less is the confusion and the greater is the ability to recognize the change.

8. The originals that are most often displayed by their imitations are just as well identified as those that are least often displaced.

9. The individuals who were *aware* of the presence of the imitations at the beginning of the experiment were affected in a variety of ways by them. Some noted their presence after meeting a few, and others did not even at the end of the experiment.

CHAPTER IV

RESULTS OF THE RELATIVE POSITION EXPERIMENT

1. Results for the Separate Trade-Marks

In this experiment a number of individuals judged the deceptive similarity between the trade-marks with the object of ascertaining the relative differences, not the exact amount of confusion. "The ability to perceive degrees of difference and to arrange objects in an ordered series," Professor Woodworth writes,[1] "is a fundamental and significant fact in psychology." While an individual may be unable to state the correct reasons for his judgments or to completely analyze the similarities and differences between the trade-marks, yet his judgments are significant. In the words of William James: "In ethical, psychological, and esthetic matters, to give a clear reason for one's judgment is universally recognized as a mark of rare genius."[2] The single judgment of any individual is a subjective and variable fact, so that an individual's judgment may not be the same at one time as at the next time, nor like that of his neighbor's. In order therefore to obtain reliable results in our experiments the relative position of each pair of trade-marks is determined by the average of the independent gradings by 50 individuals. In carrying out this experiment the test material is not presented to the subject in the same manner as in the recognition experiments. In the two kinds of experiments some of the mental processes involved, and the reactions of the subject, are different. But for all that there are certain resemblances in the mental processes, for recognition implies some judgment, and judgment implies some recognition.[3] We must not now, however, dwell on this subject for our interest lies chiefly with the objective results and their comparison, rather than on the psychological principles underlying the methods.

Table VI. presents the relative degrees of confusion of the trade-marks in the relative position experiment and the scores corresponding to them in the recognition experiments. The fifth, sixth, and seventh columns give the results of the former. The fifth gives the

[1] *Op. cit.*, p. 3.

[2] "Principles of Psychology," 1890, II., 365.

[3] R. S. Woodworth, mimeographed lecture notes on *Perception*, 1917, p. 1.

TABLE VI

THE ORIGINAL TRADE-MARK, THE IMITATIVE, THE NAME OF THE ARTICLE, THE ORDER, THE AVERAGE GRADE OF CONFUSION, ITS PROBABLE ERROR AND THE RECOGNITION SCORES

Decision	Trade-Marks: Original	Trade-Marks: Imitative	Name of Article		Relative Position: Order	Relative Position: Grade	Relative Position: P. E.	Recognition Groups: Uniformed: Per Cent. Confused	Recognition Groups: Informed: Per Cent. Recognized As: Identical	Informed: Changed	Informed: New	Control: Per Cent. of Correct Iden. of Originals
	Syphon	Black Diamond	Refrigerator	D[1]	1	.1	.02					
	Fits-U	Ariosa	Eyeglasses	R[2]	2	.2	.04					
	O-Zell	Lustro	Gelatine	D	3	.2	.04					
	Hammermill Bond	Globe-Wernicke	Paper	R	4	.2	.04					
	Globe Trotter	Certain-teed	Automobile	D	5	.3	.06					
	Everstick	Herringbone	Rubbers	R	6	.7	.12					
	Hensoldt	Buckhead	Prism Binocular	R	7	.7	.11					
	White Rock	Stonetex	Table Water	R	8	1.	.11					
N	Holeproof	Knotair	Hosiery		9	2.	.21	10	10	80	10	95
N	Royal Irish Linen	Royal Vellum	Writing Paper		10	2.4	.20	30	15	65	20	85
I	Our Little Samson	Samsoncalf	Shoes		11	3.5	.26	10	0	50	50	75
I	Rubberset	Rubber-vulc	Brush		12	3.5	.22	20	0	75	25	80
N	Sozodont	Kalodont	Tooth Paste		13	3.5	.24	45	5	80	15	85
I	Beats-All	Knoxall	Lead Pencil		14	4.1	.27	35	0	65	35	95
N	Union Leader	Union World	Tobacco		15	4.2	.24	30	5	60	35	70
I	Six Little	Six Big	Tailors		16	4.3	.26	45	20	55	25	90
N	Bestyette	Veribest	Raincoat		17	4.4	.26	40	20	55	25	95
N	Bear Lithia Springs	Great Bear Springs	Company		18	4.5	.22	45	30	40	30	85
N	Walkeasy	Waulkwell	Shoes		19	4.6	.26	10	0	100	0	80

[1] D = very dissimilar mating. [2] R = random mating. [3] Id = Identical mating.

TABLE VI (*Continued*)

Decision	Trade-Marks: Original	Trade-Marks: Imitative	Name of Article	Relative Position: Order	Relative Position: Grade	Relative Position: P. E.	Recognition Groups, Uninformed: Per Cent. Confused	Recognition Groups, Informed: Per Cent. Recognized As: Identical	Changed	New	Control: Per Cent. of Correct Iden. of Originals
I	Sorois	Sartoris	Shoes	20	4.7	*.28*	45	5	70	25	90
N	Ma-Le-Na	Man-a-lin	Medicine	21	4.7	*.28*	45	15	60	25	90
I	Seafoam	Sodafoam	Baking Powder	22	4.7	*.22*	45	35	50	15	60
I	Carbolineum	Creo-Carbolin	Preserving Paint	23	4.7	*.21*	50	20	30	50	80
I	Shipmate	Messmate	Galley Stove	24	5.0	*.25*	40	15	50	35	65
I	Green River	Green Ribbon	Whiskey	25	5.3	*.24*	50	15	50	35	95
I	Nitro	Nitro-Hunter	Firearms	26	5.3	*.21*	60	10	35	55	100
N	Social Register, Newport	Newport Social Index	Directory	27	5.5	*.28*	50	20	40	40	80
I	Yusea	U-C-A	Incandescent Lights	28	5.8	*.35*	20	10	70	20	80
I	Maizena	Maizharina	Corn Flour	29	5.8	*.22*	45	30	40	30	80
I	Pep-Kola	Pepko	Tonic	30	6.2	*.18*	35	40	45	15	80
N	Electric	Electric Light	Flour	31	6.2	*.21*	50	0	70	30	75
I	Amber Bead	Amber	Beer	32	6.3	*.22*	45	40	40	20	75
N	Muresco	Murafresco	Wall Covering	33	6.3	*.18*	65	25	60	15	90
I	Trenton	Trenton Style	Pork Roll	34	6.4	*.22*	65	55	10	35	85
N	Old Country	Our Country	Soap	35	6.5	*.24*	45	20	40	40	75
I	Dyspepticure	Dyspepticide	Medicine	36	6.6	*.23*	75	35	60	5	95
I	Cyco	Cyco Prize	Carpet Sweeper	37	6.7	*.21*	45	10	55	35	75
I	Uno	Ino	Medicine	38	7.0	*.19*	30	20	60	20	100
N	S. B.	B. & S.	Cough Drops	39	7.1	*.22*	35	30	70	0	80
I	Cottolene	Cottoleo	Substitute for Lard	40	7.1	*.21*	70	40	35	25	80
I	Worth	Our Worth	Edge Tools	41	7.3	*.18*	40	20	55	25	85
N	Magic	Magico	Cleanser	42	7.6	*.19*	45	40	30	30	60

TABLE VI (*Concluded*)

Decision	Trade-Marks: Original	Trade-Marks: Imitative	Name of Article		Relative Position: Order	Relative Position: Grade	Relative Position: P. E.	Recognition Groups: Uninformed: Per Cent. Confused	Recognition Groups: Informed: Per Cent. Recognized As: Identical	Recognition Groups: Informed: Per Cent. Recognized As: Changed	Recognition Groups: Informed: Per Cent. Recognized As: New	Recognition Groups: Control: Per Cent. of Correct Iden. of Originals
I	Ceresota	Cressota	Flour		**43**	7.9	*.15*	80	25	65	10	85
I	German Sweet	Sweet German	Chocolate		**44**	8.0	*.14*	50	25	50	25	95
I	Nubia	Nubias	Cigarettes		**45**	8.3	*.15*	85	40	30	30	90
I	Capital	Capitol	Coffee		**46**	8.8	*.11*	45	45	40	15	90
I	Johnston's	Johnson's	Chocolates		**47**	8.9	*.09*	65	75	15	10	75
	Drinket	Drinket	Coffee	Id[3]	**48**	10.	*.00*					
	Quickwood	Quickwood	Collars	Id	**49**	10.	*.00*					
	Whiz	Whiz	Stove Polish	Id	**50**	10.	*.00*					

order of confusion from least to most as determined by the average grades, the sixth gives the average grade of confusion, and the seventh the probable error of the grade. Columns 8–12 give the results of these trade-marks in the recognition experiments. Where the grades are equal their order is determined by the amounts of confusion in recognition. When they are arranged in an ordered series two pairs of trade-marks may differ slightly or greatly in confusion. The grades and their probable errors indicate such inequalities in the spacing of the series. When the grades of two pairs of trade-marks differ little, we may conclude that their amounts of confusion really differ little, and vice versa. The variability of the grade as shown by the probable error permits us to draw a similar conclusion. The probable error of a measure states the unreliability of the measure, or the probable approximation of the true measure (calculated from an infinite number of cases) to the obtained measure (calculated from 50 cases in this experiment). The probable error of the average shows within what limits the chances are even that the obtained average is correct. The probable error 0.25 of the grade 5.0 of "Shipmate—Messmate" indicates that the chances are even that the true average lies between 4.75 and 5.25, or outside it. The chances of true average being far outside this range decreases very rapidly. When the probable error is large, the true position of the pair of trade-marks is not well established, the reason for this probably being the small difference in confusion between this pair of trade-marks and those adjacent to it in the series. For a small difference is more likely to be misjudged than a large difference. Thus the larger probable errors indicate that the order is less certain, and the difference between one pair of trade-marks and the next on the list is less. The probable errors in Table VI. are generally small showing that the order is well established. Ordinary arithmetic does not apply to measures by relative position. We cannot say that grade 7.0 is twice as high as grade 3.5, or that the second grade from the highest plus the fifth is equal to the third plus the fourth.

Glancing down columns 6, 8, and 9 it will be observed that in general the higher grades have the higher scores of recognitive confusion, tho the correspondence between the 3 orders is not exact. Every pair of trade-marks, dissimilar, random, and litigated received some rating of confusion. The grades do not divide into two separate groups, suggesting those likely to cause confusion and those not likely to cause confusion; their distribution forms a continuum, as in the case of the confusion scores. The grades of the litigated trade-marks (from "Holeproof—Knotair" to "Johnston's—Johnson's" inclusive) spread over about eight tenths of the series, and show that there are no wide gaps in relative degrees of confusion, the widest being 1.1 (between grades 2.4 and 3.5, and between 8.9 and 10.0).[4] The difference between the grades are generally only a few tenths of a unit. In several instances there are two to four equal grades. The frequencies of the grades resemble a normal distribution.

Examining the litigated trade-marks we find that "Holeproof—Knotair" with the grade of 2.0 has the lowest grade; "Johnston's—Johnson's" with the grade of 8.9 the highest; and "Capital—Capitol" with the grade of 8.8 the second highest, the difference between the latter two being slight, being respectively 1.1 and 1.2 below the ratings for the identical pairs. It is easy to see why the last 10 pairs received the highest positions. While some pairs seem to deserve as high a rating as those above them, it should be remembered that the average grades representing the combined judgments of 50 individuals possess more validity than any single one. "Capital—Capitol" got perhaps a lower score in the Uninformed than it really deserves; but in the Informed it is satisfactory. These two trade-marks are peculiar in that their resemblances have been particularly noted before the experiment by many. In school the attention of children is usually directed to the similarity in spelling and the related significance of these words. In the relative position experiment the existence of this fact does not influence the judgments; we may therefore consider its results in this case more valid. Furthermore, the results of the relative position experiment on the whole are the more reliable as they come from 50 subjects and those in each of the recognition experiments from 20.

The artificial matings hold distinctive positions at both extremes of the scale. As might easily have been foretold the three identical matings were assigned to grade 10 by every observer. No litigated pair of marks received this average grade, altho some individuals placed various other trade-marks in this grade occasionally. The dissimilar and random matings occupy the lower limit of the scale. They are all lower than the lowest grade of the litigated marks. The three dissimilar matings do not have all equal scores, nor do the random ones. The range of the former is 0.1 to 0.3, of the latter 0.2 to 1.0. The lowest artificial mating is the dissimilar one, "Syphon—Black Diamond" with the grade of 0.1; the highest

[4] It is not believed that this difference of 1.1 in these two cases is really equal, the latter is probably larger.

artificial mating is the random one, "White Rock—Stonetex" with the grade of 1.0. Even a glance at the dissimilar and random matings shows that "White Rock—Stonetex" has quite markedly more deceptive similarity than the others. It is 1.0 lower than "Holeproof—Knotair" the lowest litigated pair.

It has already been pointed out that the mental processes of the observer recognizing an imitation are substantially similar to those of a purchaser presented with an imitation in daily life. Likewise, the mental processes of an observer judging and grading trade-marks in the relative position experiment bear a certain resemblance to those of a judge deciding the question of infringement of trade-marks in a case at the bar. A comparison of the mental processes of the observer and of the judge will show the superiority of the experimental procedure over the legal.

When the court has to decide on the question of confusion between two word trade-marks its judgment is usually based on the basis of three sources of facts. It hears the briefs of the lawyers for and against; it decides whether there is any likelihood of confusion between the trade-marks; and it considers how these new marks compare with previously litigated infringements and non-infringements. The judge in the last performance is doing something similar to that which the observer in the experiment is doing. But the judge has only two categories in which to register his judgment, namely: those that are likely to cause confusion and those not likely to cause confusion; the observer has 11. Consequently the judgments made by the latter are much finer. When as a result of an error a trade-mark is assigned to the wrong category by a judge, the error is large and significant; in the experiment the finer grading reduces it greatly. Strictly the legal category "not likely to cause confusion" would be located on our scale at or near grade 0; and the legal category "likely to cause confusion" anywhere from grade 0 to and inclusive of grade 10. The fact that the legal categories are ill-defined and without quantitative significance cause them to be variously interpreted by different judges. Hence, the inconsistency of many legal decisions is partly accounted for.

The experiment is mathematically more valid or accurate than the court in that its results are determined by a larger number of individuals, whereas the basis of the decision is the opinions of only a few judges. Inasmuch as the experiment employs a far greater number of observers it necessarily yields results that are much less likely to be the outcome of the chance bias of a few individuals. In the average of a large number of judgments chance bias is neutralized, and the result represents the tendency of all the observers and is nearer the truth. The court c nnot help being influenced by the counsels of plaintiff and defendant, thereby tending to partiality to one or the other. A very able trade-mark lawyer would undoubtedly have a far better chance of winning a case than would his less able opponent. Such unfairness is entirely eliminated in an experiment scientifically conducted.

2. Psychological Principles for Determining Dangerous Imitations

Infringers employ a number of ways to copy a trade-mark. A few are bold enough to make the imitation identical with the original, while most try to make it more or less similar. In judging whether a trade-mark is an illegal imitation the courts have no rules or principles to guide them. An analysis of the methods of making imitations can furnish general principles that may be of assistance to the courts. An examination of the imitations that have proved most dangerous in the experiments will enable us to formulate such principles, and tho the principles cannot be followed rigidly it will certainly be helpful to bear them in mind. In addition to the important matter of familiarity, confusion of trade-marks may be broadly analyzed as dependent on visual appearance, sound, linguistic formation (fancifulness, oddity, infrequency, structure,—compound or hyphenated), meaning, and any combination of these factors.

Among the imitations are found substitutions, transformations, reversals, omissions, and additions of words, syllables, letters, hyphens, and capitals. These changes are made at the beginning, middle, end, before, or after the word or words. The letter added, omitted, or changed may be identical, similar, or dissimilar. In some cases the meaning of the imitation may be identical, similar, or even the opposite of the whole or of part of the copied original. A few of the imitations aim at similarity in meaning alone, as in "Holeproof—Knotair";[5] most adopt the other features of the original also. Similarity in one of these features usually includes similarity in the others.

The analysis of the methods of constructing the following 10 most confusing pairs of trade-marks will illustrate many of these points. In "Johnston's—Johnson's" the imitation is the same as

[5] Most persons pronounce the imitation "Knotair," "No tear," but a few pronounce it "Not air," and thus can see no similarity in meaning between it and the original.

the original with the omission of one internal consonant. In "Capital—Capitol" the imitation is the same as the original with the exception that one internal vowel is replaced by another, the imitation being similar in sound and related in meaning. In "Nubia—Nubias" the imitations makes the original plural by the addition of a consonant. In "German Sweet—Sweet German" the imitation simply reverses the positions of the two words in the original. In "Ceresota—Cressota" the imitation reverses the position of two adjacent internal letters, and a consonant (the same kind as the following one in the original and in the imitation) is substituted for a vowel (the same kind as the one of the two reversed letters in the imitation, and which is the second letter preceding it in the original); thus the imitation is one syllable shorter than the original. The words resemble each other in sound and partly in significance. In "Magic—Magico" by suffixing a final vowel a syllable is added, the similarity in meaning being retained. In "Worth—Our Worth" the imitation simply places a personal pronoun before the original. In "Cottolene—Cottoleo" the imitation takes the first seven letters, and substitutes for the 2 final letters a vowel, thereby adding another syllable. In "S. B.—B. & S." the two initials are reversed and an ampersand is inserted between them. In "Uno—Ino"[6] the imitation substitutes another vowel, similarity in meaning being also suggested. In Table IV. among the 10 most confusing pairs that are not in Table VI. are the following, In "Siphon—Siphon System" a word is added after the original. In "West End—East End" the imitation substitutes for the first word a word of the opposite meaning which is similar in length and somewhat so appearance. In "Willoughby Lake—Willoughby Ridge" for the second word is substituted another similar in length, belonging also to geographical terminology.

It may be seen from this that a definite classification of the methods of imitating is not always possible, as many imitations are the results of a number of different devices. In "Ceresota—Cressota" for example, the imitation omits a letter, adds one, reverses two, and is one syllable shorter than the original. The length, or the number of letters, syllables, and words in a trade-mark is an important consideration in judging confusion. Other things being equal, a slight change in a long trade-mark is more deceptively simi-

[6] The marks are generally pronounced "You know—I know"; but one individual well versed in English etymology, saw no similarity in meaning as he pronounced them "Oo know—In know." This instance and the other just noted show that similarity in meaning may be conveyed to one individual and not to another, according to the pronunciation of the original and the imitation.

lar than in a short one, and hence will cause greater confusion. Feingold wisely considers this a matter of much importance.[7] It should also be remembered that trade-marks may differ greatly in length, while a part that is unappropriated by the imitation may be identical with an essential part of the original.

To determine whether imitations are dangerous the following principles are suggested by Tables IV. and VI. as a useful general guide.

1. *Omissions.*—Imitations differing from the originals only by omitting 1, 2 or 3 letters, 1 or 2 syllables, or 1 word.

2. *Additions.*—Imitations differing from the originals only by adding 1, 2 or 3 letters, 1 or 2 syllables, or 1 word.

3. *Substitutions.*—Imitations differing from the originals only by substituting 1, 2 or 3 letters, 1 or 2 syllables, or 1 word in the same or different positions in the trade-mark.

4. *Changes in Positions.*—Imitations differing from the originals only by transpositions or reversals of 1, 2 or 3 letters, 1 or 2 syllables, or 1 or 2 words.

Many imitations will have to be examined in connection with more than one of these statements. Because of the variety of features and factors operative in some cases of confusion, the above classification cannot be expected to be as accurate as a mathematical formula. In some cases, as may be seen from Tables IV. and VI., they will apply; in others they will not. It should be borne in mind too that the length of a trade-mark must be taken into account in connection with the above principles.

Finally, the names of the commodities in Table IV. show roughly that the trade-marks most often imitated are those of household articles, then in decreasing order those of foods, medicines, clothing,—firm names and liquors in about the same proportion. The most reliable way, however, of finding which articles are most frequently imitated would be to consult a far greater number of court records.

3. Scale of Relative Position Confusion

The purpose of constructing a scale of relative position confusion is the same as that of recognitive confusion, namely to supplant the present judicial procedure. The relative position scale, tho it too is not ready for adoption, should be considered in connection with the remarks made on Scale I. Scale II. is derived from Table VI. It contains 18 pairs of trade-marks, representing 18

[7] *Op. cit.*

degrees of confusion. In view of several circumstances it was not possible to make the distances between the steps equal. The smallest is 0.3, the largest 1.2, and the average of all 0.58, or nearly six tenths of a step. Scale II. is superior to Scale I. in one respect; the grade of confusion being based on the trade-marks, not as in Scale I. on both the trade-marks and the name of the article. When the Uninformed subjects mistook "Grenade" for "Grenadine" the appearance of the word "Syrup" in both connections tended to cause some of the confusion. The application of Scale II. follows the second method described under Scale I.

SCALE II

SCALE OF RELATIVE POSITION CONFUSION OF TRADE-MARKS

		Order	Confusion Grade	P. E.
Syphon	Black Diamond	1	.1	*.02*
Everstick	Herringbone	2	.7	*.12*
White Rock	Stonetex	3	1.	*.11*
Holeproof	Knotair	4	2.	*.21*
Royal Irish Linen	Royal Vellum	5	2.4	*.20*
Rubberset	Rubber-vulc	6	3.5	*.22*
Beats-All	Knoxall	7	4.1	*.27*
Bear Lithia Springs	Great Bear Springs	8	4.5	*.22*
Shipmate	Messmate	9	5.0	*.25*
Social Register, Newport	Newport Social Index	10	5.5	*.28*
Pep-Kola	Pepko	11	6.2	*.18*
Dyspepticure	Dyspepticide	12	6.6	*.23*
Uno	Ino	13	7.0	*.19*
Magic	Magico	14	7.6	*.19*
German Sweet	Sweet German	15	8.0	*.14*
Nubia	Nubias	16	8.3	*.15*
Capital	Capitol	17	8.8	*.11*
Drinket	Drinket	18	10.	*.00*

Let us suppose that the trade-marks "Syphon—Black Diamond" to "Royal Irish Linen—Royal Vellum" inclusive are non-infringements, and all above the latter infringements. An imitation receiving then a grade of confusion above 2.4 may be imagined to be illegal, one receiving 2.4 or less legal. When a new case comes up the pair of trade-marks can be compared with the various steps on the scale. A number of individuals are asked to match independently as closely as possible the new pair to one on the scale, and rating it the corresponding grade. From the average grade of confusion thus obtained, it may be determined whether the imitation is an infringement or non-infringement. If the average grade of these ratings is 2.4 or less the imitation is a non-infringement, if more an infringement. If for instance "Holeproof—Knotair" received an average grade of 2.0, it would be a non-infringement, and if "Our Country—Old Country" received 6.5, it would be an infringement.

Separate scales devised on the basis of each of the three factors contributing to confusion would be of theoretical and practical interest. A comparison of the gradings of the same trade-marks in the different scales would tell the correlation between deceptive similarity in appearance, similarity, meaning, and all combined. The practical importance of the scales would lie in the fact that the trade-marks could be rated on the particular factors that were the cause of confusion. There are quite a number of trade-marks that are likely to be confused usually under visual conditions, or in sound. By grading the former on the scale for visual confusion, and the latter for auditory confusion, more detailed information is obtained. In determining the question of infringement these grades might also be considered in connection with those on the scale of general confusion.

4. Comparison of the Relative Position Results with Those of Recognition

One of the main objects of this chapter, still to be considered, is the relation between confusion as determined by relative position and recognition. Inspection of Table VI. gives a general impression that there is a positive correspondence between the two orders. By working out correlations and by epitomizing these results we may obtain, however, a better impression. The latter is done in Table VII., by averaging in four groups the relative position and recognition scores of the trade-marks "Holeproof—Knotair" to "Johnston's—Johnson's" inclusive in Table VI. Table VII. is formed and reads like Table V. The 9 trade-marks in the first row are "Holeproof—Knotair" with the grade of 2.0 to and including "Bestyette—Veribest" with the grade 4.4, etc.

The average grades of confusion run in column 4 from 3.5 to 7.8, and in volume 6 the corresponding scores of confusion in the Uninformed run from 29 per cent. to 55 per cent. with steady increase at each step. In column 8 the corresponding scores of confusion in the Informed run with quite a regular increase from 8 per cent. to 36 per cent. Column 10, likewise in agreement, shows at each step that the higher the grade of confusion the lower the ability to recognize the change in the imitation. Column 12 shows that the incorrect discriminations do not depend on deceptive similarity. Column 14 shows also that the identifications of the originals are not related to the degree of confusion. In other words, the order as determined by the recognition scores of confusion agrees with the order by the relative position grades. It will be noticed that the orders of the recognition scores of confusion and change are in somewhat closer agreement with each other and with that of the relative position grade than with themselves in Table V. This is an indication that the order established in Table VII. by the relative position grades is more valid for relative differences than that of confusion in the Uninformed in Table V.

TABLE VII

THE AVERAGE GRADES OF CONFUSION, AND THE RECOGNITION SCORES IN THE UNINFORMED, THE INFORMED AND THE CONTROL GROUPS, WITH THE PROBABLE ERRORS FOR DIFFERENT TRADE-MARKS OF VARIOUS AMOUNTS OF CONFUSION

	Relative Position				Uninformed		Groups Informed: Av. Per Cent. Recognized As:						Control	
No. of Trade-Marks	Range of Grades of Confusion	Difference between Extremes of Range[8]	Av. Grade	P. E.	Av. Per Cent. Confused	P. E.	Identical	P. E.	Changed	P. E.	New	P. E.	Av. Per Cent. of Correct Iden. of Originals	P. E.
9	2.0–4.4	2.4	**3.5**	*.14*	**29**	*3.1*	**8**	*2.0*	**65**	*2.5*	**27**	*2.5*	**86**	*2.0*
10	4.5–5.5	1.0	**4.9**	*.08*	**44**	*2.0*	**17**	*2.1*	**52**	*3.8*	**31**	*3.2*	**83**	*2.7*
10	5.8–6.7	.9	**6.3**	*.06*	**49**	*3.1*	**26**	*3.6*	**49**	*3.7*	**25**	*2.5*	**81**	*1.4*
10	7.0–8.9	1.9	**7.8**	*.15*	**55**	*4.4*	**36**	*3.2*	**45**	*4.0*	**19**	*2.2*	**84**	*2.2*

The relation between the various orders may also be seen from their correlations. They are all positive. According to Spearman's formula[9] the order of confusion in the Informed group correlates + 0.70 (P.E. 0.061) with that of the grade of relative position, that of confusion in the Uninformed correlates + 0.59 (P.E. 0.076), and that of correct recognitions of change correlates inversely + 0.47 (P.E. 0.090), and that of incorrect discriminations correlates + 0.19 (P.E. 0.109). The last is nearly a chance correlation. In Table VII. the differences in the regularity of the changes in the scores indicate also which orders correspond more closely. The probable errors of most of the correlations overlap.

All the correlations are attenuated by the fact that the names of the commodities contributed to causing confusion in the recognitions experiments, whereas the relative position experiment is free from

[8] These differences could have been made more equal if we had taken six steps instead of four, but then the average would not have been so reliable.

[9] The method of differences in relative positions or ranks. $\rho = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$

this source of confusion. It is perhaps due partly to this that the order of confusion in the Informed received the highest correlation. Tho the name of the article appeared there it was not so much of a disturbing factor because the subjects knew that the change would be only in the trade-marks. In the Uninformed group, on the other hand, it did contribute to cause confusion between the trade-marks, and thereby lowered its correlation with the relative position grades. Finally, while the correlations demonstrate that measurement by relative position is useful as a recognition method for studying confusion, they tell against recent criticisms denying its practical value.

5. Introspective Notes

How did the individual perform the task of judging and grading deceptive similarity? Did they all work in the same manner? Or did some adopt their own methods to aid them? No one line of thought was followed by all. A number of individuals put different emphasis on each of the four factors noted in the directions. The judgments of some were based mainly on similarities, while those of others on differences. Furthermore, imaging concrete situations in everyday life seemed to help quite a few. This information was obtained by asking each individual after he had arranged the trade-marks to state how he had proceeded. Only a general statement was required, not an elaborate introspection. To many of the replies obtained do not really merit the name of proper introspection, they indicate quite a variety of mental performances. Below are given 16 different statements by the 50 subjects. It would be interesting to know how much the variability of the grades depends on these differences in method.

1. Confusion.
2. Confusion and similarity.
3. Similarity in appearance, sound, and meaning.
4. Graded in order of importance by similarity in appearance, sound, and meaning.
5. Similarity in sound given high confusion rating.
6. Noted similarities more than differences.
7. Observed differences systematically.
8. Considered the ease of associating the rival trade-marks.
9. The most important words given most weight.
10. Imagined whether he would mistake the imitative trade-mark when read in a magazine advertisement for the original seen first in another part of the magazine.

11. Played the rôle of a purchaser looking at the imitation.
12. Imagined how he would react to the imitative trade-mark, having only a casual acquaintance with the original.
13. Tried to imagined the trade-mark on packages in a store.
14. Considered whether others would be confused.
15. Fancied himself a clerk in a store trying to fool the purchaser with the imitation.
16. Direct and quick comparison of the trade-marks.

6. Summary

1. Every imitation, non-infringements as well as infringements, receives some grade of confusion.

2. The relative position grades of the imitations do not divide into two distinct groups, suggesting those likely to deceive and those not likely to deceive, but their distribution forms a continuum. The last two findings are in agreement with those in (4) and (5) of Chapter III. telling against the present legal procedure in regard to imitations.

3. Even the random and dissimilar matings of trade-marks receive a grade of confusion, all being lower than the litigated pairs.

4. The analysis of the methods of making imitations suggests certain general psychological principles for determining dangerous imitations. These principles include omissions, additions, substitutions, and changes in positions.

5. There are fairly high correlations of the order of confusion of the relative position grades with the two orders of confusion in the Uninformed and Informed groups.

6. In judging and grading the individuals did not follow a single line of thought, but employed methods of their own to aid them.

CHAPTER V

PSYCHOLOGICAL TESTS OF THE ACCURACY OF JUDICIAL DECISIONS[1]

1. The Tests

When the court renders a decision as to the likelihood of confusion of trade-marks, instead of establishing a fact it merely states an opinion. Unless, there is a dissenting judge the court probably believes that its decision is correct. But no court has undertaken to see whether its decision was right, nor to see whether the decision did in fact conform with the cited authorities. Scientific investigation would further suggest that even the authoritative cases be subjected to a test. As long as these things remain undone the courts are traveling in the dark. The recognition and relative position experiments that we have been examining furnish us with the means of determining the legal and psychological accuracy of the decisions. The scanty information that the decision conveys is one of its serious defects. It rarely advances beyond stating that there is or is not a possibility of confusion, and making a comparison between the trade-marks in the pending case with litigated pairs. How much more accurately the psychological methods can answer these questions has already been demonstrated.

There were 40 subjects in the recognition experiment, all of them Uninformed. There were 40 in the relative position experiment; the serial form, not the group form, being used. The nine pairs of trade-marks studied were not very familiar; as may be seen in Table VIII. There the results of the recognition and relative position experiments are given. The first column in the table shows whether the imitation in the decision was held to be an infringement (I) or a non-infringement (N). The other columns are self-explanatory.

The procedure and material has been described in Chapter II. However, an additional word might be said about the nine decisions, which are to be checked up. Five were adjudications of infringement, in which the use of the imitation was enjoined, and four were

[1] The results of this chapter appeared in a previous article by the writer, "A Psychological Study of Confusion between Word Trade-Marks," *Bull. of the U. S. Trade-Mark Assn.*, 1915, 11, 101–114.

TABLE VIII[2]

THE PER CENT. OF INDIVIDUALS CONFUSED, THE GRADE OF CONFUSION AND PROBABLE ERROR OF EACH OF THE INFRINGEMENTS AND NON-INFRINGEMENTS

Decision	Trade-Mark Original	Imitative	Per Cent. Confused	Av. Grade	P. E.
N	Sozodont	Kalodont	28	3.6	*.27*
I	Nox-all	Non-X-Ell	28	4.9	*.23*
I	Club	Chancellor Club	35	2.7	*.31*
N	Bestyette	Veribest	35	4.1	*.30*
N	Mother's	Grand-Ma's	38	3.2	*.30*
I	Au-to-do	Autola	40	4.3	*.31*
N	Peptenzyme	Pinozyme	43	5.2	*.30*
I	Green River	Green Ribbon	50	5.7	*.29*
I	Ceresota	Cressota	63	7.9	*.19*

adjudications of non-infringement, in which injunctions against the use of the alleged illegal imitations were refused. The results obtained from these 9 trade-marks test the accuracy of the decisions. No other circumstance or reason, as far as the records showed, determined the point of infringement or non-infringement in the decisions. This requirement had to be insisted upon, because it is only the question of likelihood of confusion of the word trade-marks that the experiments measure. If various other factors such as the question of the validity of the trade-marks, unclean hands, similarity of the type, color, or other features of the label or package entered in the decision and operated to influence the point of infringement, the experimental results of these complicated decisions could not be properly compared with each other, nor with simple decisions of confusion of just word trade-marks. It is obvious that decisions determined by confusion of word trade-marks plus certain other circumstances are not in the same legal nor psychological categories as those de-

[2] Ceresota, Cressota, flour; North Western Consolidating Milling Co. v. Mausser & Cressman, 162 Fed. Rep., 1004 (U. S. Cir. Ct.).

Nox-all, Non-X-Ell, hats; Nox-All & Gotham Co. *v.* Denzer Goodhart & Co., 2 Trade-Mark Rep., 356 (U. S. Dist. Ct.).

Green River, Green Ribbon, whiskey; Lang *v.* Green River Distilling Co., 148 O.G., 280 (Ct. of App. D. C.).

Club, Chancellor Club, cocktails; In re Herbst Importing Co., 134 O.G., 1565 (Ct. of App. D. C.).

Au-to-do, Autola, cigars; In re Wilcox Co., 162 O.G., 539 (Ct. of App. D. C.).

Bestyette, Veribest, raincoats; New York Mackintosh Co. *v.* Flam, 2 Trade-Mark Rep., 324 (U. S. Dist. Ct.).

Sozodont, Kalodont, tooth paste; Sarg Sohn & Co. *v.* Hall & Ruckel, 165 O. G., 732 (Ct. of App. D. C.).

Mother's, Grand-Ma's, waxing pads; Bromund Co. *v.* Columbia Wax Products Co., 200 O.G., 1115 (Ct. of App. D. C.).

Peptenzyme, Pinozyme, digestant; Reed & Carnick *v.* Waterbury Chemical Co., 200 O.G., 279 (Ct. of App. D. C.).

termined simply by the confusion of just the word trade-marks. Nor could the former be properly given as authorities governing the latter, nor vice versa.

Let us now compare the recognition scores of the infringements and non-infringements. The two most confusing imitations in the recognition and relative position experiments are infringements, the lowest in the recognition is a non-infringement and in the relative position is an infringement. The non-infringement "Peptenzyme—Pinozyme" is more confusing than 3 of the 5 infringements; the non-infringement "Mother's—Grand-Ma's" is more confusing than 2 of the infringements; and the non-infringement "Bestyette—Veribest" is equal in confusion to one infringement and more so than another. Of the 2 imitations that stand at the bottom one is legal and the other illegal. The grades of confusion in relative position show likewise that the non-infringement "Peptenzyme—Pinozyme" is more confusing than the same three infringements. The three remaining non-infringements are more confusing than one infringement. In other words some of the imitations declared to be legal actually confused more individuals than some imitations declared to be illegal, and vice versa. Both experiments prove, therefore, that some of the decisions are inaccurate and inconsistent. In several of the above comparisons the differences between the scores are not large. It should also be observed that many of the pairs of trade-marks differ only little in confusion. It will be seen that the orders of confusion in recognition and relative position correspond fairly well.

The average grades of confusion here are not comparable with those in Chapter IV., tho their relative amounts are. Some recognition scores of the trade-marks recorded here and in the two previous chapters show divergences due to the different classes of subjects used and their small number.

TABLE IX

THE AVERAGE PER CENT. OF CONFUSION, THE AVERAGE GRADE AND PROBABLE ERROR OF THE FIVE INFRINGEMENTS AND FOUR NON-INFRINGEMENTS

Number of Decisions	Decision	Av. Per Cent. Confused	Av. Grade	P. E.
5	Infringement	43	5.1	.51
4	Non-infringement	36	4.0	.26

If we compute the averages of the 5 infringing and 4 non-infringing imitations in Table VIII. we get the figures in Table IX. The average percentage of confusion (43) of the infringements is 7 per cent. higher than the average (36) of the non-infringements.

The advantage here is as it should be. This is also the case with the average in relative position. The average grade (5.1) of the infringements is 1.1 higher than the average (4.0) of the non-infringements; their probable errors are 0.51 and 0.26 respectively.[3] Tho the differences between the average grades show that on the whole the courts were correct in their judgments, the inaccuracies of the individual cases are covered up. The most important characteristic of these differences is their small amount. Let us consider again the recognition scores in Table VIII. In view of the fact that the largest difference between 2 infringements, "Ceresota—Cressota" and "Nox-all—Non-X-Ell" is 35 per cent., and that between 2 non-infringements, "Peptezyme—Pinozyme" and "Sozodont—Kalodont" is 15 per cent., the difference here of 7 per cent. between the 2 averages becomes insignificant. The grades show this discrepancy more decidedly. Here the largest difference between 2 infringements, "Ceresota—Cressota" and "Club—Chancellor Club," is 5.2, and that between 2 non-infringements, "Peptenzyme—Pinozyme" and "Mother's—Grand-Ma's" 2.0, whereas the difference between the 2 averages is only 1.1. The scores of the two supposedly distinct decisions are thus seen to overlap throughout most of their range. The differences within both classes of decisions outweigh so much the difference between the classes that for practical purposes the difference in confusion between the infringing and non-infringing imitations may be disregarded. Therefore, some of the judicial decisions are unreliable and do not represent two really different legal or psychological categories.[4]

2. Summary

1. Some of the imitations declared to be legal by the courts actually confused more individuals in the recognition experiment than some imitations declared illegal, and vice versa. The results of the relative position method confirm this. Therefore, some of the judicial decisions were inaccurate and inconsistent.

2. The scores of the imitations of the two supposedly distinct decisions overlap throughout most of their range, so that the decisions do not represent two really different legal or psychological categories.

[3] The probable errors are calculated from the average grades of the 9 trade-marks; if they had been calculated from the individual grades they would have been much smaller, probably one half as large.

[4] Several other experiments by the writer, as yet unpublished, confirm this conclusion.

3. Concluding Remarks

In conclusion it may be said that the adoption of the experimental method of measuring confusion of trade-marks will insure several advantages. Trade-mark legislation will become more efficient in having an objective standard by which to define an illegal imitation. Judicial decisions will become more accurate and will be freed from being blindly governed by past erroneous decisions. Business will save money by not being deprived of legal trade-marks thru inaccurate decisions, by cutting down the fees of trade-mark lawyers, and by the elimination of the opportunity of frequently appealing from the decisions. Furthermore, the state and federal governments will economize in time, money, and labor spent by Equity judges and the officials in the Trade-Mark Division of the Patent Office. Finally, while opening a new field for applied psychology the experimental method will aid in perfecting the machinery of law and its administration, and in purifying some of the unfair practices in business.

B.

INDIVIDUAL DIFFERENCES AND FAMILY RESEMBLANCES IN ANIMAL BEHAVIOR

I

A STUDY OF HABIT FORMATION IN VARIOUS STRAINS OF MICE

BY

HALSEY J. BAGG

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. WOODWORTH

No. 43, June, 1920

Columbia University Contributions to Philosophy and Psychology

Volume XXVI, No. 4

NEW YORK

THE SCIENCE PRESS

Agents: G. E. STECHERT & CO.; London (2 Star Yard, Carey St., W. C.); Leipzig (Koenigstr., 37; Paris (16, rue de Condé).

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

ACKNOWLEDGMENTS

The writer is indebted, above all, to Professor J. McKeen Cattell for suggesting the problem, and for helpful criticism and encouragement throughout the experiment. In the preparation of the manuscript, and in the evaluation of the statistical material, thanks are also due to Professors Robert S. Woodworth, Edward L. Thorndike and Henry A. Ruger.

CONTENTS

I. HISTORICAL

In the present thesis an effort has been made to combine in a single study three main points of investigation; first, the genetic study of behavior; secondly, the subject of individual differences; and thirdly, a consideration of the exact method of habit formation employed by the mice that have been tested in the experiments that are now to be reported.

Practically no experimental work has been done upon individual differences and family resemblances in animal behavior. In most cases, the behaviorist has been content to study the mass reaction of a group of animals to external stimuli, and in the main, has not attempted to treat the variability of his group because of the relatively small number of animals tested. Professor J. McKeen Cattell, about fifteen years ago, began to apply the methods of genetics to the study of conduct, but the results obtained by him and his students were not published, and the problem has been given to me. Yerkes devotes a chapter of his book on "The Dancing Mouse,"[1] to differences in behavior, and there brings together results for variability in ". . . general behavior, rapidity of learning, memory, and discrimination." His results showed the existence of a considerable amount of individual differences in the behavior of the dancing mouse, and no family resemblance in the litters he obtained. He does not give quantitative results, but confines himself to a general discussion of the individual peculiarities of the animals he had tested. The following quotation from his text expresses this point: "I noted, in this test of the animals' ability to learn, that while one individual would be scurrying about trying all ways of escape, investigating its surroundings, looking, sniffing, and dancing by turns, another would devote all its time to whirling, circling, or washing itself. One in the course of its activity would happen upon the way of escape, the other by reason of the limited scope of its activity, not the lack of it, would fail hour after hour to discover even the simplest way of getting back to its nest, to food, and to its companions." Concerning the "inheritance of forms of behavior," Yerkes found that certain lines of descent exhibited a pronounced tendency to whirl to the left, while others reacted in the opposite direction. When two such strains were crossed the offspring showed an equal frequency of left and

[1] Yerkes, Chapter 17, "The Dancing Mouse."

right whirlers. It was also found that there was no "inheritance of individually acquired forms of behavior." Apparently the descendants of animals that had been previously trained to learn a certain task were given no advantage over ordinary individuals from untrained stock.

G. V. Hamilton in his monograph entitled "A Study of Perseverance Reactions in Primates and Rodents,"[2] found that there were definite types of behavior exhibited by the various subjects he used. These consisted of twenty children, a baboon, four monkeys, and five kinds of rodents, comprising, one mouse, five gray rats, five black rats, ten white rats and six gophers. The reactions of the monkeys and the baboon presented a considerable range of individual differences, which determined the experimenter in the selection of his subjects, as indicated in the following quotation: "The marked individual differences presented by the five infra-human primate subjects reflect a policy of selecting subjects in whom oddities of general reactive equipment had been observed." Later on in the investigation the author refers to the presence of individual differences, as follows: "When a mammalian is confronted by a series of situations for which he is unable to discover and stereotype a specifically adequate and invariably successful mode of response he tends to vary his response in a manner which is less a species than an individual characteristic."

The writer has had the opportunity of going over the original data of Basset's work on white rats,[3] and finds that a certain amount of individual difference occurs in the animals he tested. Some animals did consistently better work than others, but as Basset himself points out, his numbers were too few to make possible any conclusions from the differences that were observed.

One might mention a large number of isolated cases where the literature of comparative psychology gives evidence of individual differences. It would not be worth while to treat them here, however, because they generally deal with relatively few animals, and are given merely as side issues of experiments planned to bring out other factors.

[2] Hamilton, "A Study of Perseverence Reactions in Primates and Rodents," Behavior Monograph Series, No. 13, 1916.

[3] Basset, "Habit Formation in a Strain of White Rats with Less than Normal Brain Weight," Behavior Monograph Series, No. 9, 1914.

II. STATEMENT OF THE PROBLEM

THE plan of the experimental work presented in this investigation is to measure individual differences in behavior, to determine the extent to which the animal that departs from the average in one direction will depart in others, to measure the resemblances in families and in lines of descent, and to determine whether kinds of conduct can be established in family lines by selection. In a previous publication,[4] of which this thesis is a continuation, it was found that individual differences occurred in the ability of various strains of mice to learn a simple maze, and also that a family resemblance existed among mice of the same litter, that amounted to a coefficient of correlation in the neighborhood of 0.50. Certain mice, and even whole lines of descent, showed marked variations from the average, some taking more than twice the average time to learn a given task. These differences were well beyond the limits of the probable error. The results were obtained from testing 90 mice, as determined by the time required to find their way through a maze. Since then, these mice and their offspring have been tested in other ways, and further experiments are now in progress with the F^7 and F^8 generations. In the present investigation there are described the individual differences and family resemblances of 93 mice, in addition to the 90 already reported on in the previous article. These mice have been tested in the same maze as were the previous ones; in addition, they have been given an interference test, a retention test and have been studied in a second maze, as described below.

[4] Bagg, "Individual Differences and Family Resemblances in Animal Behavior," *The American Naturalist*, April, 1916.

The present material was submitted for publication in June, 1918.

III. METHODS OF THE EXPERIMENT

THE first maze employed was designed by Professor Cattell, the plan of which is shown in Fig. 1. The animal has, in the first compartment, the alternative between two gates, one of which can be pushed open, while the other is locked, with an identical situation in a second compartment. The path that the animal must follow can be altered by varying the position of the open gates. "Unit construction" is used in the dimensions, which are adjusted to the size of the animals used, and which permit the addition of any desired number of standard units.

Preliminary tests were made with albino rats, but later mice, which are more active and more easily handled, were substituted. The mice were given one trial each day at as nearly the same time as possible. Light was found to play but a minor rôle in the tests, daylight and artificial light serving equally well. At the outset the age of the mice when first tested was not always known, but later, when the various litters were obtained, the young mice were tested at, or about, the age of four weeks.

The mice were rewarded for a successful trial by a mixed diet of milk, bread, oatmeal and sometimes meat. A little dry bread was always in their cages. Besides satisfying their hunger, the mice had the additional reward of a place to exercise and the companionship of the mice that had just been tested. The order of the tests was varied day by day. In case the way through the maze was not found in 360 seconds the animal was removed and tested again the following day. The maximum record for a single trial is 360 seconds.

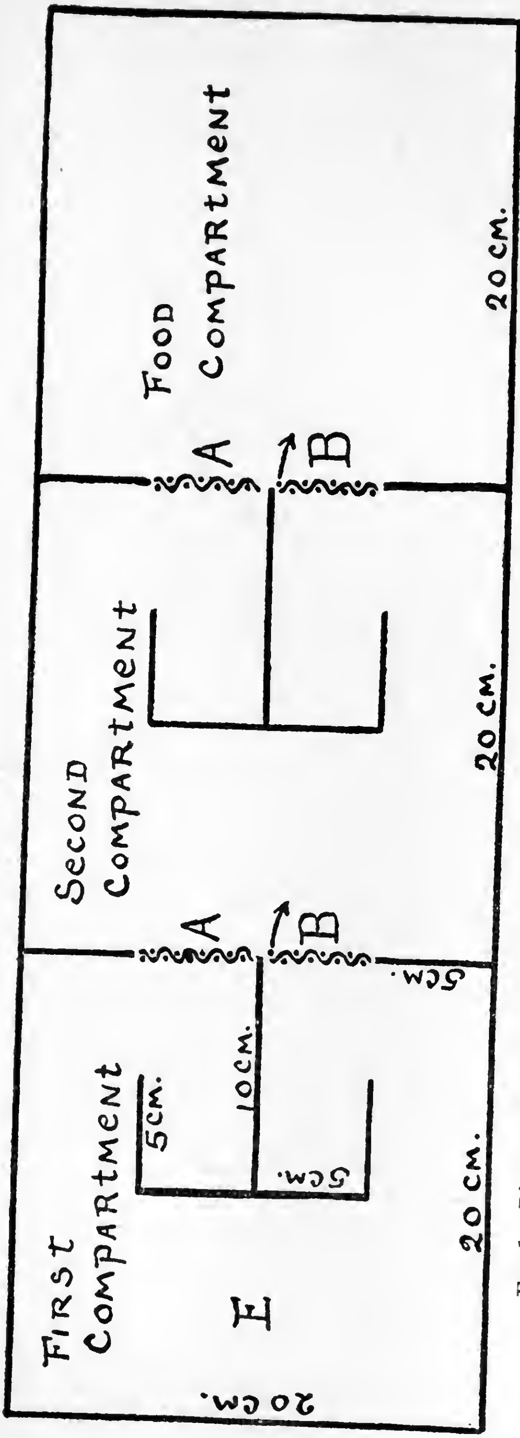


FIG. 1. Diagram of Maze. *A* and *A* = closed wire gates; *B* and *B* = open wire gates; *E* = entrance from above.

IV. TESTS OF LEARNING ABILITY

(*a*) The Maze Test

In the maze test, the mice were first given an initial learning test of seventeen trials. This was a desirable number for two reasons; first, because it was sufficient for the average mouse to learn the maze, and secondly, because the seventeen trials could be divided into three groups representing comparatively distinct stages in the course of learning. The first stage includes the first two trials which are largely affected by chance, and although given here for completeness, are not averaged in the final ratings for each individual. The second group, including the next five trials, represents a period of continued rapid, but less variable learning; the third group, including the following ten trials, covers the period of slow or nearly completed learning. The first group of 90 mice was given only this initial learning test of seventeen trials. The behavior of the latter group of 93 mice was more completely studied. After the seventeen trials the gates were changed, so that the ones that had been open were locked, and those closed were open. This was an interference test designed to give a measure of the adaptability of the animals. It was necessary for the mouse to break the old habit, and learn to get through the maze by way of the previously closed gates. In the first trial after this change, the interference effects were pronounced, although in the eleven subsequent trials the mice rapidly learned the new order. The interference test was divided into two groups. The first two trials were put in one group, and the remaining ten trials in a second group. The first group of two trials gives a measure of the direct effect of the interference upon the animal's behavior, while the second group of ten trials shows the degree to which this interference is carried over into the following trials, and, in a fashion, indicates the adaptability of each animal and the flexibility of its behavior.

(*b*) The Multiple Choice Test

Upon finishing the interference tests the mice were immediately started upon a second experiment which will be here called the multiple choice test. A plan of the apparatus is given in Fig. 2. The mouse was placed in the maze, through the door marked "E," within the first compartment. Here it had a choice of one of four gates, marked in the diagram 1, 2, 3, 4. Three of these gates are blue and one is red.[5] Punishment was given at the blue gates, but not at the red. The gates are the same size as those used in the maze test just

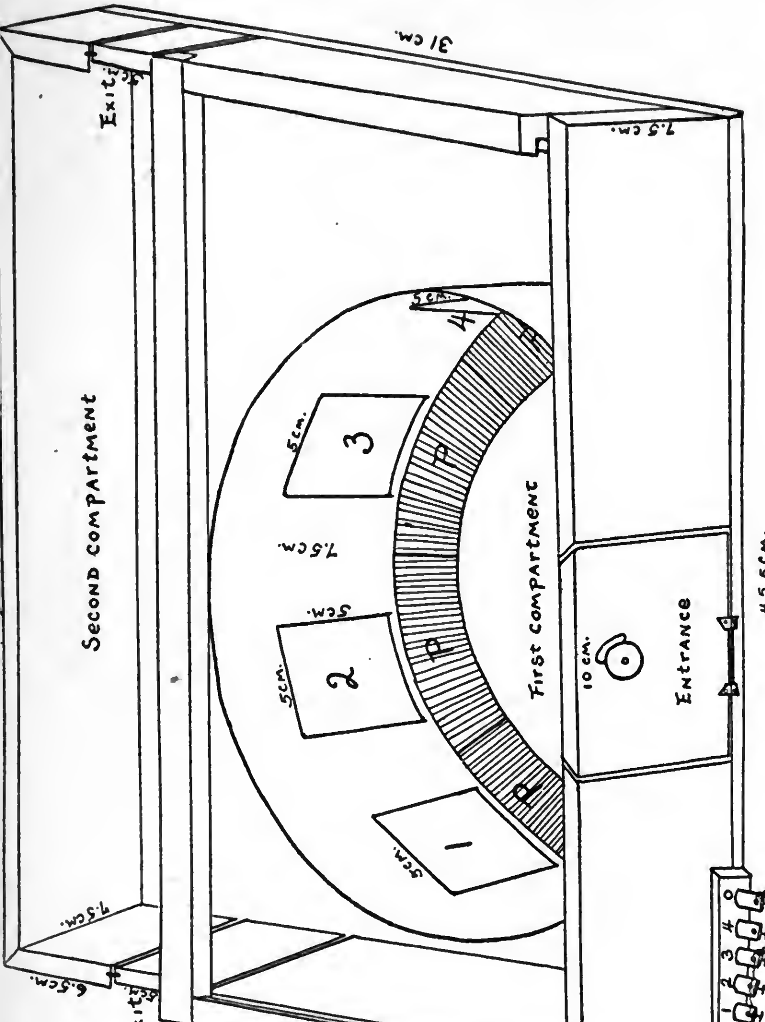


FIG. 2. Diagram of the Multiple Choice Apparatus. The first compartment is covered with glass, which extends as far back as the exit doors. The second compartment is covered by wire netting, hinged at the back and which may be lifted upwards. The electric connections for the punishment pads, "P," are made at the lower left hand corner of the apparatus. The doors, 1 to 4, are set on a piece of sheet iron, bent in a curve as indicated in the diagram, and which is painted a neutral gray.

[5] The Milton Bradly papers, red and blue were used, and, in so far as the experiment was not designed to test the color sense of the animals, the brightness value of the colors was not determined. The papers were changed from time to time, but the odor factors were not eliminated. These clews, if they existed as such, for the animals, were purposely retained, and the mice were allowed to make use of them in solving their problems.

described, the red one being the only one which could be pushed open, and through it the mouse could enter the second compartment. If the mouse attempted to pass through one of the blue doors it received a slight electric shock from the punishment pad "P," which was on the floor of the maze directly in front of the door. The position of the red gate was changed every day, being now in one place and now in another, following a program which had previously been made out. After making the successful choice, the mouse entered the second compartment where he was free to go directly into its nest cage by means of one of the exit doors, where the usual reward was given. Twenty-five trials were made with each individual, and, as in the previous maze, the trials were divided into three somewhat homogeneous groups. The first group, irregular on account of the great play of chance, includes the first two trials; the second group, including the next five trials, marks the period of rapid learning, and the third group of eighteen trials represents the period when the learning was practically completed. It has been found, as will be discussed later on, that the last period of trials was longer than necessary; in fact, the mice showed very little improvement during this period. The day after completing the trials in the multiple choice test the animals were given a series of trials in the maze test first considered. This was a retention test, designed to give a measure of the permanence of association for the previously learned task. In this case ten trials were given, and the gates were opened in the same order as during the last ten trials of the interference test.

V. EXPERIMENTAL RESULTS

(a) Time and Error Averages for All Tests

Tables I. to VII. give the time and the number of errors, *i.e.*, the number of cases in which the mouse tried to go through a locked gate, which is a measure of the activity of the animal. In this paper the average of the last fifteen trials is used as an index of performance for the first set of seventeen trials in the maze test, and the last 23 trials are used as an index for the 25 trials in the multiple choice test, in each case the first two trials being eliminated for irregularities previously mentioned. The groups of two and ten trials each in the interference test, and the group of ten in the retention test are used as indexes for the respective cases. The above tables give the complete records of the 183 mice tested, grouped in families as described below. The average time is 54.12 ± P.E. 2.3 seconds per trial for the last fifteen trials in the maze test; 60.26 ± P.E. 4.7 seconds for the last ten trials of the interference test; 52.81 ± P.E. 4.7 seconds for the retention test, and 39.47 ± P.E. .08 seconds for the last 23 trials of the multiple choice test. The distributions of the individuals in both experiments is shown in Fig. 3.[6] The distribution for the animals in the maze test, based on the average speed attained in the last fifteen trials is indicated by the solid line, and the distribution for those in the multiple choice test, based on the speed attained in the last 23 trials, by the broken line. In the maze test 65 animals took less than 20 seconds, in 47 cases the time was between 20 and 40 seconds, and there were 71 cases between 40 and 360 seconds. But one mouse failed to learn the maze. The distribution in the multiple choice test gave 30 cases in which the time was under 20 seconds, 20 cases between 20 and 40 seconds, and 26 cases between 40 and 280 seconds. None of the mice failed to learn the multiple choice test.

(b) Difference in Learning between Families of White and Yellow Mice

As reported in the preliminary account of this experiment, it was found that certain strains of mice took considerably longer time to learn the maze than others tested at the same time. Among the

[6] Seventy-six of the 183 cases were tested in both the mazes given in these distributions.

colored mice then tested, several yellow ones made poor records. These mice were mated, and they and their offspring compose a group of 27 individuals, whose average time and error record is con-

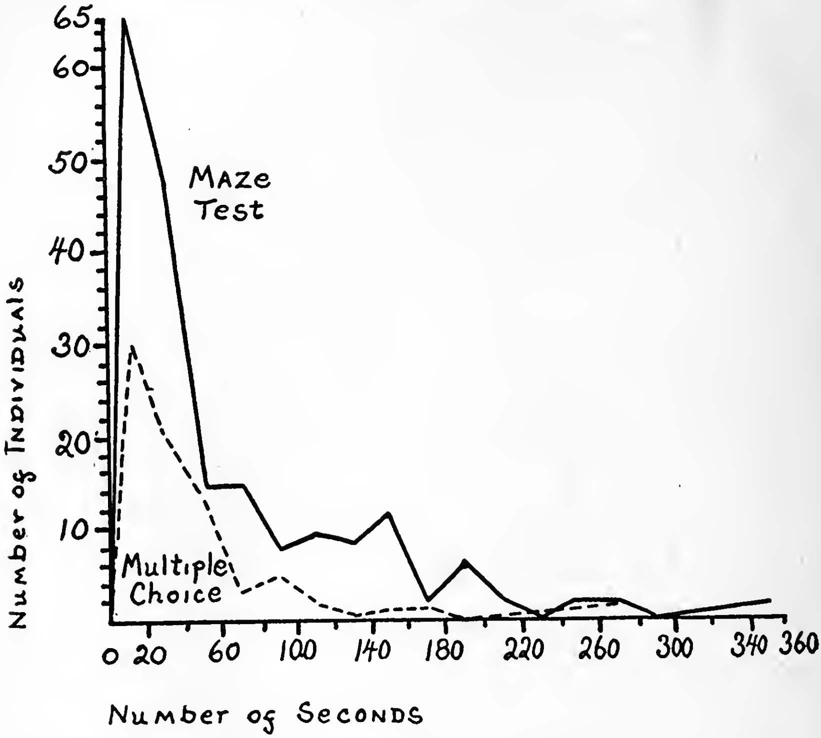


FIG. 3. Total Distribution Curves for Maze Test and Multiple Choice Test.

siderably poorer than the normal for the entire population.[7] The yellow group gave an average time of 83 ± P.E. 7.0 seconds, and an average of 2.0 errors for the last 15 trials. The remaining group of 63 mice, mostly white, gave an average time of 27.5 ± P.E. 2 seconds and .9 error per trial. The yellow mice were thus found to take, on the average, at least three times as much time, and to make twice as many errors as did the white mice. The distribution curves for the white group is skewed, most of the individuals falling between 0 and 20 seconds. The curve for the yellow family is nearly flat, there being about the same number of individuals between 0 and 20 seconds, as between 60 and 80 seconds, and between 140 and 160 seconds.

[7] This group of 27 mice was composed (see Tables I, IV and VII) of Nos. 20 and 26, and their seven offspring; No. 27, the sister of No. 26; a litter of five mice, Nos. 32, 33, 34, 36 and 37 and their ten offspring, and finally two unrelated yellow mice, Nos. 2 and 3, that were used at the beginning of the experiment. The 63 remaining mice of the white group bring the total to 90.

TABLE I

Complete Time and Error Records for the Yellow Family in the Maze Test

No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average	No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average
20Y ♀	360	228	58	115	1.6	38Y ♂	360	277	98	157	3.6
21W ♂	247	41	11	21	.2	39Y ♂	360	143	25	64	2.3
22W ♂	234	28	14	19	.4	40Y ♂	360	47	9	21	.8
23AG ♂	280	43	6	19	.5	41Y ♀	210	20	8	12	.5
24AG ♂	182	47	8	21	.9	57Y ♂	360	171	109	130	3.7
25AGW ♀	177	136	26	63	1.7	58Y ♂	357	130	51	77	2.4
26Y ♂	360	206	172	183	3.7	59Y ♂	360	275	98	156	3.3
27Y ♀	201	10	14	13	.5						

In the first column is given the catalogue number, color and sex of the animals. In the second are the time averages (in seconds) for the first two trials; in the third, for the next five trials; in the fourth, the last ten trials, and in the fifth column the average of the two preceding columns. The error average for the last 15 trials is given in the last row of figures. This order is followed in all the subsequent tables, but in Tables III and VI averages are added for the last group of five trials, and in addition these tables give the time and error averages for two interference tests of two and ten trials respectively; a retention test of ten trials, and finally the averages for the multiple choice test of 23 trials, which is divided, first, into a group of the two first trials, next the following five trials, next the last 18 trials, next the last 23 trials, next the last five trials, and finally the error averages for the last 23 trials.

One day's record has been omitted for mice Nos. 27, 28, 29 and 31 because the poor records for that day were obviously due to a constant error, on account of traveling, etc. These are the only cases where such a condition has occurred.

(c) Results for the Maze Test

Fig. 4 gives the complete record curves for all the tests given in the maze, showing curves based upon the average and the median record for each day, and, as indicated in the drawing, these may be divided into three main parts: first, an initial learning period of seventeen trials, second, the interference groups, consisting of two and ten trials, and finally a retention test of ten trials. 183 mice were tested in the first group of seventeen trials, and 71 in each of the succeeding groups.[8] Two daily record curves were calculated for teach test, and in the upper curve (represented in the figure by a solid line and marked "average") the records for all the individuals in each group were averaged for each successive trial, and the

[8] The tests that followed the initial learning period of seventeen trials were not instigated until the experiment was well started and the writer had become familiar with the peculiarities of the behavior of his subjects. Some animals died during the rather long period in which they were observed, and their incomplete records, although given in the tables, are not averaged in Fig. 4.

probable error calculated for each point in the curve. In accordance with a plan proposed by Professor Cattell, the limits of the probable error are shown by the broken lines. The chances are even that with a greatly increased number of cases the time would have remained between these limits, and a nearly smooth curve can be drawn within them. When the gates were changed at the eighteenth

TABLE II

COMPLETE RECORDS FOR THE WHITE FAMILY IN THE MAZE TEST

No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average	No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average
12W♂......	360	91	56	67	1.9	103W♂.....	108	5	10	8	.5
13W♀......	100	159	37	91	1.4	104W♂.....	174	12	13	12	1.0
15W♀......	316	26	17	21	.4	105W♀.....	99	12	11	11	.8
48W♂......	156	16	7	10	.5	106W♀.....	284	22	12	15	.9
49W♀......	185	75	25	42	1.9	109W♂.....	101	69	8	28	.9
50W♀......	246	25	18	20	.7	110W♂.....	93	11	7	8	.4
51W♀......	56	13	7	9	.5	111W♂.....	67	13	12	12	1.0
52W♀......	169	12	10	11	.9	112W♂.....	151	33	8	16	.9
53W♀......	196	77	34	48	1.7	113W♂.....	88	11	7	8	.7
65W♀......	285	87	8	35	.9	114W♀.....	75	13	5	8	.4
66W♂......	360	306	17	113	1.2	115W♀.....	70	13	11	12	1.0
67W♀......	360	173	25	75	1.7	116W♂.....	183	13	8	10	.7
76W♀......	222	58	17	31	.9	117W♂.....	318	11	4	7	.5
77W♂......	317	162	20	68	1.0	118W♂.....	81	14	10	12	1.0
78W♂......	360	186	91	122	2.9	119W♂.....	115	40	9	20	1.1
71W♂......	149	121	18	53	.9	120W♀.....	118	19	7	11	.5
72W♀......	360	304	30	121	1.7	121W♀.....	64	17	9	12	.7
74W♀......	89	29	21	24	1.2	122W♂.....	76	84	7	33	.7
86W♀......	84	19	13	15	1.2	123W♀.....	66	39	29	32	2.2
87W♀......	58	16	14	15	1.2	124W♀.....	25	23	10	14	.9
88W♀......	86	14	7	9	.8	125W♂.....	189	15	7	9	.5
89W♀......	110	16	10	12	.9	126W♂.....	90	19	9	12	1.0
91W♂......	252	47	33	38	1.6	127W♀.....	72	22	11	15	.9

trial, an interference effect occurred that resulted in a rise of the time curve to 118 seconds at the first trial. This was just half the number of seconds it took the average mouse to go through the maze for the first time. The rest of the interference test showed an average time curve that was above the curve for the last few trials of the initial learning test, except for the sixth day of the interference test when the curve dropped to 40 seconds. The curve based on the average for the interference test began to follow the usual course of learning until the sixth day when the maximum speed record was reached, but at that point, for some unknown reason, a retarding factor occurred that caused a decided rise in the curve from then to the end of the test. There are two possible explanations for this phenomenon; first, that at the lowest point of the curve the mice had reached their maximum speed and efficiency and thereafter they

TABLE III

CONTINUED RECORDS FOR THE WHITE FAMILY FOR THE MAZE AND MULTIPLE CHOICE TESTS

No.	Records for Maze Test												Multiple Choice Test					
	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Last 5 Trials	Error Average Last 15 Trials	Inter-ference 2 Trials	Error Average	Inter-ference Next 10 Trials	Error Average	Reten-tion 10 Trials	Error Average	First 2 Trials	Next 5 Trials	Last 18 Trials	Last 23 Trials	Last 5 Trials	Error Average Last 23 Trials
135W ♂	99	29	6	14	3	.9	73	7.0	10	.5	17	.9	48	7	63	51	15	3.7
136W ♂	67	22	7	11	6	1.0	50	2.0	11	1.0	4	0	31	14	10	11	9	2.3
137W ♀	94	38	10	20	8	1.3	92	4.5	14	1.2	4	.1	49	12	7	8	8	1.7
138W ♀	304	110	117	152	62	3.6	194	4.5	60	2.4	55	2.9	57	137	10	38	8	3.9
139W ♂	73	23	7	13	6	.4	20	1.5	20	1.2	74	2.5	43	113	22	42	16	2.6
140W ♂	76	32	13	19	12	1.4	58	4.5	10	1.0	17	1.1	45	15	20	19	20	3.3
141W ♂	172	61	20	34	16	1.5	52	2.0	22	1.5	22	1.0	49	40	33	35	33	3.7
142W ♂	50	40	33	36	30	1.9	24	1.5	36	2.0	33	.9	94	39	34	35	32	3.1
143W ♀	214	62	164	130	218	3.5	150	6.5	71	2.4	72	2.9	66	16	143	115	50	5.3
144W ♀	62	8	54	38	62	1.6	51	3.0	52	1.9	23	.8	60	9	12	11	6	2.0
145W ♂	360	360	49	57	12	3.7	38	3.5	27	1.8	Died		84	67	109	100	135	6.3
146W ♂	77	21	7	12	4	.6	10	1.0	15	1.1	Died		186	105	52	63	33	5.5
147W ♀	360	360	360	360	360	8.2	360	5.0	360	6.9	360	9.6	360	177	110	159	25	11.2
148W ♀	66	32	23	26	16	1.2	24	1.5	55	3.1	141	5.0	72	16	78	65	35	7.6
150W ♂	115	20	7	11	6	.7	74	4.0	196	4.3	153	1.4	91	94	35	48	42	4.5
151W ♀	36	20	14	16	7	.6	54	3.5	41	1.7	22	.6	139	37	28	30	44	3.3
152W ♀	141	38	51	47	33	2.2	18	2.0	9	.9	20	1.4	44	20	31	29	21	3.5
153W ♀	121	91	113	106	111	3.3	44	2.5	42	2.2	Died		46	16	43	37	35	4.5
154W ♀	192	183	118	140	80	2.6	50	2.0	33	1.6	30	1.3	40	14	45	39	31	4.3
155W ♂	59	73	13	33	16	1.4	25	1.5	38	.6	81	1.7	26	13	121	98	100	5.3
157W ♀	95	33	70	57	124	2.3	33	1.5	156	4.5	101	4.1	35	17	49	42	56	3.8
158W ♀	108	45	11	22	8	.7	21	1.5	66	3.3	66	2.9	46	59	40	44	88	4.0
159W ♀	146	52	20	31	26	1.5	60	2.5	46	2.6	59	2.7	48	70	34	42	38	4.5
160W ♀	64	128	69	88	20	2.2	149	6.0	73	3.1	107	3.9	35	101	81	85	126	5.3
161W ♂	176	29	17	21	12	1.3	138	7.5	16	.5	9	.1	28	21	9	12	5	2.2
162W ♂	75	29	40	37	16	2.2	152	6.0	77	2.8	81	3.6	118	67	94	88	36	4.9
163W ♂	147	40	11	21	11	.7	49	3.5	28	2.1	10	.7	23	15	11	12	6	2.6
164W ♀	43	17	7	10	7	.9	114	5.5	41	1.8	14	1.2	65	8	15	13	6	2.3

TABLE III (*Continued*)

No.	Records for Maze Test												Multiple Choice Test					
	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Last 5 Trials	Error Average Last 15 Trials	Inter-ference 2 Trials	Error Average	Inter-ference Next 10 Trials	Error Average	Reten-tion 10 Trials	Error Average	First 2 Trials	Next 5 Trials	Last 18 Trials	Last 23 Trials	Last 5 Trials	Error Average Last 23 Trials
165W ♀	154	32	140	104	240	4.0	346	11.5	67	3.0	24	1.2	43	28	23	24	21	2.6
166W ♀	84	21	18	19	29	.6	32	2.0	52	2.7	24	1.7	44	22	21	21	15	3.1
171W ♂	346	19	214	149	194	2.3	339	12.0	53	2.5	8	.8	26	9	13	12	9	2.6
174W ♀	352	86	43	58	51	.7	52	3.5	Died									
175W ♀	257	54	266	195	326	4.5	360	9.5	184	5.3	40	1.9	48	15	102	83	15	6.2
176W ♂	70	71	21	38	7	1.3	37	3.0	13	1.3	23	1.3	43	14	7	9	11	1.7
177W ♂	282	82	30	47	13	1.5	42	3.5	11	1.1	15	.6	52	24	23	23	40	3.4
178W ♀	277	227	111	150	50	2.0	91	3.5	92	2.3	238	3.0	44	23	53	47	124	3.0
180W ♂	280	61	15	30	8	1.7	23	2.0	27	2.0	Died		29	13		Died		
181W ♂	190	180	85	116	52	3.1	131	5.0	32	1.7	9	1.2	260	18	6	9	4	1.9
182W ♂	65	19	9	12	4	.3	56	5.0	45	2.1	15	1.1	48	4	9	8	6	1.9
183W ♂	86	34	16	22	13	1.2	49	2.5	23	1.0	Died		87	74	23	34	23	3.3
185W ♂	112	23	20	21	9	1.1	33	2.0	48	2.0	Died		47	20	17	18	15	3.2
186W ♂	150	40	18	25	14	1.8	32	1.5	68	1.5	46	1.9	35	13	18	16	7	2.2
187W ♀	360	136	101	113	81	3.7	141	9.0	65	3.5	198	10.3	72	6	21	18	36	2.6
188W ♀	95	106	184	158	221	3.3	197	10.5	149	5.5	11	.6	46	39	7	14	6	2.1
189W ♀	113	46	29	35	26	1.9	275	12.5	60	2.9	34	2.2	122	20	20	20	9	2.8
190W ♂	53	13	6	9	8	.8	57	5.0	12	1.0	24	1.4	55	30	6	11	2	1.4
191W ♂	58	39	27	31	42	1.4	67	4.0	19	1.2	78	4.2	48	34	20	23	14	2.4
192W ♂	219	36	13	21	19	.8	72	5.5	48	2.6	15	1.1	111	146	29	55	27	4.2
193W ♂	81	18	14	15	14	1.2	14	1.5	17	1.1	12	.6	285	24	14	16	14	1.8
194W ♀	82	21	11	14	14	.7	18	1.5	19	2.0	21	.9	273	135	30	53	81	3.0
195W ♀	140	290	271	277	312	6.6	221	4.5	360	5.7	343	7.0	320	116	70	80	78	5.5
200W ♂	90	78	12	34	10	1.3	25	1.0	15	1.4	10	1.0	43	28	13	16	9	3.2
201W ♀	47	93	153	133	85	2.6	76	2.5	28	2.2	99	2.7	87	12	16	15	16	1.6
202W ♂	154	76	31	46	25	1.2	106	7.0	17	1.3	11	.3	34	27	11	14	12	2.5
203W ♂	90	125	24	58	11	1.6	168	12.5	9	.9	31	1.7	38	20	14	15	18	2.3
204W ♂	206	73	45	54	53	1.4	100	4.5	37	2.3	12	1.0	30	16	14	14	15	3.3

TABLE III (*Concluded*)

No.	Records for Maze Test												Multiple Choice Test					
	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Last 5 Trials	Error Average Last 15 Trials	Interference 2 Trials	Error Average	Interference Next 10 Trials	Error Average	Retention 10 Trials	Error Average	First 2 Trials	Next 5 Trials	Last 18 Trials	Last 23 Trials	Last 5 Trials	Error Average Last 23 Trials
205W ♀	279	186	217	207	142	3.8	240	5.0	133	2.8	13	1.0	55	135	51	70	16	4.5
206W ♀	233	94	139	124	233	4.0	49	2.5	98	3.4	62	4.5	28	33	14	18	11	2.4
207W ♀	213	172	192	185	280	4.8	204	8.1	75	3.0	85	5.4	44	18	19	19	14	2.8
208W ♂	250	174	40	74	31	2.1	34	1.5	8	.9	Died		41	26	50	44	76	2.8
209W ♂	275	218	61	113	73	2.7	189	4.5	9	.6	9	.9	19	16	11	12	10	3.1
210W ♀	46	271	130	244	164	1.0	94	3.5	36	2.3	20	1.5	69	22	31	28	53	2.8
211W ♀	239	150	165	160	82	2.8	20	.5	12	.9	12	.8	28	18	69	58	119	3.0
212W ♀	194	166	128	141	153	2.5	26	1.0	14	1.2	28	1.1	29	16	19	18	16	2.2
213W ♀	92	10	13	12	20	.2	26	1.5	11	1.6	8	1.2	40	12	22	20	21	2.6

became "tired" of the problem and lost their adjustment, or in other words, the test had been carried too long; secondly, the following explanation may be given: the drop in the sixth day of the test is

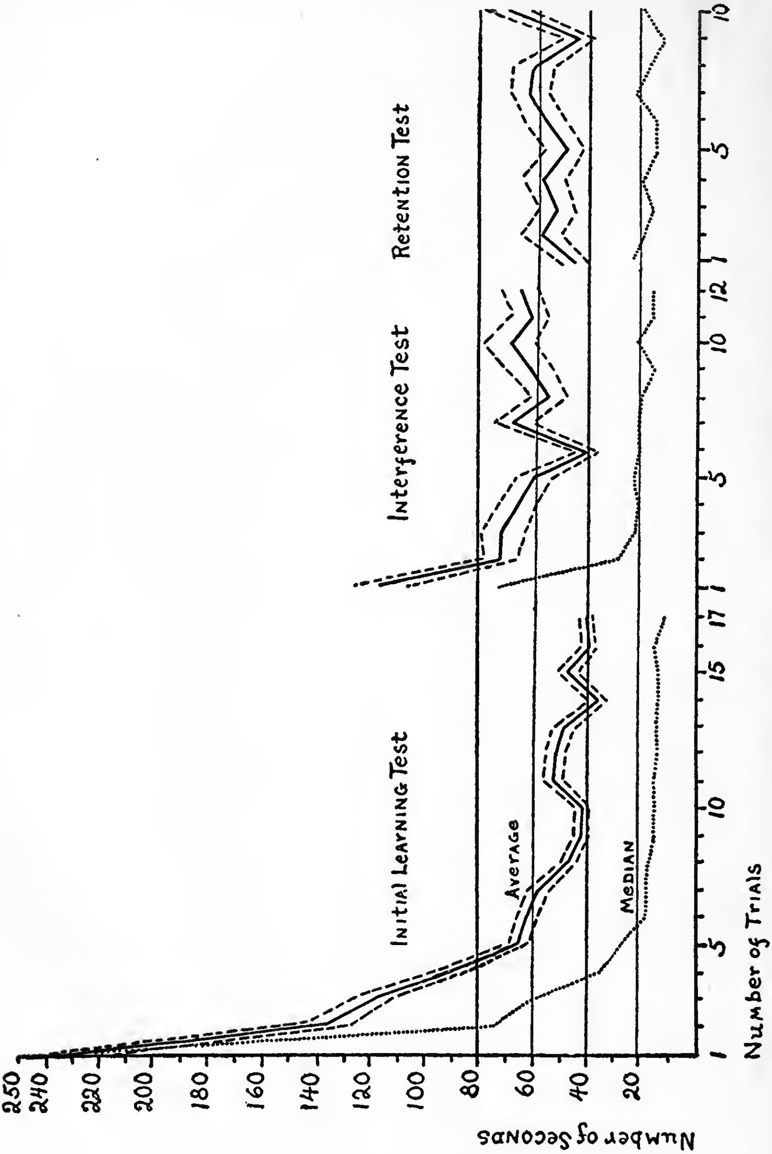


FIG. 4. Complete Record Curves for Mice in the Maze Test, showing curves based upon the average and the median for each day. (The limits of the probable error for the average curve are indicated by the broken lines.)

not significant and represents a chance irregularity which would disappear from the curve if a still greater number of animals was used. As there is no evidence to support the first of these explanations, it is probable that the second solution is the correct one. We can, therefore, say that the interference as shown in this experiment was not a transitory thing, being evident for only one or two trials after its application, but instead, it exerted an influence over a number of trials and prevented the animals from attaining the same degree of proficiency that they had previously shown in a similar task. However, if the interference test had been continued it is possible that the curve would have reached as low a record as that given by the initial learning test.

TABLE IV

COMPLETE RECORDS OF A FAMILY CONSISTING MOSTLY OF YELLOW INDIVIDUALS IN THE MAZE TEST

No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average	No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average
32Y♂	225	109	71	83	1.7	61Ch♂	360	202	30	87	1.0
33Y♂	154	186	42	90	.5	62YW♂	312	90	18	42	1.5
34Y♂	186	88	39	56	1.0	63YW♀	243	130	38	69	1.2
36Y♂	354	55	28	37	.9	64Y♀	360	182	113	136	2.3
37Y♀	137	20	16	17	1.2	68YW♀	112	71	65	67	2.0
54W♀	360	36	34	35	.7	69Y♂	177	41	9	19	.7
55Y♂	360	242	103	150	3.5	70Y♀	234	223	75	124	2.3
56Y♀	360	130	103	112	2.0						

TABLE V

COMPLETE RECORDS OF A SMALL FAMILY SHOWING GOOD RECORDS IN THE MAZE TEST

No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average	No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average
29W♀	229	9	8	8	1.3	45W♀	142	13	5	8	.5
30Gr♂	58	43	33	36	1.0	46Gr♂	141	9	4	6	.5
44W♀	297	16	7	10	.5	47Bl♂	150	34	29	31	1.0

It is interesting to note that Hunter and Yarbrough found that interference occurred between an old habit and the formation of a new one, in their study of the auditory habits in the white rat.[9] The following is quoted from their results: "Habit interference occurs in the white rat between a first habit and the formation of the second one." "Interference is most marked between the end of the perfected habit and the beginning of the new habit." "Habit interference may serve greatly to slow up the formation of a new habit." The results of this investigation agree, in the main, with the statements that have just been quoted, but, although the interference effect caused a sharp rise in the time curve, and the interference was

[9] Walter S. Hunter and Jas. N. Yarbrough, "The Interference of Auditory Habits in the White Rat," *The Journal of Animal Behavior*, Vol. 7, No. 1.

evident throughout the subsequent trials, yet it did not appear to slow up the formation of a new habit to as great a degree as might be expected from the nature of the behavior that was tested. The quick recovery from the interference effect suggests that the behavior of the animals of this investigation shows a considerable amount of flexibility. As was previously stated, the animals took only half as much time to go through the maze on the day when the interference was given as they did on the first day they were tested, and this shows that an accelerating transfer effect from the previous training was operating to counteract the slowing up of the interference.

In order to more carefully analyze the behavior of the mice the median record was calculated for each day's performance, and a curve based upon the same is represented by the dotted lines in Fig. 4. It may be noted that for each test the curve based on the median falls considerably below the corresponding curve for the average. The essential character of the curves, found by these different methods, is the same; except that the curve based upon the median is more regular than the one found from the average. In the interference test the curve for the median records is much more regular than that of the average curve for the same test, while no disturbance at all is to be noted at the sixth day of that test. The character of the curve based on the median supports what has already been said concerning the permanence of the interference effects over a number of trials.

The retention test shown in Fig. 4 was given immediately after the mice had been tested in the multiple choice test, and may represent a certain amount of training acquired there. The average time for the last ten trials of the interference test is $60.26 \pm$ P.E. 4.7, while the average time for the ten trials in the retention test is 52.81 $\pm$ P.E. 4.7. The superiority of the average retention test in time is nearly twice its P.E., indicating a fair reliability. The superiority of the retention test may be due to the presence of one or more of the following conditions: (1) A mere carrying over of capacity attained in the interference test. (2) The dying out of bonds developed in the negative test (which in this case was the initial learning test), and the strengthening of bonds developed in the interference test. (3) The transfer of capacity developed in the multiple choice test. Now 1 is not likely because the average record made in the retention test was much better than that in the interference test even after a considerable interval of time. It is possible that the condition in 2 may account for the facts, but there is no direct con-

TABLE VI

COMPLETE RECORDS OF A FAMILY COMPOSED OF COLORED INDIVIDUALS FOR THE MAZE AND MULTIPLE CHOICE TESTS

No.	Records for Maze Test												Multiple Choice Test					
	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Last 5 Trials	Error Average Last 15 Trials	Interference 2 Trials	Error Average	Interference Next 10 Trials	Error Average	Retention 10 Trials	Error Average	First 2 Trials	Next 5 Trials	Last 18 Trials	Last 23 Trials	Last 5 Trials	Error Average Last 23 Trials
99Y ♀	77	10	7	15	4	.3												
100Y ♂	310	46	25	32	10	1.5												
101AG ♀	201	41	17	25	6	1.8												
102GR ♀	76	17	85	63	159	2.2												
128AG ♂	122	14	3	7	1	.6	12	1.0	4	.5	8	.8	37	19	34	31	19	4.1
129AG ♂	178	23	5	11	4	.6	6	.5	8	.6	7	.9	41	9	46	38	10	4.3
130AG ♂	124	24	4	11	2	.8	17	2.0	25	1.7	6	1.0	244	42	18	23	6	2.9
131AG ♂	360	19	3	8	3	.7	16	2.0	3	.2	5	.6	55	9	8	8	4	1.4
132AG ♀	187	34	4	14	3	.9	36	4.0	4	.9	16	1.3	360	250	13	65	5	3.6
133AG ♀	268	143	13	56	11	1.3	47	2.5	40	2.1	89	2.9	191	63	18	28	10	3.6
134AG ♀	200	18	2	7	1	.5	20	2.0	19	1.5	76	4.2	360	271	280	278	138	7.2
167Y ♂	221	161	143	149	110	2.7	51	4.5	148	4.3	17	6.8	25	21	7	10	6	2.3
168Y ♀	305	70	71	71	72	2.3	19	2.0	151	4.3	62	3.3	43	87	34	46	5	4.4
169B1 ♂	146	53	128	103	159	2.5	95	6.0	38	1.6	38	2.6	55	28	7	11	6	2.1
170GR ♂	88	33	75	61	73	1.5	154	7.5	121	3.4	27	1.5	29	12	6	7	3	1.7
196GR ♂	309	82	13	36	10	1.0	172	5.0	68	2.6	11	.4	55	18	52	44	87	3.3
197GR ♀	254	42	13	23	13	1.7	78	6.0	134	4.3	65	3.0	154	22	20	21	18	3.4
198Y ♀	346	231	19	90	11	1.8	58	6.5	115	4.0	170	6.8	96	45	196	163	207	6.7
199AG ♀	166	28	10	16	13	1.2	15	2.5	9	1.2	Died							

firmatory evidence from the data, and it must be remembered that the original test, although older than the interference test, still received a greater amount of repetition. It appears from the results of the experiment that the condition in 3 is most likely to account for

TABLE VII

Complete Records of the Unrelated Individuals in the Maze Test

No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average	No.	First 2 Trials	Next 5 Trials	Last 10 Trials	Last 15 Trials	Error Average
1Ch ♂	212	48	11	23	1.0	5W ♀	74	22	8	12	.6
2Y ♀	285	213	103	140	5.1	28W ♂......	291	38	5	17	1.0
3Y ♀	316	55	63	61	3.3	31Gr ♀	130	78	15	38	1.0
4W ♂	77	18	14	16	.7	85W ♂......	131	14	9	11	.8
						90Ch ♀	202	81	13	35	1.0

the facts; that is, the transfer of capacity developed in the multiple choice test accounted for the superiority of the retention test, by virtue of the better adjustment that the animals received to the experiment as a whole, and by practise in the elimination of fruitless movements. The number of seconds required for the average mouse to complete the first trial of the retention test is below all but one of the records that the same mice made in the interference test, and it is noteworthy that the curve for the retention test based on the daily average, is not similar to either of the preceding curves, but is nearly flat, except for a slight rise at the last trial. Since the average animal did not start with a high time record, it is evident that it did not need to learn the task all over again, but showed a considerable amount of permanence of association for the previously learned task.

The curve for the retention test, based on the median record for each day, confirms, in the main, what has already been said concerning this test. The curve for the median is also flat, the four high points all reaching to about a score of twenty seconds, and it again shows the relative superiority in the record for the first day of the test. Also, taken day by day, the records for the retention test, with a single exception, show the retention test with daily records superior to the corresponding ones of the interference test.

(*d*) Results for Multiple Choice Test

Fig. 5 gives complete record curves for the same 71 mice as tested by the multiple choice, and the limits of the probable error, for the curve based on the daily average, are indicated in the same manner as explained for the previous curves. The average time for the first trial in the maze test is 236 seconds as compared with 91 seconds as the average for the first trial of the multiple choice test. How much this difference is due to what the average animal acquired in the first experiment can not be determined since the tests themselves are markedly different.[10] In the first place, the distance to be traversed

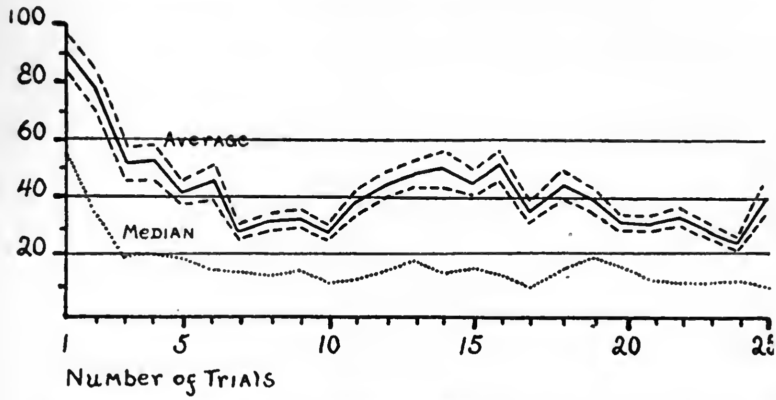


FIG. 5. Complete Record Curves for Mice in the Multiple Choice Test, showing curves based upon the average and the median for each day. (The limits of the probable error for the average curve are indicated by the broken lines.)

in the multiple choice test is much shorter than in the maze test, and the intensity of punishment is greater in the former case because there the electric shock was used. Also color clews, and the fact that it was necessary for the animal to go through only one door, tend to lessen the average time in the multiple choice test. As indicated in the previous discussion, there was no doubt a transfer effect in learning from the first task to the second, that would again result in shortening the time in the multiple choice test. This was due; first, to a better general adaptation to the experiment as a whole, which was carried over from the maze test; and secondly, to the fact that an important element was common to both tasks, namely, that in each case the animal learned to escape from confinement by means of a door that could be pushed open. It is interesting to note at this point that Yerkes in his study of the dancing mouse[1] (see page 263) found that experience in one labyrinth made the learning in a second labyrinth much easier. "Those individuals whose first labyrinth training was in (labyrinth) *C* made their first correct trip as the result of 19.7 trials, whereas those which had previously been trained in labyrinth *B* were able to make a correct trip as the result of only 7.0 trials. Similarly the table shows that training in *C* rendered the

[10] This point is being more closely studied in an experiment, now under way.

subsequent learning in *B* easier." The average curve in Fig. 5 indicates that there is practically no increase in learning between the seventh trial and the twenty-fifth. A fairly uniform decrease in time occurs from the first to the seventh day, when an average speed of twenty-six seconds was made. The only other performance that bettered this record was made on the twenty-fourth day, when an average time of twenty-four seconds was recorded. The number of trials in the multiple choice test could have been much less and still sufficient for the average mouse to learn the maze. It was impossible to determine that more trials were given than were necessary until the experiment was well started. In fact, the experimenter arbitrarily chose twenty-five as the number of trials because he thought it would take the average mouse longer to learn the second task than the first. As explained above, the results of the experiments disapproved this assumption.

In the multiple choice test, the curve based on the median record for each day, as in the previous tests, is considerably below the corresponding curve for the average. The curves here are essentially similar, and it may be noted that the curve for the median, after the second trial, never rises higher than twenty seconds or lower than ten.

VI. CORRELATIONS IN LEARNING RECORDS

VARIOUS correlations have been calculated for performance in one task with performance in another, and between groups of trials within a single task. The correlation have all been positive, varying from 0.11 to 0.85 as described below. A complete list of the correlations is given in Table VIII. The Pearson formula,

TABLE VIII

RESULTS OF CORRELATIONS

No.	Things Correlated	Correlations
1.	Time in last 15 trials of maze test with error average in same task	+0.85
2.	Time in first group of 5 trials in maze test with the time in last group of 5 trials in maze test	+0.46
3.	Time in last 15 trials of maze test with the time in retention test.	+0.35
4.	Time in last 15 trials of maze test with the time in the first 2 trials of the interference test............................	+0.55
5.	Time in last 15 trials of maze test with the time in the last 10 trials of the interference test............................	+0.49
6.	Time in last 23 trials of Multiple Choice test with error average made in same test....................................	+0.82
7.	Time in first group of 5 trials in Multiple Choice test with the time in last group of 5 trials in Multiple Choice test........	+0.25
8.	Time in last group of 15 trials in maze test with the time in last group of 23 trials in Multiple Choice test.................	+0.11

$$r = \frac{\Sigma(x \cdot y)}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma y^2}},$$

was used for the correlations numbered in the table as 1, 2, 6 and 7. Because the remaining correlations, 3, 4, 5 and 8 were made between two groups both from asymmetrical distributions, the ranking method was used with the following formula:

$$r = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}.$$

The ranking method was employed so that undue weight would not be given to the few extreme cases in the skewed distributions. This disadvantage did not occur in the correlations made, between time and error, and performance at the beginning and end of the tests, so in these cases the Pearson formula was used.

The correlation between the time in the initial learning period in the maze test with the errors made during that performance amounted to + 0.85. A similar coefficient of correlation of + 0.82 was found to hold between the last 23 trials of the multiple choice test and the errors made in that task. These results are what one might naturally expect to find, namely, that the animal that made many errors required more time. The conclusion need not hold, however, for the type of behavior that this investigation deals with, because it might very well be that a stupid animal was one likely to sit in a corner of the maze and make a poor time record, but a good error record. These two high correlations are thus seen to give a measure of the activity of the animals in each task, and they show the value of the tests that were used, in that each was adapted to the instinctive behavior of the animals tested.[11] The native equipment of the mice no doubt adapted them to finding their way through narrow passages and forcing their bodies through small openings.

A low positive correlation of 0.11 was found to hold for the relation between the initial learning period in the maze test and the last 23 trials in the multiple choice test. As previously discussed, the time values in the two tasks that were correlated are not comparable, and this fact may account for the lowness of the coefficient of correlation.

A correlation was made between performance at the beginning and close of the trials, and in this case the first group of five trials was correlated with the last group of five. For the maze test this correlation amounted to + 0.46, which represented a fairly strong correlation. A similar correlation between the first group of five trials and the last group of five trials in the multiple choice test gave a lower correlation of + 0.25.

The correlation between the initial learning period in the maze test and the retention test for the same task was found to be + 0.35. This indicated a fairly close relation between proficiency in a given task and the amount of association retained. A similar correlation between the same initial learning period in the maze test and the first

[11] Guinea pigs have been used as subjects of another investigation and when tested in the maze, that apparatus was found to be very well suited to their instinctive type of behavior. They have not as yet been tried in the multiple choice apparatus which was used in these experiments. The maze test, and the multiple choice as well, were found to be unsuited to the average native equipment of a litter of seven English Bulldog puppies that also have been tested. Four out of the seven dogs failed completely to learn the task, sitting in a corner of the maze and barking disconsolately at the gates. They held back from passing through an opening through which an average mouse would at once venture. Two of the dogs learned very slowly, while one made a good record. The dogs were more interested in looking for the experimenter and responding to the slightest noise he would inadvertently make, than in paying attention to the maze problem itself.

two interference trials was + 0.55; and for the same period in the maze test with the last ten trials of the interference test the correlation amounted to + 0.49.

The above correlations tend to show: first, that the animal that does well in any one task is likely to retain more than one that does not do well; secondly, time and error are closely related in the types of behavior that this investigation deals with, and one is a measure of the other; thirdly, an animal that does well in the beginning of a task is more likely to do well at the end that an animal that is slow in learning; fourthly, the animal that did well in the maze task exhibited greater adaptability in behavior than one that did not do well, as shown by a comparison of the interference records.

VII. FAMILY RESEMBLANCES

In the preliminary report of these experiments it was found that the mean variation of the entire group of 90 mice, including mice from all the strains that had been tested up to that date, amounted to 35.6. This means that any mouse picked at random from that group would be likely to vary from the average by 35.6 seconds. In order to find whether mice of the same litter vary less than unrelated individuals, the mean variations for each of the eighteen families was calculated, and these, when weighted for size of family, were found to be 20.2. It was, therefore, stated that the resemblance in behavior between mice belonging to the same litter was nearly twice as great as between unrelated individuals, and that this corresponded to a coefficient of correlation in the neighborhood of 0.5 for brothers, as found by Pearson, Thorndike and others. It is probable, however, that in this case the mean variation was increased for the unrelated individuals, and the family resemblance correspondingly made to appear greater, due to the fact that the mixed group was made up of two fairly distinct strains of mice, a quick white group and a slow yellow group.

As previously stated the above calculation was made from the results in only one task, namely, the initial learning period of fifteen trials in the maze test. When the mean variations were calculated for the four distinct tests that were later used, and in which considerably more animals were tested, there was apparently no greater resemblance, judged by this method of variation, between animals of the same litter than between unrelated individuals. In order to discount the inevitable reduction in mean variation that resulted from the grouping of the individuals into family lots, in which case the mean variations were calculated from the family averages, all the mice that belonged to the family groups were rearranged according to chance into other groups of exactly the same size. Then the mean variations for the groups formed by chance arrangement were calculated in exactly the same manner as was previously done for the true family groups. A comparison of the two results showed a strict similarity in mean variation in each of the four tests; so that the mean variation for the groups of related individuals was not less than, but practically identical with the variation that obtained for the same individuals grouped by a chance arrangement. Later in this thesis, when a detailed study of the family histories is taken up, the records for the various litters will give further light on this subject of family resemblances, and individual variations.

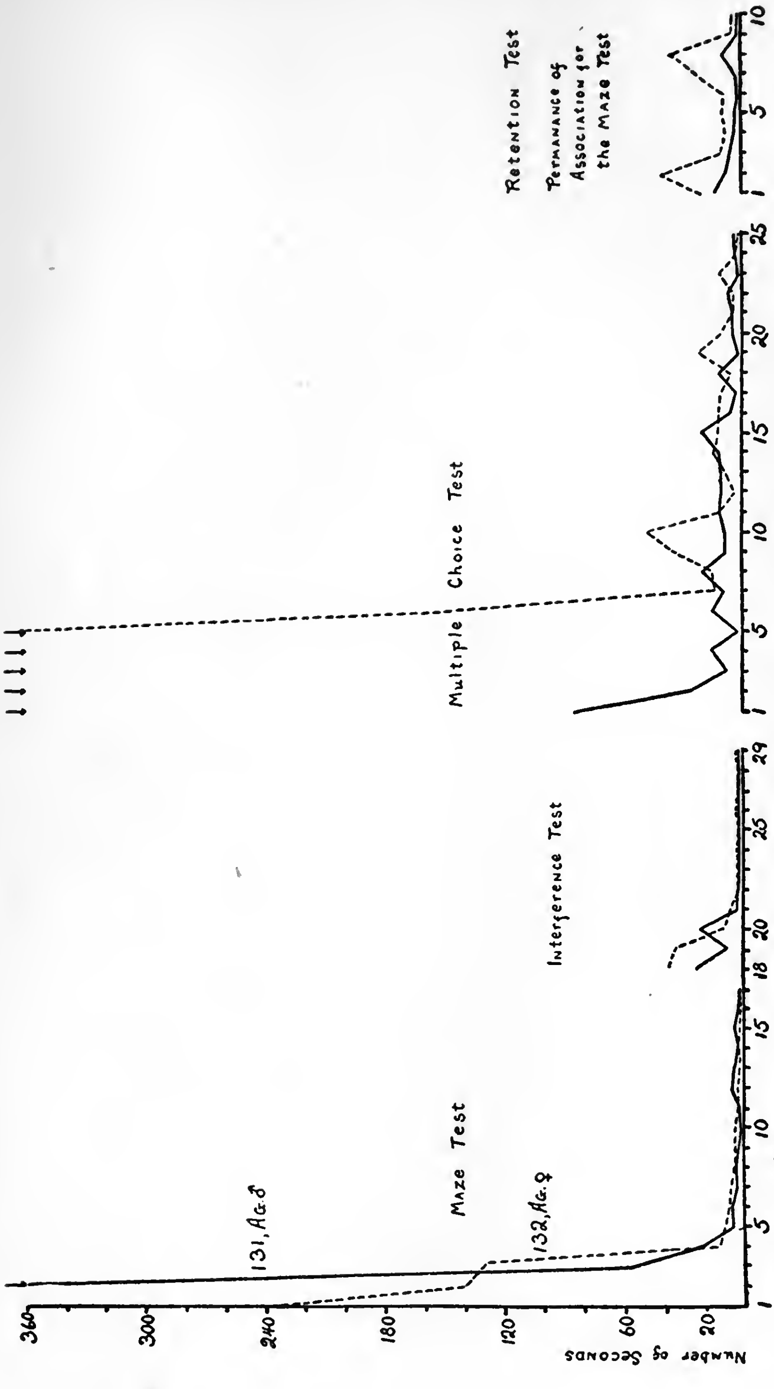


FIG. 6. Daily Record Curves for 131, Agouti ♂ and 132 Agouti ♀.

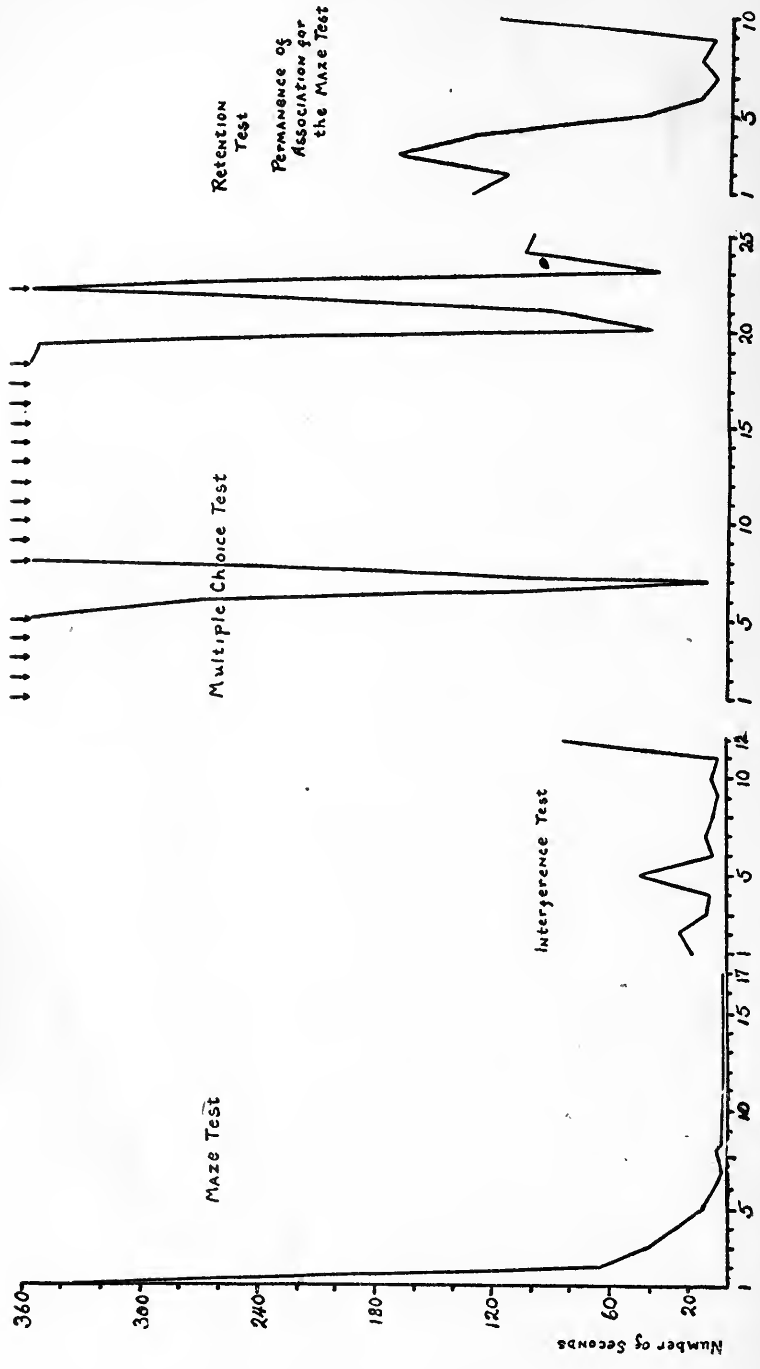


FIG. 7. Daily Record Curves for 134, Agouti ♀.

VIII. SEX DIFFERENCES

(*a*) Differences in Time and Error Records

In Table IX. the males and females are grouped separately, and their average times and errors are given for the various tasks that were used. The preliminary report of this investigation showed that there were small sex differences for performance in the initial learning period in the maze test, and the differences that did exist were

TABLE IX

Averages for Sex Differences and Sex Variability

Tests	No. and Sex	Average No. of Secs. per Trial	Mean Variations in Seconds	Probable Error	Average No. of Errors per Tria
Initial Learning Period, Maze Test ..	93 ♂	43.80	33.5	±2.9	1.2
	90 ♀	64.77	53.3	±4.7	1.7
Interference Test, 10 Trials.........	34 ♂	37.76	29.0	±4.3	1.6
	37 ♀	80.96	56.8	±8.0	2.8
Retention Test	34 ♂	27.58	21.8	±3.2	1.4
	37 ♀	76.00	59.1	±8.3	2.9
Multiple Choice Test	34 ♂	25.76	17.0	±2.5	2.8
	37 ♀	52.08	35.1	±4.9	3.9

well within the limits of the probable error. With a larger number of animals, 93 males and 90 females, it was found that in the initial learning period in the maze test, the males did considerably better than the females. The males made an average time per trial of 43.80 ± P.E. 2.9 seconds, and 1.2 error per trial, while the females made a corresponding record of 64.77 ± P.E. 4.7 seconds and 1.7 error per trial. As previously stated, there were 71 animals that were tested in the interference, retention and multiple choice tasks, and of these 34 were males and 37 females. As indicated in the table for sex differences, in all three of the above mentioned tests the females took, on the average, twice as many seconds to learn the tests and made considerably more errors per trial than did the males. In the interference test of ten trials the males made an average time of 37.76 ± P.E. 4.3 seconds and 1.6 error per trial, while in the same task the females took 80.96 ± P.E. 8.0 seconds and 2.8 error per trial. The retention test records gave a similar result; the males making an average of 27.59 ± P.E. 3.2 seconds per trial and 1.4 error per trial, and again the females took more than twice as long to perform the same task, namely, 76.00 ± P.E. 8.3 seconds and 2.9 error per trial. In the multiple choice test the males took 25.76 ± P.E. 2.5 seconds and 2.8 error per trial, while the females made an average of 52.08 ± P.E. 4.9 seconds and 3.9 error per trial. The final average of the males for all four tests amounted to 33.72 ± P.E. 4.9 seconds per trial, while that of the females was 68.45 seconds. This shows that the females in the total record took slightly more than twice as long to learn the tests as did the males.[12]

(*b*) Comparative Distributions of the Males and Females

In Fig. 8 are given the comparative distributions for the 34 males and 37 females that have been tested in all four of the tasks for which time averages are given in Table IX. The distributions for the males and females, in the initial learning period of the maze test, show that the curve for the males is skewed, for, although there is an equal number of individuals, namely, 12, between 0 and 20 seconds, and 20 and 40 seconds, still from that point on the curve drops quickly, while no individuals are to be found beyond 160 seconds. The distribution curve for the females shows that the mode for that sex is about the same as that found for the males, also that several females are distributed toward the right end of the figure, which indicates that they were the animals that made poor time records. Eight females are to be found between 0 and 20 seconds, and while the curve for this sex drops steadily until 80 seconds is reached, still it rises steadily from that point until it reaches a maximum in the region between 140 and 160 seconds, where five females are to be found. This almost gives the curve a bimodal appearance, but there is no evidence at hand which leads the writer to believe that there were two distinct classes of females among the mice that were tested. The important point to be noted in Fig. 8 is that there were six females that exceeded any of the records made by the males; there

[12] It might be expected that this sex difference would have its effect on the validity of the correlations that have already been given, since they show that the total population was made up of two groups, one, the males, which were relatively quick learners, and the other the females, relatively slow learners, but the accompanying correlations that were made for the separate sexes show that they follow rather closely the combined correlations that have been given in Table VIII. Correlating the last 15 trials of the maze test with the first two trials of the interference test gave the following correlations: ♂s = + .60, ♀s = + .54. The correlations for the same 15 trials of the maze test with the last 10 trials of the interference test gave: ♂s = + .35, ♀s = + .49. Again the same maze trials when correlated with the retention test gave the following correlations: ♂s = + .18, ♀s = + .27. The last sex correlation was made between the last 15 trials of the maze test and the last 23 trials of the multiple choice test, and here the following correlations were found: ♂s = — .26, ♀s + .16.

were two females between 180 and 200 seconds; one between 200 and 220 seconds; one each between 240 and 260, and 260 and 280 seconds; and finally one female that failed completely.

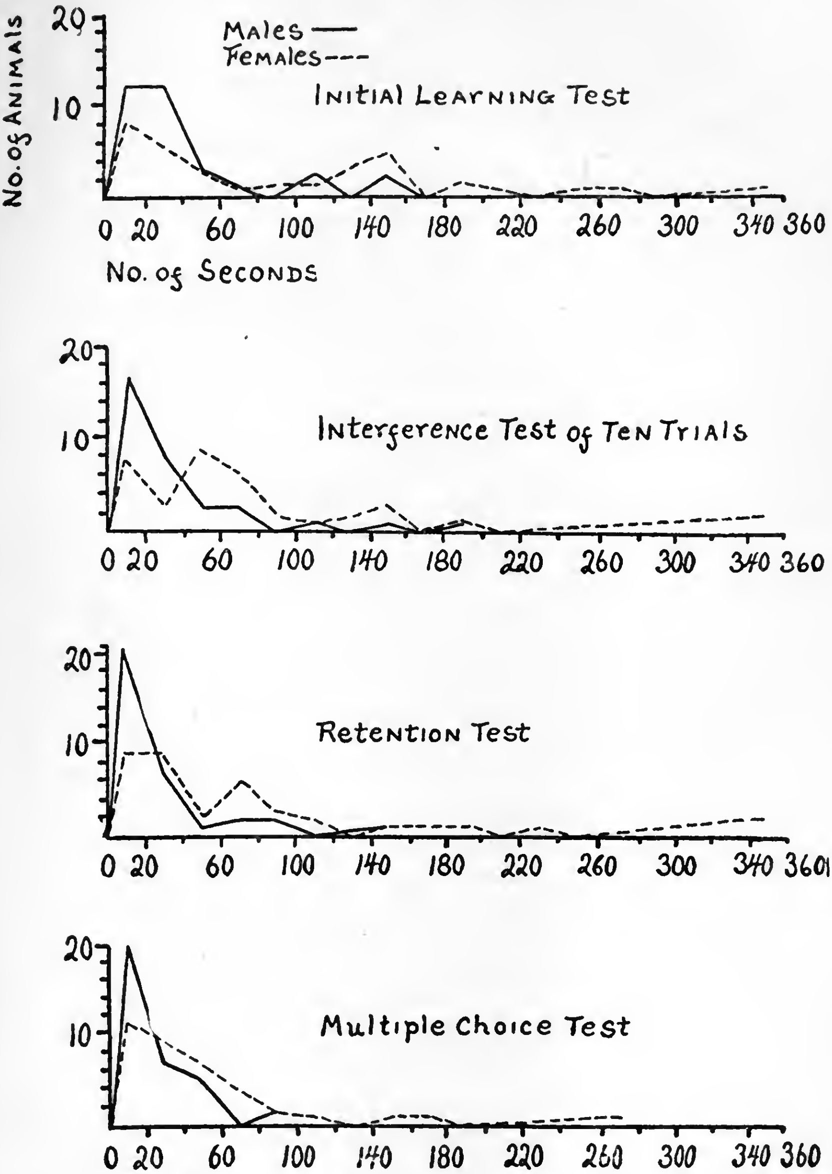


FIG. 8. Distribution Curves for the Males and Females for all Tasks. In each case the number of animals is given by the ordinate, while the abscissae represent the number of seconds The males are represented by the solid lines and the females by the broken lines.

The distribution curves for the males and the females, for the interference test of ten trials in the maze, are also given in Fig. 8. Here it is to be seen that the curve for the males is distinctly skewed. Seventeen males are grouped between 0 and 20 seconds; eight between 20 and 40 seconds, and from there on the curve is similar to the one made by the same animals in the initial learning period of the maze test, but in the present case no males are to be found that took, on the average, longer than 200 seconds to complete the test. This figure shows that the mode for the females is to the right of the mode for the males; the greatest number of females, namely nine, are to be found between 40 and 60 seconds. The curve for the females is irregular, again showing a superiority in favor of the males, while here two females failed to learn the test.

The distribution curves of both sexes for the retention test show that the curve for the males is similar to the one above that we have just examined. It is skewed, with the mode between 0 and 20 seconds, where twenty-one animals are to be found, while no records for the males exceed 160 seconds. The curve for the females is less skewed than that of the males; an equal number of individuals, namely nine, are to be found between 0 and 20 seconds, and 20 and 40 seconds; the curve then drops quickly, but rises again in the region between 60 and 80 seconds, where there are six animals. Then the curve for the females gradually falls, but again shows the same characteristic that is to be found in all the comparative distributions, namely, that a number of females are distributed beyond the maximum records made by the males in the same test. In this case five females made longer average time records than the males, and two of them failed to learn the test.

The distribution curves for the males and females in the multiple choice test are given at the bottom of Fig. 8. Here the mode for both sexes is to be found between 0 and 20 seconds; but the curve for the males is steeper than the corresponding one for the females; twenty males being found between 0 and 20 seconds; seven between 20 and 40 seconds; five between 40 and 60 seconds, and finally two between 80 and 100 seconds, which is the maximum time record for the males. The distribution curve for the females shows eleven individuals between 0 and 20 seconds, and then, as shown in the figure, the curve drops at a regular rate; there being nine individuals in the next time group, seven in the next, four in the next, and two in the next, where the curves for both sexes meet. It is to be noted, however, that the same thing occurs here that was found in the three previous distributions, namely, that the females exceeded the maximum time records of the males that were tested with them at the same time. In this case one female is to be found in each of the following time regions: 100 to 120 seconds; 140 to 160 seconds; 160 to 180 seconds, and 260 to 280 seconds.

(c) Variability of the Sexes

An inspection of the probable errors of Table IX. indicates that the behavior of the females was much more variable than that of the males. The table gives in detail the variability of the sexes. In the initial learning period of the maze test the following record was made: mean variation males, 33.5 seconds; mean variation females, 53.3 seconds; while in the interference test of ten trials the mean variation for the males was 29.0 seconds; and for the females 56.8 seconds. The corresponding records in the retention test gave a mean variation for the males of 21.8 seconds and for the females 59.1 seconds. In the multiple choice test the mean variation for the males amounted to 17.0 seconds, while for the females 35.1 seconds. The average of the variations for all the tasks amounted to 25.3 seconds for the males and 51.0 seconds for the females. These figures show that the variability in the behavior of the females was about twice as great as that of the males.

The above results in sex differences and variations agree with those of Hubbert in her work on habit formation in the albino rat.[13] The following is quoted from her results: "The general averages for an equal number of males and females show the males superior to the females in all points save one, that of absolute time. They finished in fewer trials, required less total time, and covered a smaller amount of distance in learning the problem than did the females, while their speed was slightly higher. . . . The mean variation from the time average is less for the males at all ages, their distance variation is less at the age of sixty-five days and three hundred days."

Yerkes found that for the behavior of the dancing mouse in the black-white discrimination tests: "The males almost invariably acquired a perfect habit quicker than the females . . . (but) . . . in the labyrinth test the female is as much superior to the male as the male is to the female in the discrimination tests. . . . A degree of proficiency in labyrinth 'B' attained by the males after 7.0 trials was equaled by the females after 6.2 trials. In labyrinth 'C' the males acquired a habit as a result of 18.7 trials; the females, as a result of 13.8. And similarly in labyrinth 'D,' 6.1 trials did no more for the males than 2.9 did for the females."

[13] Hubbert: "The Effect of Age on Habit Formation in the Albino Rat," Behavior Monograph Series, No. 11, 1915.

The results of the present investigation agree with those found by Yerkes concerning his discrimination test, but do not agree with his results for the labyrinth test. The multiple choice used in the present experiments was primarily designed to test discrimination, while tending to destroy position habit, and so it may be considered comparable to the discrimination test used by Yerkes. Here there is agreement, in that the males did better than the females, but in the maze test, which corresponds to Yerkes' labyrinth, the results of this investigation agree with Hubbert's and disagree with those of Yerkes.

IX. INDIVIDUAL DIFFERENCES AND METHODS OF HABIT FORMATION

We may now take up in more detail a study of the records made by a few of the animals that show typical or exceptional types of behavior, and also discuss the observations made on the methods of habit fixation. In Figs. 6 and 7 are given samples of practise curves for several mice, showing the daily records in each group of trials, in the order that they were given. The arrows at the highest points on the curve indicate that the mouse did not pass through the maze or multiple choice test within 360 seconds.

In Fig. 6 are given the daily record curves for two agouti mice, 131 Ag. ♂ and 132 Ag. ♀. Their average time records are given in Table VI. They are mice from a single litter and their records show, in both cases, very rapid and consistent learning in the maze test, remarkably slight interference effect at the eighteenth trial, followed by a complete recovery and very speedy time records from the twenty-first to the twenty-ninth trial. No. 131, whose record is indicated by the solid line, made a record in all the tasks which was considerably better than the average. It showed a very strong retention for the maze test, while the records in the multiple choice test, although very good, show a more irregular performance than in the maze test. This irregularity of performance in the multiple choice test appears typical for the mice that have been tested in this investigation, and the irregularity may be due to the factors peculiar to the test itself, or, as previously discussed, to interference effects from previous training. No. 132, whose record is indicated in Fig. 6 by a broken line, failed to get through the multiple choice test for the first five days; succeeding on the sixth day, it made fairly good records thereafter. These two mice were taken as examples, because they typify the characteristic manner in which good time records were made by the mice that were tested in these experiments. The writer refers to the method of learning by rhythm of movements, which also has been noted by Watson, Basset and others for white rats. It was noted, for instance, that mouse 132 Ag ♀, in the initial learning period in the maze test, never went to the closed gate in the first compartment after the first two days of training. For a week of learning, from the fourth trial to the tenth inclusive, this mouse passed the first open gate successfully, but instead of going on to the open gate in the second compartment, the animal invariably made a detour

TABLE X

Rhythm in the Production of Errors in the Maze Test

Record for 131 Ag. ♂				Record for 132 Ag. ♀			
	No. of Days	No. of Times to Gates in Each Compartment: Closed	No. of Times to Gates in Each Compartment: Open		No. of Days	No. of Times to Gates in Each Compartment: Closed	No. of Times to Gates in Each Compartment: Open
Initial Learning Period	1	8	1 12	Initial Learning Period	1	1 1	3 2
	2	10	6		2	2 4	4 4
	3	1 2	2 2		3	4	1 3
	4	1	1 1		4	1	1 1
	5	1	1 1		5	1	1 1
	6	1	1 1		6	1	1 1
	7	1	1 1		7	1	1 1
	8	1	1 1		8	1	1 1
	9	1	1 1		9	1	1 1
	10		1 1		10	1	1 1
	11		1 1		11		1 1
	12		1 1		12	1	1 1
	13		1 1		13		1 1
	14		1 1		14	1	1 1
	15	1	1 1		15		1 1
	16		1 1		16		1 1
	17		1 1		17		1 1
Interference	18	1 3	1 1	Interference	18	4 1	1 1
	19		1 1		19	3 	1 1
	20	1 1	1 1		20	1	1 1
	21		1 1		21	1 	1 1
	22		1 1		22	1 	1 1
	23		1 1		23	1 	1 1
	24		1 1		24	1 	1 1
	25		1 1		25	1 	1 1
	26		1 1		26	1 	1 1
	27		1 1		27	1 	1 1
	28		1 1		28	1 	1 1
	29		1 1		29		1 1

	Record for 131 Ag. ♂			Record for 132 Ag. ♀		
	No. of Days	No. of Times to Gates in Each Compartment		No. of Days	No. of Times to Gates in Each Compartment	
		Closed	Open		Closed	Open
Retention	1	1	1	1	2	2
		1	1		1	1
	2	1	1	2	1	1
			1			1
	3	1	1	3	1	1
			1			1
	4	1	1	4	1	1
			1			1
	5		1	5	1	1
			1			1
	6		1	6	1	1
			1			1
	7		1	7	1	1
			1			1
	8	1	1	8	2	2
			1			1
	9		1	9	1	1
			1			1
	10		1	10	1	1
			1			1

by first going to the closed gate in the second compartment before completing the day's record. After the position of the open gate had been changed, from the right to the left in this case, at the interference point, *i.e.*, the eighteenth day, the mouse tried to pass through the previously open gates, but from the third day after the interference had been given a new rhythm was set up, which lasted until the last trial, where a perfect record of no errors was made. The rhythm in this case reversed the previous rhythm by the mouse trying to pass through the closed gate in the first compartment, but failing to pass this gate it quickly ran around to the successful gate and dashing through, continued on to the open gate in the second compartment. As will be shown in an accompanying table, it is interesting to note that this mouse, except for the first three days of training and the first two interference days, made only a single error on any one day. In Table X. are given the daily error records for the same two mice, Nos. 131 and 132, that were made in the maze test. The left half of the table gives the complete record for No. 131, while at the right is given the record of No. 132. The first column gives the number of days and indicates the extent of each learning period; the second column, the number of times that each mouse tried to pass through the closed gates in each compartment (if two numbers are given for any one day the upper number represents the errors in the first compartment and the lower one errors in the second compartment); while the third column gives the number of trials made at the open gates. The errors are thus given in the second column for each mouse, while the number 1 in the third column merely means that the animal has successfully passed through the open gates. The record indicated in the table by: 1 “1” 1 and found to occur on the fourth day for both mice, is repeated several times for each animal. This rhythm may be interpreted by saying that the animal successfully passed the first open gate, without making any errors in the first compartment, but ran around to the closed gate in the second compartment and was thus credited with an error before it finished the day's record. By a glance at the table it will be seen that the single error that was made in the second compartment was not eliminated until the tenth day for mouse No. 131, and one day later No. 132 made a perfect record. This same error occurred once more on the fifteenth day for No. 131, and of the twelfth and fourteenth day for No. 132. These examples are typical in that they show that errors in an animal's behavior are not suddenly discontinued, but are gradually eliminated. Further, it may be noted that after only two days of the interference test, the eighteenth and the twentieth, No. 131 changed its entire movement habit; after only six errors it was able to make the correct turn to the left in each compartment, while before the interference was set the correct turn had been to the right. This record shows a rather unusual amount of flexibility of behavior; only a few animals have been observed to make a similar record. In fact, if such records were found to be common in animal behavior we would have to alter our conception of a kinesthetic, or muscle sense movement; for the quick readjustment as exemplified in the behavior of mouse No. 131 would hardly be compatible with our idea of such a sense. The observations of the writer lead him to believe that kinesthesis was the all important element in adjusting the movements of the animals that were observed in this investigation. It is also to be remembered that in the case of No. 131 any visual, olfactory or gustatory clews, if they existed as such for the animal, were so left under the experimental conditions, that they would favor the production of errors during the interference period and not tend to their elimination. The behavior of mouse No. 132 in the interference period, with numerous errors, appears to be the more general type of animal reaction.

Turning again to the subject of the fixed rhythmic type of behavior it may be seen in Table X. that, in the case of No. 132, the record which is indicated by 1 “1” 1 in the initial learning period,

"1 1"
changed to 1 in the interference test, and this particular performance persisted nearly a month later in the retention period.

It has been noted from the careful study of each animal's behavior, that individual errors, or tendencies to make such errors, are, as a rule, gradually overcome, that a wrong movement made in the first few trials persists throughout several succeeding trials in a gradually diminishing extent. One mouse persisted in turning once to the left gate in the first compartment, when the gates were opened on the right side. It continued to do this for several days, when it was noted that it began to go only part way to the left gate; then turning around it took the successful path. This mouse never completely broke the habit, but in the end, the turn to the left had degenerated into a quick whirl around in a circle, followed by a dash through the proper gates. This type of behavior has been noted by other observers; the following is quoted from page 32 of Basset's monograph:[3] "As in the maze experiment, many of the inbred rats were subject to errors which persisted throughout the experiment. In particular may be mentioned one rat that invariably formed a loop in the course from the entrance to the point of operation."

Fig. 7 gives the daily record for mouse No. 134, Agouti ♀, whose average time records are given in Table VI. This record shows an initial set of seventeen trials in the maze test that almost duplicates the theoretical learning curve. The daily record for eight days, from the tenth trial to the eighteenth, which marked the application of the interference test, shows that the animal approximated the physiological limit of performance; taking an average time of one second per trial and making no errors at all during that period. The interference effect for the animal is slight, as shown by the first group of two trials, the remaining ten trials of the interference are a little irregular, but the noticeable individual difference in the behavior of this mouse is to be seen from the very poor record it made in the multiple choice test. It did not make a successful trip in that test until the sixth day and then failed for eleven consecutive days, from the eighth to the eighteenth day, inclusive, and also failed on the twenty-second day. During all these trials the animal appeared to be in very good health, it was active in the apparatus as well as in its nest box and made plenty of errors in testing the colored doors. It is tempting to speculate that the poor records of this mouse in the multiple choice test were due to an interference effect carried over from the previous training in the maze, but perhaps the relatively poor retention test that followed would seem to disprove this assumption. This case indicates that an experimenter can predict the actions of an animal to only a limited degree, for, after making consistantly good records in any one task, some external or internal factors, unknown to the investigator, may become operative and break up the expected sequence of learning. It also emphasizes the importance of testing an animal in more than one task before we finally grade its behavior.

If space permitted it would be interesting to give in detail the learning curves of mice that illustrate still other kinds of individual differences. It may be well to call attention to the average records of mouse No. 147 (see Table III.), that failed completely to make a successful trip through the maze, although it was tried in every one of the tests. It finally succeeded in making the poör average record of 159 seconds in the multiple choice test, but when retested at the time the retention test was given to the other mice, it again failed completely in the maze test. Because of the great amount of individual differences among the animals, it is difficult to find the record of a single mouse that made a record that was similar to the average for each task. Only No. 133, whose averages are given in Table VI., and No. 189, in Table III., approximated an "average" record. The record of No. 142, in Table III., is interesting in that it shows consistently uniform performance throughout all the tasks. This mouse made the following averages: 36 seconds in the initial learning period, 24 seconds for the first two interference trials, 36 seconds for the last ten interference trials, 35 seconds for the multiple choice test, and 33 seconds for the retention test. Another type of animal behavior is to be seen in the case of mice that made very poor records in the initial maze tests, and good records in the multiple choice and retention tests.[14]

The multiple choice test, which did not favor the production of a stereotyped form of reaction, nevertheless showed some interesting types of individual and group responses. It was found that the animals exhibited three types of behavior, which were not definite, but merged one into the other, so that an animal might use one of them on a certain day and another type on the next. The types of reactions were as follows:

Type 1.—Response by slowly and carefully "examining," visually or otherwise, each of the colored doors, very suddenly becoming oriented and dashing through the successful red gate.

Type 2.—The animal would react by trying the gates in a definite

[14] See records of the following mice in Table III., Nos. 138, 153, 154, 165, 171, 178, 181, 187, 188, 206, 207, 209, 210, 211, 212, and in Table IV., No. 169.

order, from right to left, or left to right, but going through the unlocked door as soon as it was reached.

Type 3.—Responding by trying the gates once each, in an irregular manner, and finally going through the right gate apparently by chance.

It was noted that the majority of the animals that were observed in this investigation used types 2 and 3 interchangeably. A few animals used type 1 for several days in succession and appeared to actually discriminate before they made their final choice of the gates. This type 1 reaction was noted by Burtt in his experiments entitled, "A Study of the Behavior of the White Rat by the Multiple Choice Method."[15] In the case of one rat he says: "In problem 1 she would at times become oriented very suddenly, dash to the right end and then across to the correct door." Types 2 and 3 of this investigation correspond to the *B* and *C* types found by Hamilton in the study of primates and rodents.[2]

[15] Harold C. Burrt, "A Study of the Behavior of the White Rat by the Multiple Choice Method," *Journal of Animal Behavior*, May, 1916, Vol. 6, No. 3.

X. HEALTH CONDITIONS AND LEARNING ABILITY

It is generally conceded that bodily changes in the general health of an animal may greatly tend to interfere with the formation of habits. This investigation, however, has shown that a number of ani-

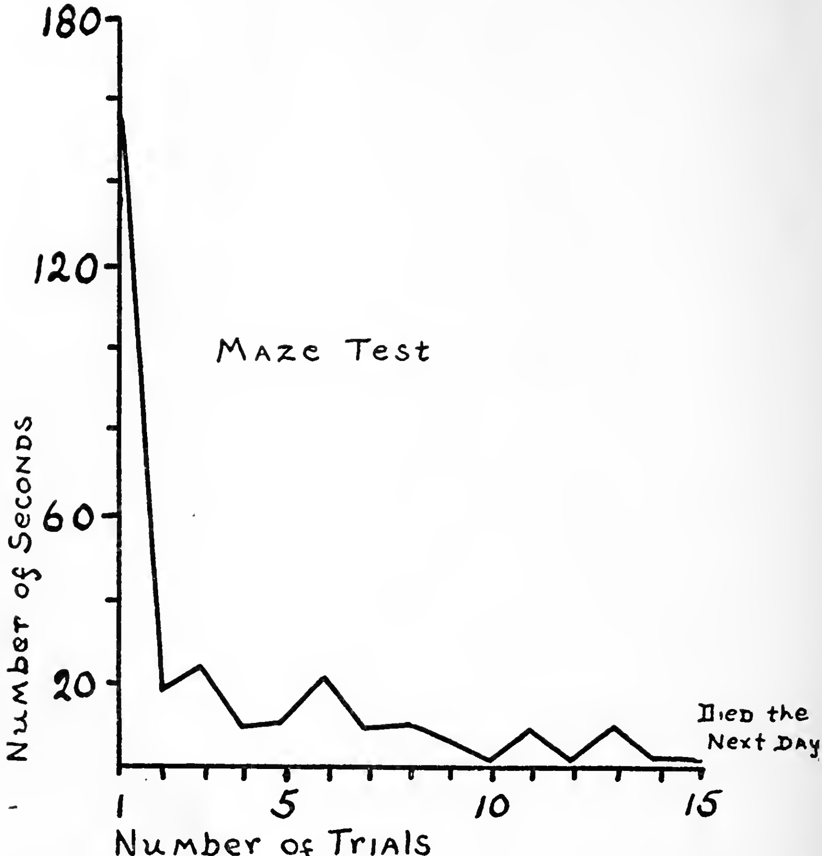


FIG. 9. Daily Record Curve for 156, White ♂, that gave good records although stunted and very ill throughout all the test.

mals have made exceedingly good records although they were apparently in very poor health. They were put through the tests, more out of curiosity to see what they would do than for anything else, and the experimenter was surprised to find them making very good records up to the very day they died. An example of such a case is shown in Fig. 9, which gives the daily record of No. 156, White ♂, who although stunted and very ill throughout the experiment, nevertheless made good records. This mouse did not seem to care for the reward, which was the food to be found at the end of a successful trip, and when it returned to its next box it suddenly became very inactive, and did not appear to arouse itself again until it was tested the following day. The tables also indicate the incomplete records of several mice that made good records up to the time of their death, and show that when a habit is once firmly fixed it may resist a good deal of distracting influences.

XI. FAMILY HISTORIES

We may now take up in detail the family histories. Fig. 10 gives a graphic representation of matings, from which were selected two mice, No. 20 Y ♂ and No. 26 Y ♀, that made the unusually poor records of 115 and 183 seconds respectively, though the other mice in the same litters had good records. The parentage of Nos. 20 and 26 was unknown; they were mated and gave two litters, each composed of three males and one female. Three mice in these two litters gave unusually slow records and made considerably more errors than normal. Two other mice gave poor records; two gave good records, while one died before it was tested. It is unfortunate that both females in these litters died before further offspring could be obtained. Table 1 gives the complete record of both time and error averages for these mice. It is a question whether or not selection of parents having poor records tended to produce more than the normal number of offspring slow to learn. Further investigation can alone afford an answer.

The mice whose records are given in Tables II. and III. are graphically represented in Figs. 11 to 14 inclusive. They have been carried down to the eighth generation and are still being tested. These animals are related and compose a group which is here called the white family. The matings began from an original pair of albino mice that were purchased from a dealer in the spring of 1913. The records for the parents and the first six generations are given in Fig. III., and it is to be noted that the time averages for the animals, whose catalogue numbers are lower than 128, are based upon performance only in the initial learning period in the maze test. The time records for all the animals of the F^7 and F^8 generations, and one family, Nos. 135 to 138 in the F^6 of Fig. 12, are based on an average performance in all the tests as previously explained.[16] The parents of the white family, Nos. 8 and 9, were quick to learn, making averages of thirteen and eighteen seconds respectively. The first generation was composed of four mice, one was not tested, two took considerably longer to learn the test than either of the parents, while the remaining animal made a quick record. The family average equalled 59.7 ± P.E. 15.7 seconds. A successful mating was ob-

[16] The average time for the 71 mice that was calculated upon performance in all tasks was 55.9 ± P.E. 3.5 sec., and this was practically identical with the average time made by the 183 mice that was based on the records for only the initial learning period of the maze test; namely, 54.1 ± P.E. 2.3 sec.

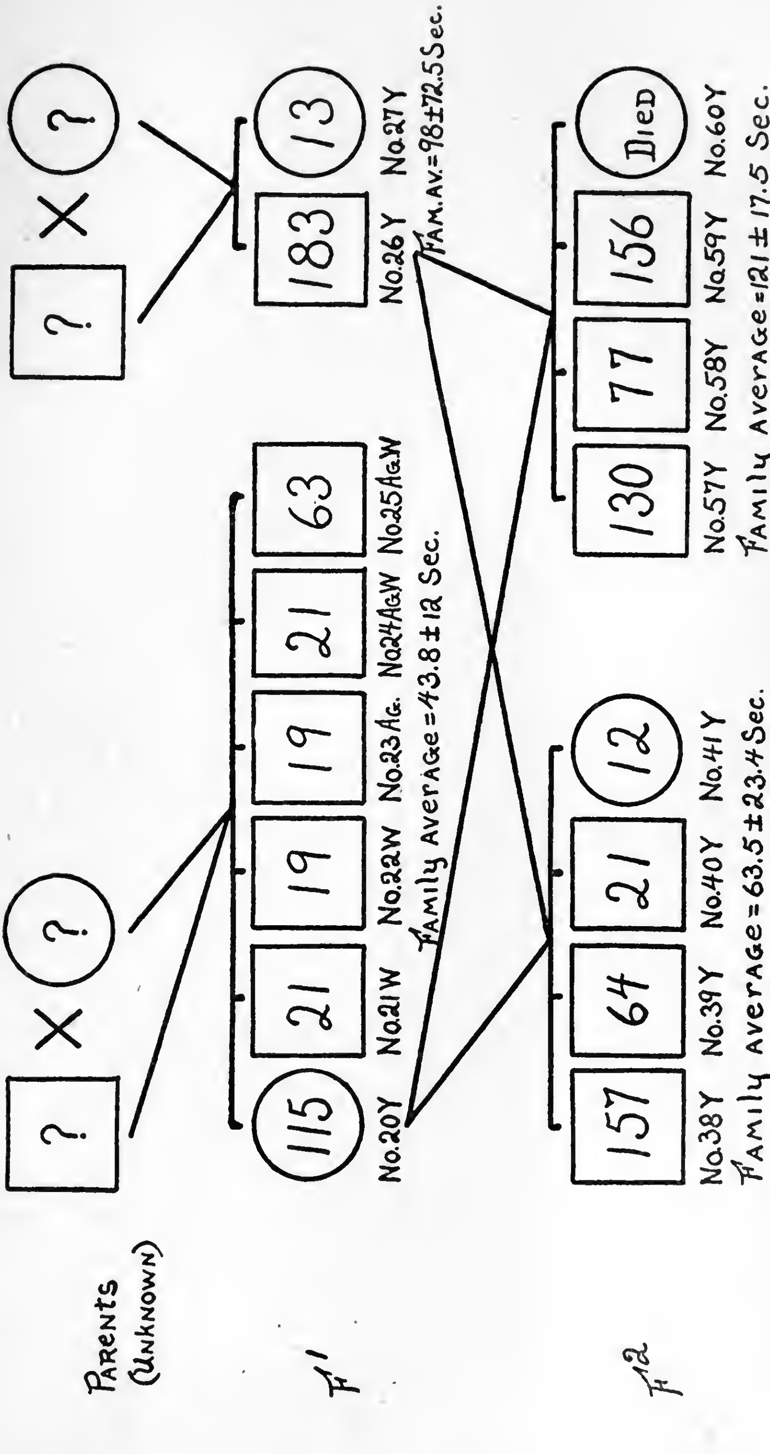


Fig. 10. Descent of a Yellow Family.

tained from the two mice that made the slow records, and the resulting litter of five females and one male composed the second generation of the white family. All the mice of this litter gave time records that were better than the average, so that the family time was very good, amounting to 23.3 ± P.E. 5.6 seconds. A glance at the chart will show that the majority of the mice of this second generation tended to resemble their grandparents rather than their parents. No. 48, the male of the litter, was quick to learn and made an average time of ten seconds. He mated with two of the females of the same litter, namely, No. 50, that made an average of 20 seconds, and No. 51, with an average of 9 seconds. From these two matings two litters were obtained that compose the third generation of the strain. All the mice in these litters made poorer records than either of their parents, and thus reversed the condition that occurred between first generation parents and second generation offspring, that we have just examined. The offspring of No. 48 × No. 50 were two females and one male. The male, No. 66, made the very poor average of 113 seconds, while one female took 35 seconds, and the other 75 seconds. Four offspring of this generation, three males and one female, resulted from the mating of No. 48 × No. 51, while here again one of the males, No. 78, made the poor record of 122 seconds, one other male, No. 77, took 68 seconds, and the female, No. 76, made an average of 31 seconds. The remaining male died before it could be tested. It will thus be seen that from the mating of No. 48 with two females, two litters were obtained that gave almost identical family averages; the average for Nos. 65, 66, and 67 was 74.3 ± P.E. 15.9 seconds, while the average for Nos. 76, 77, and 78 was 73.7 ± P.E. 19.5 seconds. The only litter that was successfully raised to continue the fourth generation of the strain was obtained by back crossing No. 66, who made the poor average of 113 seconds, with his own parent No. 50, with the quick average of 20 seconds. Their offspring composed a litter of two males and two females. One of the males died; the other made an average record of 53 seconds; one female made a good record of 24 seconds; while here again, as in the two previous litters, a single mouse, No. 72, made a poor average of 121 seconds. The family average for these three mice was 66 ± P.E. 22.3 seconds. No. 71 mated with No. 72, but only a single male, No. 91, was saved from their litter to stand as the sole representative of the fifth generation. No. 91 made an average time of 38 seconds, and was destined to become the paternal parent of a considerable number of offspring. As neither of the females of the fourth generation mated with No. 91, he was mated successfully with four unrelated white females,

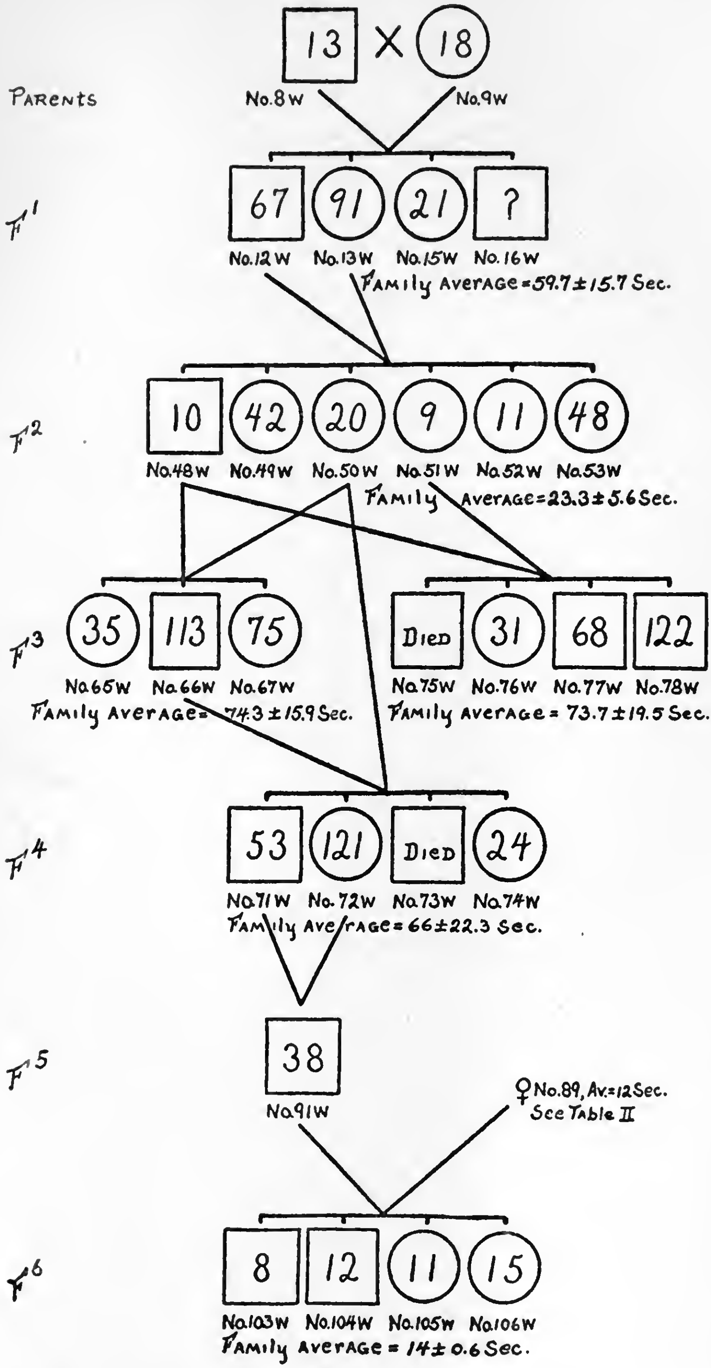


FIG. 11. Descent of a White Family.

Nos. 86, 87, 88, and 89. These females had been previously tested and found to give exceptionally good records as indicated in Table II. Twenty-seven offspring resulted from these matings. Their records are remarkably uniform and the family averages are among the lowest so far obtained. The result of crossing No. 91 with No. 89, who made the quick average of 12 seconds, is shown as the sixth generation in Fig. 14, which is composed of two males and two females. These mice were uniformly quick to learn, making averages of 8, 12, 11 and 15 seconds, with a family average of 14 ± P.E. 0.6 seconds.

Fig. 12 represents a continuation of the history of the white family resulting from the mating of No. 91 with No. 87, that made a good time record of 15 seconds. These mice produced, in the sixth generation, two litters, the first of which was composed of five males and two females that made exceptionally quick and uniform records, and showed, as in the previous mating that we have just discussed, a considerable amount of family resemblance. Their family average amounted to 13.1 seconds with the low P.E. of 1.7. It is to be noted that three of the mice of this litter made an average of eight seconds, two an average of twelve seconds, one an average of 16 seconds, and another 28 seconds. The second litter of No. 91 × No. 87 gave two males and two females, and again good records were made by all but one of the mice, No. 138, that made an average of 76 seconds. The other mice gave average time records of 23, 9 and 11 seconds respectively, while the family average amounted to 29.7 ± P.E. 11.4 seconds. The seventh generation of this particular branch of the white family was composed of three different matings of the sixth generation mice. No. 115, with an average of 12 seconds, was crossed with No. 118 that made a similar average. Their offspring are shown at the left of Fig. 15, and two litters were obtained from the same mice. The first litter gave a family average of 41.1 ± P.E. 7.5 seconds, and the members of this family made fairly low and uniform records, except in the case of No. 143, that made a poor time average of 100 seconds. Four of the other mice made records in the neighborhood of 30 seconds and one took 16 seconds. It is a curious fact that in each of the litters of the seventh generation of this branch of the white family there was a single female that made time records that were considerably poorer than any of the records made by the males of the same litters. It will be also found that this condition holds to a limited extent for the two other branches of the strain that will be given later in Figs. 13 and 14. This irregularity in the behavior of the females tended to increase the mean variations for the

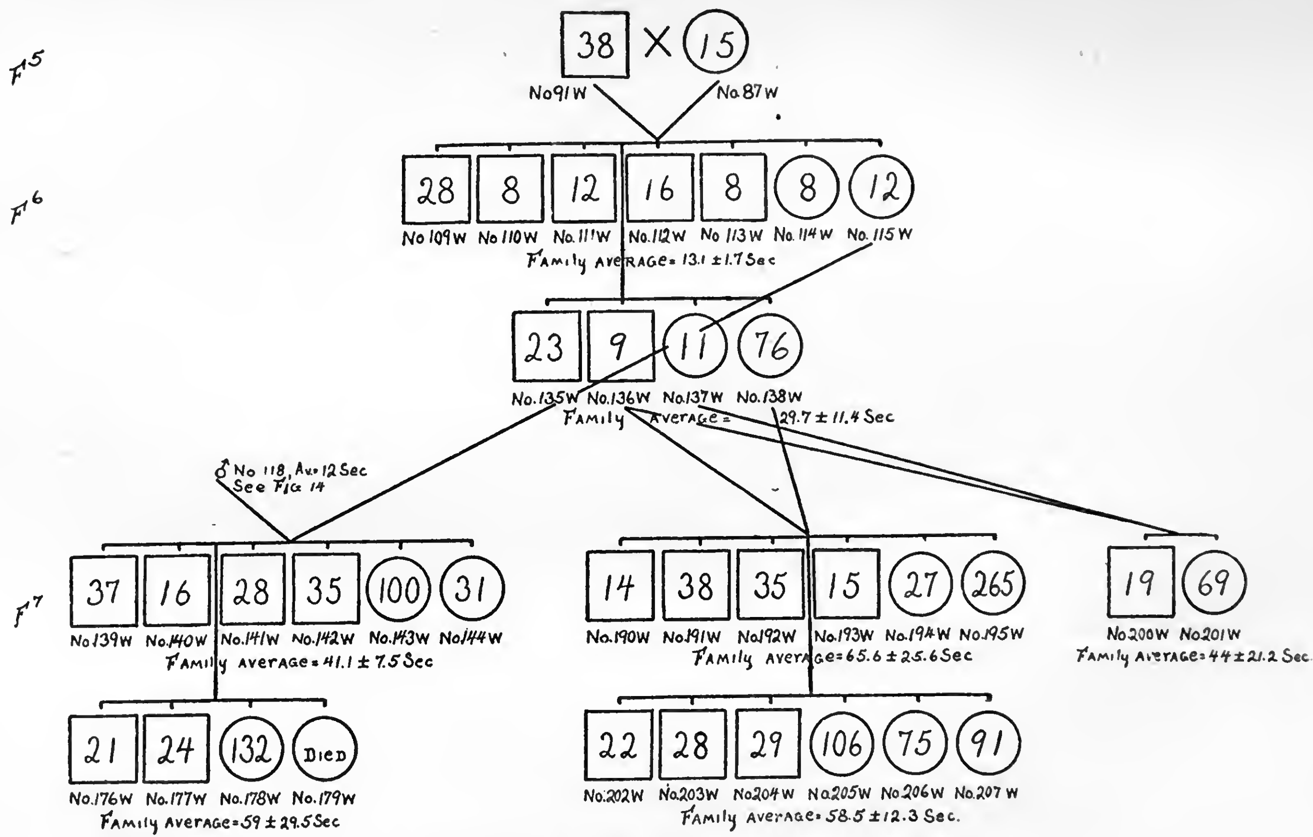


Fig. 12. Continuation of the White Family, No. 91 mated with No. 87.

litters and to decrease the family resemblances, while it accounts for the rather large sex difference that has already been discussed. The second litter, resulting from crossing No. 115 with No. 118, gave two males that made averages of 21 and 24 seconds respectively, one female that died, and again a very slow female that made an average of 132 seconds. The family average was 59 ± P.E. 29.5 seconds. The middle group of seventh generation mice that is given in Fig. 12, resulted from two successful matings of a male, No. 136, with an average record of 9 seconds, with No. 138 whose average was 76 seconds. It may be seen that in the first litter, Nos. 190 to 195 inclusive, all the males and one female made good records, while again, as was found in the previous litters of this generation, one female made exceptionally poor records. This was the case of No. 195 that made an average of 265 seconds. The family average amounted to 65 ± P.E. 25.6 seconds. The litter that was subsequently obtained from the same parents gave three males that made good records, 22, 28 and 29 seconds, respectively, and three females that made time averages considerably slower than the average for the entire population. One female made an average of 75 seconds, another 91 seconds, and still another 106 seconds. Again it may be noted that the females fell behind the males in quickness of learning. The family average of the second litter is similar to the first, and amounted to 58.5 ± P.E. 12.3 seconds. It is hoped that sufficient data may be later obtained to show whether or not there is a closer resemblance between litters belonging to the same parents than between unrelated litters, and also to determine whether or not mice belonging to the first litters of any two young parents tend to be superior or not to the mice that are produced from these same parents when they are considerably older and have produced several litters of mice. The number of double litters that has been obtained in these experiments is hardly large enough to warrant any conclusion on this point, but it may be noted that several of the cases that have been found show that the family averages of first and second litters overlap when due consideration is given to the size of the probable error that is attached to each average. No. 136 and No. 137, that made very good time records, were mated, and they added two more mice to the seventh generation of the white family; one of them, a male, made an average of 19 seconds and the other, a female, took 59 seconds. Their combined average amounted to 44 ± P.E. 21.2 seconds.

In order to study another branch of the white family we must again go back to the fifth generation and note what resulted from mating No. 91 with another white female, namely No. 86, that also made a quick time record of 15 seconds. The records for this branch of the strain are given in Fig. 13. As in the other branch we have just discussed, No. 91 again produced an exceptionally quick and uniform family in the sixth generation. This litter was composed of three males and three females, Nos. 122 to 127 inclusive. Four in the litter made time averages of fifteen seconds or lower, while the other two took 32 and 33 seconds each. Their family average amounted to 19.1 ± P.E. 3.4 seconds. No. 125, with an average of 9

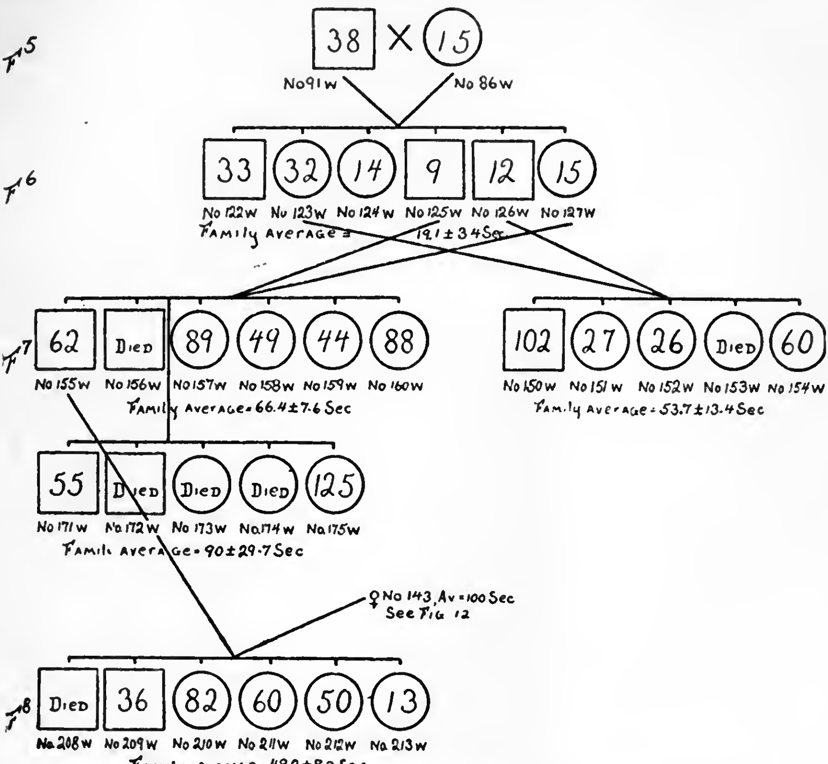


Fig. 13. Continuation of the White Family, No. 91 mated with No. 86.

seconds, and No. 127, whose average was 15 seconds, were mated and they produced in the seventh generation two litters. The first was composed of six mice, Nos. 155 to 160 inclusive; two were males and four females. All these mice made averages slower than any of their sixth generation kin. One male made an average of 62 seconds, two females gave average records of about 90 seconds, one of 44 seconds and another of 49 seconds. The family average here was fairly high, amounting to 66.4 ± P.E. 7.6 seconds. The second litter was composed of two males and three females, but Nos. 172 and 173 died before they could be tested, while No. 174 did not complete its record. The incompleted records of this mouse are given in Table III. The one male that lived, No. 171, gave an average of 55 seconds and the female made an average of 125 seconds. Their combined average equalled 90 ± P.E. 29.7 seconds. Once again it may be noted that the female was the one that made the comparatively poor time record. The other litter of the seventh generation was obtained from mating No. 123, whose average was 32 seconds, with No. 126, whose average was 12 seconds. Their offspring, Nos. 150 to 154 inclusive, compose a litter of one male and four females, one mouse, No. 153, did not quite complete its record (see Table III.). The male of this litter, No. 150, made a poor average record of 102 seconds, and it is worth calling attention to the fact that its case is the only one, among all the sixth, seventh, or eighth generations, where a male has made a very poor record. Two females, Nos. 151 and 152, made good records of 26 and 27 seconds each, while the remaining female made an average of 60 seconds. The family average in this case was 53.7 ± P.E. 13.4 seconds. The matings for the eighth generation have only just been started; one litter, however, Nos. 208, to 213 inclusive, has been obtained by crossing No. 155, a male, with a time average of 62 seconds, with No. 143, whose average was 100 seconds. Their offspring were six in number, two males and four females. One male, No. 208, died, but its incomplete records are given in Table III., and the remaining male, No. 209, made an average record of 36 seconds. One female, No. 213, gave a good average of 13 seconds, while the other three females ranged from 50 to 82 seconds. The average family record amounted to 48.2 ± P.E. 8.0 seconds.

The fourth and last branch of the white family is given in Fig. 14. Here it may be seen that No. 91 was mated with another female, No. 88, that made the very quick average record of 9 seconds. The sixth generation litter of four males and two females that was obtained from this mating made the best family record that has so far been obtained, namely, 12 ± P.E. 1.0 second. None of the mice made an average time record exceeding 20 seconds. It is an interesting point that all but one of the 27 offspring of No. 91 made time records considerably superior to the average. Three matings were made from these sixth generation mice. No. 116, a male, was crossed with No. 121, and these two mice made average records of 10 and 12 seconds respectively. Two seventh generation litters were obtained from them. The first litter, Nos. 161 to 166 inclusive, was composed of three males and three females, with a family average of 34.3 ± P.E. 7.4 seconds. Two of the males made good records of 14

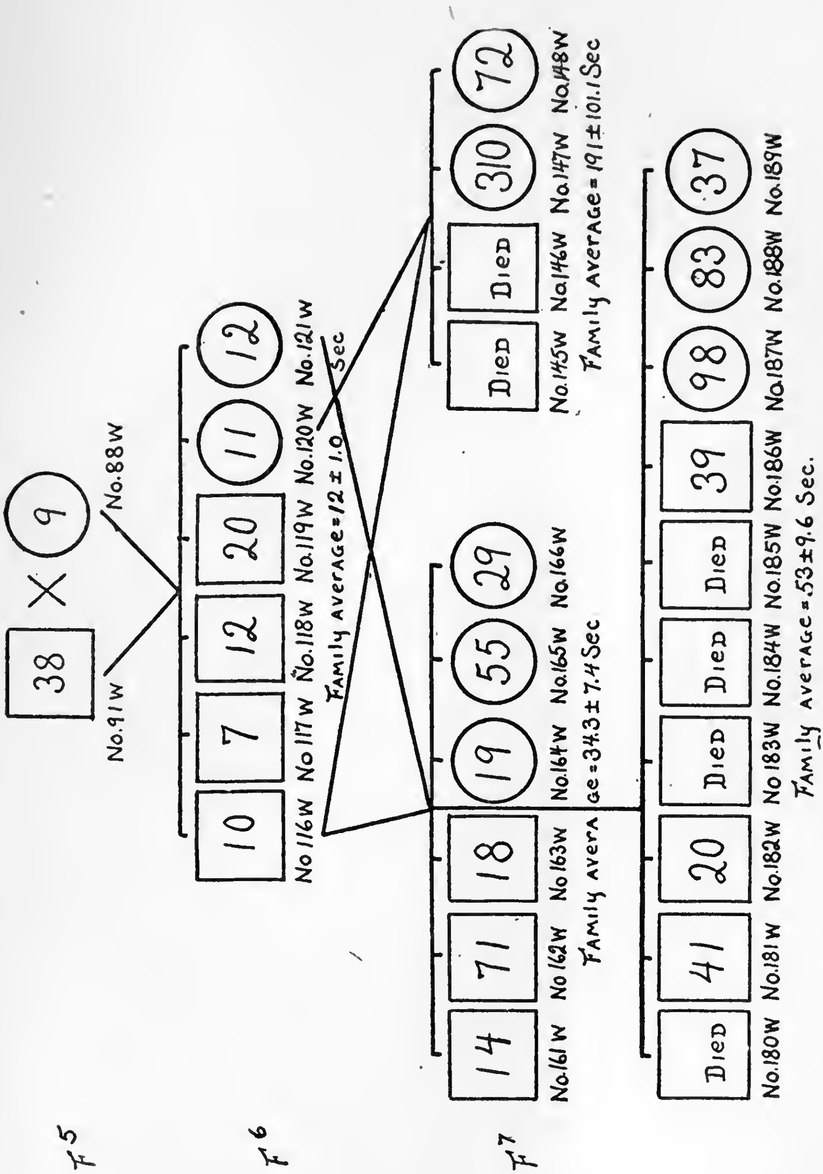


Fig. 14. Continuation of the White Family, No. 91 mated with No. 88

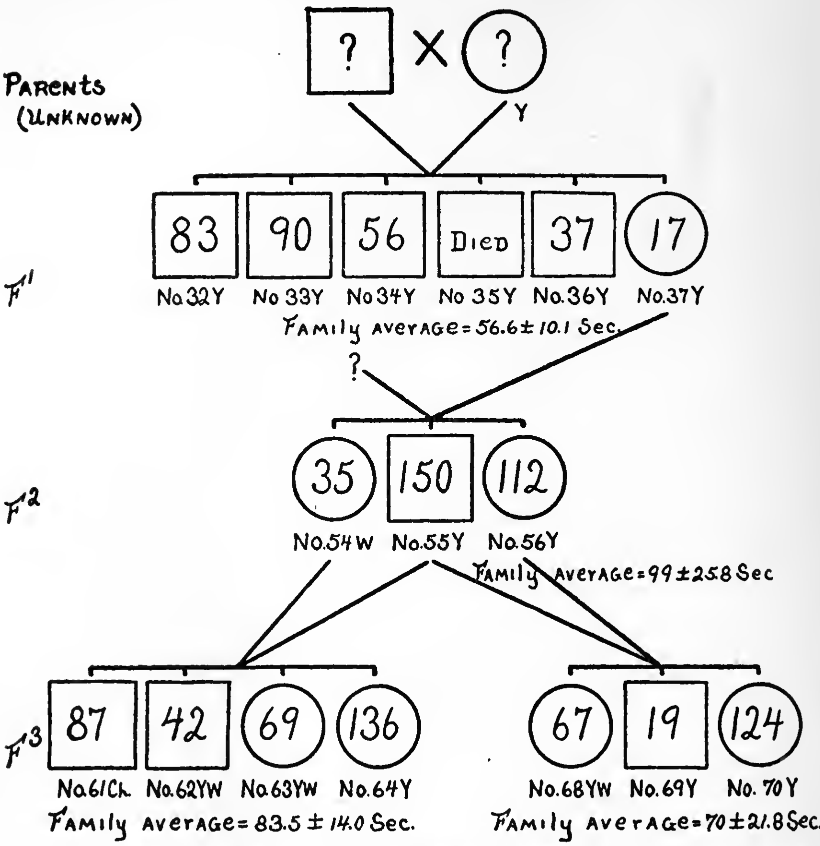


FIG 15. Descent of a Family of Mice consisting mostly of Yellow Individuals.

and 18 seconds, while one gave a record of 71 seconds; two females also made good records, and one an average record. None of the mice in this litter made records as low as either of their parents. The second litter of No. 116 × No. 121 was rather large, composed of ten mice, but three of the seven males died before their records were completed (see Table III. for records of Nos. 180, 183 and 185), and one male, No. 184, died before the tests began. Three of the males and one female that completed their records made fairly quick averages, while two females gave time averages of 98 and 83 seconds. The family average amounted to 53 ± P.E. 9.6 seconds. When No. 116 was mated with No. 120, whose average was 11 seconds, a sixth generation litter of four mice resulted; two were males and died before their records were completed (see Table III. for records of 145 and 146), and one female, No. 147, gave the very poor record of 210 seconds. This was the female that failed completely in all the maze tests, while the other female, No. 148, made a poor record of 72 seconds. It is to be noted again that in the seventh generation of this branch of the white family, four of the females made time records inferior to any of those made by the males of the same generation.

Another family, mostly of yellow mice, was derived from a yellow female and an unknown male, probably white. The first generation from this mating gave a litter of six, Nos. 32 to 37 inclusive. The records of five of these, one having died, are given in Table IV. and are graphically represented in Fig. 15. This litter gave a family average of 56.6 ± P.E. 10.1 seconds. Two of the males, Nos. 32 and 33, made poor records; No. 34, an average record of 56 seconds; while the remaining male made a fairly good record of 37 seconds. No. 37, the only female of this litter, gave a record of 17 seconds; she mated but once, and it is not known with which brother. She bore in the second generation two females and a male, Nos. 54, 55 and 56. One female, No. 54, made a record of 35 seconds, while the other female made a poor record of 112 seconds. The male, No. 55, also did poorly with an average of 150 seconds. No. 54 and No. 56

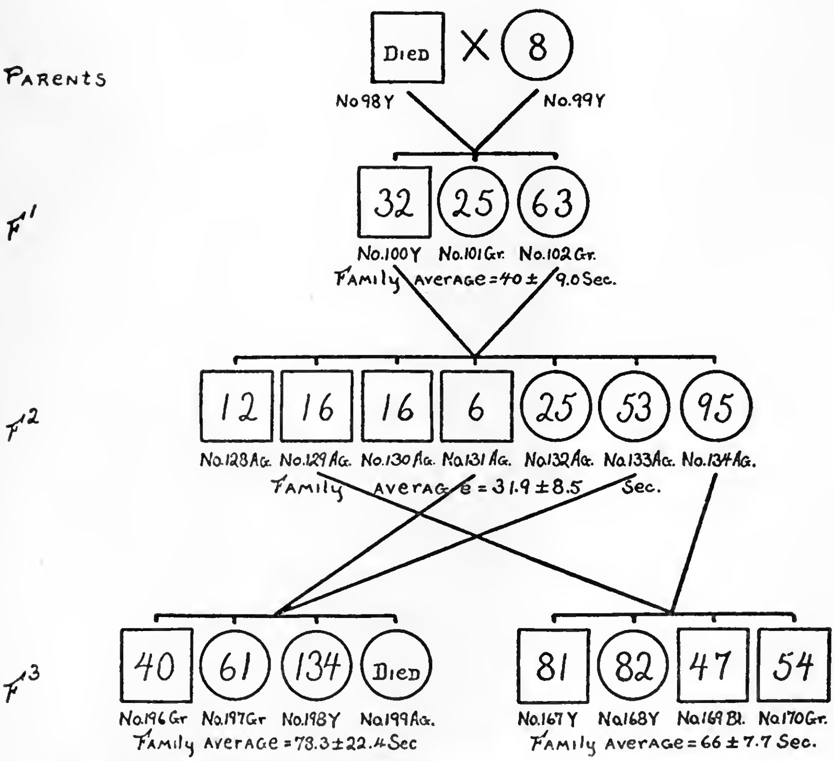


FIG. 16. Descent of a Family of Mice composed of Colored Individuals.

were both crossed with No. 55, and two litters resulted. No. 55 × No. 56 gave Nos. 68, 69 and 70 in the third generation, and No. 55 × No. 54 gave Nos. 61 to 64 inclusive. From a survey of the complete records of these mice, it is seen that although the second and third generations came from the female, No. 37, which made the exceptionally low record of 17, still, two of her young in the second generation made poor records, and Nos. 61, 65 and 70 in the following generation did the same.

The last family history that will here be discussed is composed of another strain of colored mice that began from a mating of two yellow mice. The records for these individuals are given in Table VI. and are graphically represented in Fig. 16. This family was started in order to see whether there was any real correlation between strains of colored mice and their ability to learn quickly and to adapt themselves to varied conditions. It was especially desired to test the learning ability of a greater number of yellow mice, and to this purpose the two founders of the family, Nos. 98 and 99, were both yellow mice that came from the same yellow strain that had previously exhibited individuals that were very slow to learn. It was also desired that a further study of the effects of close inbreeding be carried on independently of the other problem, and for this purpose certain strains of the family are now being bred with this object in view. The inbreeding in this investigation has not been carried on close enough, or for a sufficient number of generations, to give any definite results. It is to be noted that in the white family it was necessary to outbreed in the fifth generation. To return to the records of the family of colored mice, No. 99, of the parent generation, made a very good record of eight seconds, while the male died before he was tested. The first generation from this mating gave one yellow male and two gray females. The male and one of the females made good records, while the other female gave an average time of 63 seconds. The family average was 40 ± P.E. 9.0 seconds. The male of this generation and the normal female, No. 102, were mated and a litter of four males and three females resulted that composed the second generation. All these mice were of an agouti color in which a good deal of yellow pigment was added. All the males and one female, No. 132, made good records; a female, No. 133, made an average of 53 seconds, and No. 134 an average of 95 seconds. Here the family average was 31.9 ± P.E. 8.5 seconds. Two matings were obtained from the mice of this litter, and the resulting offspring compose the third generation; No. 131, with a very good average of 6 seconds, was mated with No. 133, with an average of 53, and gave

Nos. 196 to 199 inclusive; while No. 129, average 16, $\times$ No. 134, average 95, gave mice Nos. 167 to 170 inclusive. The litter of No. 131 $\times$ No. 133 was composed of a gray male, with an average of 40 seconds; a gray female, with an average of 61 seconds; a yellow female that made the very poor record of 134 seconds, and one agouti female, No. 199, that died before its records were completed. This agouti female, however, made very good records in the initial learning, and both interference periods of the maze test before it died. Its records may be considered similar to those of No. 156, already discussed (see Fig. 12). The family average for the first litter in the third generation was 78 $\pm$ P.E. 22.4 seconds. The second litter in this generation was composed of Nos. 167 to 170 inclusive; offspring of No. 129 $\times$ No. 134. The family consisted of three males and one female; there was a yellow male and a yellow female, a gray male, and finally a black mouse also a male. The two yellow individuals, Nos. 167 and 168, made the poor averages of 81 and 82 seconds respectively; the black colored male took 47 seconds, while the gray male 54 seconds. Their family average was 66 $\pm$ P.E. 7.7 seconds. This experiment has not extended far enough to give as yet any definite results, but it is interesting to note, that, as was found in the case of the white mice, the males continue to make, on the whole, better average records than the females, and the three yellow mice, Nos. 198, 167 and 168, in the third generation make the poorest time records in their respective litters. If there is a tendency for the yellow mice to be slow to learn, this point can only be finally determined as the work of this investigation continues.

XII. SUMMARY

1. Albino and colored mice can be used to advantage for laboratory work in animal behavior.

2. Both the maze and the multiple choice apparatus that were used in this investigation appear well adapted for this kind of work.

3. There is a marked difference in individual behavior, with a close correlation between time and error records.

4. There appears to be a fairly close relation, in the maze test of this investigation, between performance at the beginning of the initial learning period with performance at the end of the same period. The quick learners appear to make the best records at the end of the test. This relation appears to hold for the results in the multiple choice test, but in this case the coefficient of correlation is much lower.

5. There appears to be practically no correlation between performance in the maze test with performance in the multiple choice test, and this may be due to factors peculiar to each test, as has already been discussed in the body of this thesis.

6. On the whole, the animals that do well in the initial learning period of the maze test continue to make good time averages in both of the interference tests and in the retention test.

7. The quick learners appear to exhibit a considerable amount of flexibility of behavior.

8. Definite types of movements were noted in the formation of habits in both the position test and the discrimination test.

9. There appears to be a certain amount of interference effect carried over from one type of reaction to another.

10. Poor health in an animal may not necessarily be correlated with slow learning ability, for when a habit is once fixed it may resist a considerable amount of disturbing influence.

11. There appears to be no particular resemblance among individuals of the same litter, as judged by their time records for various tests.

12. There appears to be a considerable difference among different strains.

13. There appears to be a sex difference in favor of the males in all the tests of these experiments.

14. The females appear more variable than the males in their behavior.

ACKNOWLEDGMENTS

The writer wishes to thank Professors R. S. Woodworth, E. L. Thorndike and H. L. Hollingworth, of Columbia University, for their helpful suggestions and assistance throughout the experimental study. Thanks are also gratefully rendered to Professors James McKeen Cattell, Margaret Floy Washburn and A. T. Poffenberger. It was due to the kindness of the late Dr. August Hoch that the author secured the insane patients as subjects.

CONTENTS

Yerl.
A.

EXPERIMENTAL STUDIES IN RECALL AND RECOGNITION

BY

EDITH MULHALL ACHILLES, PH.D.

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. WOODWORTH

No. 44, SEPTEMBER, 1920

COLUMBIA UNIVERSITY CONTRIBUTIONS TO PHILOSOPHY AND PSYCHOLOGY
VOLUME XXVII, No. 1

NEW YORK
1920

AGENTS: G. E. STECHERT & CO.; London (2 Star Yard, Carey St., W. C.);
Paris (16, rue de Condé).

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

EXPERIMENTAL STUDIES IN RECALL AND RECOGNITION

CHAPTER I

Historical Survey

Memory has been mentioned in literature for centuries, as Burnham[1] (5) has shown in his survey of the history of the subject. The first scientific study of the subject was in the time of Aristotle; his doctrines and conceptions as well as those of Plato, St. Augustine, Leibnitz, Bacon, Locke, Hume, Hartley, Kant and others, are interesting, but not until 1885 do we find a record of any experimental work. During the years 1879–80 and 1883–84 Ebbinghaus (7) performed upon himself a long series of experiments in memorizing nonsense syllables. Experimental researches on memory have continuously increased since Ebbinghaus published his results. Here will be considered those which have been concerned with the differences between recall or reproduction and recognition.

As methods of testing memory, reproduction and recognition were distinguished by Wolfe (17) in 1886. In studying the memory of tones he found it easier to use the method of recognition and rejected that of reproduction which Ebbinghaus had employed. Baldwin and Shaw (1), in determining the accuracy of the memory for the size of a square, used three different methods—(1) Selection from variety, (2) Identification, (3) Reproduction. Memory curves were made for these methods and found to be practically parallel, but the degree of accuracy much higher by the second method than by the first. In the method of reproduction the subjects after having seen the square were requested to draw it. The reproductions were always too small. Binet has published notes on the experimental study of memory to which Warren has referred. We quote in part: (1) Method of Reproduction—this consists in making the subject *reproduce* his impression; if it is a story that he has listened to, he repeats or writes it; if it is a visible form, he sketches it; if it is a sound or an inflection, or some visible movement, he imitates it; if it is a color, he reproduces it by mixing the tints on a palette or varying the sectors of rotating discs. This is certainly the most natural method, in that

[1] The numbers in parentheses refer to the list of References on p. 76.

it follows closely our ordinary procedure; but as a method of studying the memory it is somewhat indirect; it requires a translation or transposition of the impression, and a special aptitude, *e.g.*, in sketching or painting; it should, therefore, be expressly reserved for the verbal memory. (2) Method of Selection—this consists in the recognition of an impression when it is presented again in company with others; thus a tint is exhibited at first and the subject must remember it; at the end of a certain time he is shown a graded series of the same color, comprising 15 different tints, and must recognize the one which was shown him before; the second presentation may give the whole series simultaneously or successively. This method is simpler than the preceding, since it bears more directly on the memory; there remain to be considered the sources of error which it contains. Whenever we have to make a choice from among a set of objects our attention is drawn toward the center of the series; if the impression to be recognized is the sixth in a series of 15, the seventh impression is more apt to be indicated than the fifth, because the former is nearer the center; consequently the arrangement of the series exerts some influence upon the character of the results. (3) Method of Comparison—the subject compares the remembered impression with another impression which is shown him, and answers that the latter is "equal, greater, or smaller." Kennedy (12) explains that in every experiment on memory there are of necessity two terms—*A*, a stimulus of some sort which constitutes the thing to be remembered, and *B*, an expression on the part of the reagent of the state of the memory of *A*. "The methods of experimental investigation of memory are to be distinguished from each other solely by the character of the *B* of the experiment. . . . The distinction between memory methods rests upon a difference between the general phenomena of memory, namely, those of reproduction and recognition. It is a well-known fact that former experiences are much more easily recognized than reproduced and that we understand many more words than we are able to use. As far as experimental methods are concerned, in reproduction we call up an old experience, unaided directly by any immediate expression from the outer world; while in recognition we simply feel the identity of a term already in consciousness with one which was there in the past. In the experiment itself this difference takes the following form. In the method of reproduction the subject, having been given some stimulus, is called upon after the lapse of a certain time to reproduce that stimulus without any further aid from the experimenter. . . . On the other hand, in the method of recognition, the subject, having been given the stimulus to be remem-

. . . .

"9. The effective element is very much more pronounced in recognition than in recall.

"10. The wide difference in results obtained from the different groups under different conditions shows how unreliable it is to derive general laws from small groups of subjects arbitrarily tested. Therefore the conclusions herein derived are necessarily limited to the tests described in this study, until they shall have been verified by further studies.

"11. Some sex differences are obvious:

"(*a*) The girls are superior to the boys for both recognition and recall, and much more for recall than for recognition. Their superiority for recall invariably holds true even when divided into grades, while for recognition this superiority obtains only for large groups.

"(*b*) The most striking sex difference is shown in the ratio of recognition to recall; while on the average, the boys' efficiency for recognition is three times that of recall, the ratio by the girls is about two. Likewise the total gain in per cent., in recognition efficiency over that of recall is markedly superior for the boys.

"(*c*) A higher correlation between recall and recognition obtains for girls than for boys.

"(*d*) For recall the girls seem to be more variable than boys, while for recognition the opposite holds true.

"(*e*) The girls recall more incorrect words than the boys, while the boys recognize more incorrect words than the girls.

"(*f*) Both for recall and recognition a higher percentage of the total answers given by the girls than those given by the boys are correct."

Hollingworth (11) says "the difference between recall and recognition seems to be a rather simple matter. Recall is that aspect of memory process in which a setting, a background or association-cluster, is present in clear consciousness, but a desired *focal element* is missing. . . . Recognition is, schematically, just the reverse of this process. In recognition the focal element is present, in the form of sensation, image, or feeling, and the question is whether or not this element will recall a more or less definite general sitting or background. . . . It is often said that recognition is an important part of recall memory, giving warrant to the correctness of the recalled element. This may often be the case, but it is not necessarily so. Items may be correctly recalled but not recognized as correct and rejected. Moreover, every case of recognition presumably involves recall or tendency to recall on the part of the setting. The ordinary act of 'memory' is said to be complete when focal element and setting belong together, that is, mutually recall or sustain each other.'' He calls attention to several factors which appear to influence recall and recognition memory in different degrees, but draws only tentative conclusions due to the meager experimental data.

''An Experimental Investigation of Recognition and Reproduction'' was published by Rybnikoff in 1914 in Russia. He says that the method of reproduction has been used to a great extent, probably due to the influence of Ebbinghaus, and that the method of recognition has been slightly developed. Although the title of the paper suggests an investigation in a comparison of the two processes, he uses, for the most part, recognition in his series of memory tests. Tests for reproduction were made in some cases, after the recognition test, by the *Treffermethode.* It is true that in the earlier researches on memory the method of reproduction was used. Recently, however, the investigations have employed recognition. Rybnikoff is to be praised for realizing the one-sided development, rather than adversely criticized for not being familiar with the recent work in recognition, as the recent researches in recognition had probably not reached the University of Moscow when Rybnikoff wrote.[2] Among the many researches on memory few have been concerned with recall and recognition as different methods of testing, notwithstanding their differences were pointed out early in the history of the experimental literature.

[2] The author wishes to express her thanks to the late Professor J. W. Baird, of Clark University, for his notes in English on the work of Rybnikoff.

CHAPTER II

The Problem Stated

Different problems in the field of the experimental work on memory have been examined. Often the data of various investigators on similar problems have been apparently incomparable due to the difference in methods of testing memory. For some time it has been known that one can recognize more items than one can recall, but it has not been known whether the recognition memory can be stated as two, three or four times the recall memory, or whether the position at the beginning or end of a list, or other factors, will influence recognition as much as they do recall. There are no data to prove that one who recalls well can also recognize well; nor whether the results which exist for one material, *i.e.*, words, would be equally evident for another material, *i.e.*, nonsense syllables or geometrical forms.

The present investigation has been primarily designed to examine the similarities and differences in the two methods of testing, reproduction and recognition. The questions which it has aimed to answer are:

1. What is the equivalence of repetitions in Recall and Recognition of different materials?

2. Does the factor of determination to remember influence recall and recognition in the same way?

3. Does primacy effect recall?—recognition?
   " recency " " "
   " color " " "
   " size " " "

4. Does the person who recalls one material well also recognize that material well? Is the answer the same for different materials? Is it true for both adults and children?

5. Does the person who recalls one material well also recall another material well? Is the answer the same for both adults and children?

6. Does the person who recognizes one material well also recognize another material well? Is the answer the same for adults and for children?

7. Is a material which is recalled well also recognized well?

8. Are there any sex differences in achievement among the adults in recall?—in recognition?

9. Are there any sex differences in variability among the adults or among the children in the recall of the different materials?—in the recognition of the different materials?

10. Is there any difference in recall among the children of the same age regardless of class in school?—in recognition?

11. Is there any difference in recall among the children of the same grade regardless of age?—in recognition?

12. Is there any difference in the amount recalled or recognized by children, college students, and graduate students?

13. Do tests for recall and recognition on insane subjects, who have memory defects, reveal any characteristic differences between the methods?

CHAPTER III

Preliminary Experiments

Further experimental investigation into the relation of the two memory processes, recall and recognition, was stimulated by Hollingsworth's article on "Characteristic Differences between Recall and Recognition."[1] His title suggests that there are two differences, but his conclusions are only tentatively drawn, due to the meager data. To add to the data on points which he raised was the purpose of these preliminary experiments.[2] The first questions asked were:

1. What is the equivalence of repetitions for recall and recognition?

2. What is the influence of the factor of determination to remember on recall and recognition?

3. What is the effect of the factors of primacy and recency?

1. *Equivalence of Repetitions*

"It is, of course, a matter of common experience that a single presentation may suffice to enable recognition but be quite insufficient to make recall possible," says Hollingsworth. The experiments of the present writer were designed to determine the difference in the number of repetitions necessary for the mastery of fifteen items for four materials, pictures of objects (lamp, fish, clock, etc.), geometrical forms, words (chair, hammer, car, doll, etc.) and nonsense syllables (fik, vod, deb, ruz, biv, etc.).

"Mastery" was measured for both recall and recognition. Each subject was shown fifteen words successively at regular intervals of two seconds each and then required to write those he remembered; in the case of the pictures of objects, the name of the object was written. A set of thirty (30) containing the original fifteen (15) was given to the subject from which he was to select fifteen which he thought were in the original set. The first set of fifteen was presented again as before and the subject was requested to recall those

[1] Hollingworth, H. L. "Characteristic Differences between Recall and Recognition," *Amer.* J. of Psych., 1913.

[2] Mulhall, Edith F. "Experimental Studies in Recall and Recognition," *Amer. J. of Psych.*, 1915.

he could and then to select fifteen from the thirty set. This was repeated until he was able to recall and recognize the fifteen items correctly. The same experiment was performed for all four materials on twenty-five subjects, all undergraduate or graduate students in Columbia University.

Table I shows the average number of repetitions necessary for complete recall and complete recognition, and the ratio of the two for all materials.

TABLE I

Material	Process	Av. No. Repetitions	M. V.	Ratio Av. for Recall to Av. for Recog
Pictures	Recall	3.36	.79	3.23
	Recog	1.04	.08	
Forms	Recall	3.96	.85	2.20
	Recog	1.80	.81	
Words	Recall	4.76	1.61	1.80
	Recog	2.64	.96	
Syllables	Recall	7.12	2.23	1.22
	Recog	5.80	2.11	

The data indicate that the difference between recall and recognition memory is greatest for pictures, somewhat less for forms and words, and least for nonsense syllables. The number of repetitions required for both recall and recognition is least for pictures, greater for forms and words, and greatest for syllables. In examining the materials one finds that the pictures offer the greatest richness of associations. The forms offer somewhat fewer advantages—the shape can be visualized, in a few cases they can be named, and associations, but few in number, can be made. The third material, words, all being nouns, are names and have associations, but lack any particular picture or form element. The syllables, as their name implies, are nonsense; most, if not all, are devoid of any association. Introspections of the subjects indicate that association was employed as an aid in memorizing.

The following secondary experiment shows this more definitely. To each of two subjects a mixed set of sixteen items (four pictures, four forms, four words, four syllables) was presented, one unit every two seconds. Careful introspections were called for as each item was afterwards recalled. The subjects were also required to select sixteen from a set of thirty-two (eight of each material). A few of the records are given as examples of the means the subjects used to aid memory.

Pictures Remembered	*Why and How*
butterfly	Benefit performance of Madam Butterfly to be given for the College.
car	Its many uses, with engine, auto, etc.
coat	Livery.
bridge	A particular bridge in Sullivan County.
lamp	Toaster received yesterday which I connected to a lamp.
nuts	Walnuts I bought to-day.
Words	
tree	Curious tree examined on a recent walk.
knife	Bread.
Syllables	
naf	Nap.
jod	Jodl, psychologist.

Forms: The forms reminded the observers of "a diamond," "a wind-mill," "exclamation point," "color disc," "pie," and "kindergarten shapes."

The introspections of both subjects show clearly the method of association in recalling the material and the difficulty and often inability of remembering material with few or no associations.

Since the greatest difference in recall and recognition memory is for pictures when the material has an abundance of association, less for forms and words, and least for syllables where associations are respectively less, it appears that the difference in recall and recognition memory is in part dependent on the richness of associations present.

2. *Influence of Determination to Remember*

Hollingworth's data on this point were gathered in an experiment in which each of five observers performed the "opposites" test 60 to 75 times. The task was to speak the opposite of each of fifty adjectives as quickly as possible. One or two trials were made each day but the order of the list was changed. After 60 to 75 trials each observer was asked to reproduce all the pairs of opposites that had been used. Recognition was tested by presenting one hundred pairs of opposites, containing fifty new pairs and the fifty original ones, and asking the observer to identify the original fifty. The results were:

	Rc. in 3 Min.	Rc. in 15 Min.	Rg. in 3 Min. or Less
Average of five observers	14.8	28.8	49.6

from which Hollingworth concludes: "It is not possible, on the basis of these figures, to say that recognition is not influenced by the presence or absence of the determination to remember, but it is clear that it is much less influenced by this factor than is the process of recall." No control experiments in which determination was employed were made and no definite conclusions about the influence of this factor can be drawn. The recognition score is about perfect, but the presence of this 100 per cent. recognition efficiency when no determination to remember was present, does not indicate that determination could have exerted no influence, if it had been present, but that it would have had no opportunity to show its influence or lack of influence in the scoring, since 100 per cent. is maximum efficiency.

To gather further data on the influence, if any, of intention to remember on recall and recognition two groups of twenty-five subjects each were tested. Twenty (20) photographs (15 men, 5 women) with names attached, and fifteen syllables were the two kinds of material used, one rich with associations, the other devoid of them. The first twenty-five subjects were asked to take part in an experiment in judgment. Each subject read the following instructions:

> "Arrange the 15 nonsense syllables in four piles according to ease of articulation. You will be given two minutes to do this,"

and was given a set of fifteen (15) cards on each of which was typewritten one syllable, and four cards labeling the four piles, "Very Easy," "Fairly Easy," "Slightly Easy," "Indifferent." If the arrangement was made before the two minutes had expired the subject was urged to make sure the arrangement was satisfactory. At the end of the two minutes the observer was asked to make a record of the syllables in each pile on the blank provided.

Twenty photographs with names attached and four cards labeling the piles "Very Attractive," "Fairly Attractive," "Slightly Attractive," "Indifferent," were given to each observer with the following instructions:

> "Arrange the 20 pictures in four piles as indicated. You will be given 2 minutes to do this."

After the two minutes had expired each subject was requested to record the names of the pictures in each pile on the blank provided. This was done to make sure the subject looked at the names under the pictures.

The subject's memory for both materials was tested. Three minutes were given for the recall of the syllables and then a selection of fifteen from a set of thirty, fifteen of which were in the original set, was made. The observer was asked to recall the names of photographs in three minutes. A set of forty photographs, twenty being those of the original set without names, was provided with the following directions:

> "Select 20 pictures from this group which you think were previously shown. Name any which you can."

After the subject had selected the photographs, a set of forty cards on each of which was typewritten a name (twenty were the names attached to the photographs in the original set, twenty were other names) was used for the recognition of names. Care was taken not to repeat any name, either first or last, nor to use the last name of any well-known person.

The second group of twenty-five subjects first copied the list of syllables and names on the pictures. This was done in order to control the conditions, keeping them, as far as possible, identical with those of the first group. If such a record had not been required for the group which was judging, one could not be sure that the subjects read the names. Even with this precaution, the subjects claimed they had never looked at them, until reminded they had written each on the blank.

After making these copies, the subjects of the second group were given a set of fifteen nonsense syllables with these instructions:

> "You will be given 2 minutes to look at these 15 syllables. Later you will be called upon to remember them, so look at each with a *determination to remember it.*"

After two minutes had expired, the set of photographs was presented with the directions:

> "You will be given 2 minutes to look at these 20 pictures. You will afterwards be called upon to remember the names and photographs, so look at each with a *determination to remember.*"

Recall and recognition were tested in the same way as for the first group.

Throughout the experiment an attempt was made to keep the conditions for both groups the same. Any sources of error or awkwardness in conducting the experiment, due to the difficulty of keeping the observers of the first group ignorant of the purpose of the investigation, which may give rise to criticism, will be found to be constant for both groups.

The purpose of the experiment was (1) to compare recall memory, with and without determination to remember, with recognition, with and without determination to remember, respectively; (2) to compare recall and recognition with determination and recall and recognition with no factor of determination present; and (3) (*a*) to compare recall of syllables, with and without the intention to remember, with recall of names with and without determination, and (*b*) to compare recognition of syllables with and without the intention to remember with recognition of photographs and names with and without the intention to remember.

In Table II. the total number recalled and recognized by both groups is given for both materials. Table III. shows the ratio of recall with a determining factor present to recall with no such factor present and the ratio of recognition with a determining factor to recognition without the determining tendency for both materials. Table IV. gives the ratio of recognition with determination to recall with determination and the ratio of recognition without determination to recall without determination for both materials.

TABLE II

SYLLABLES

	No Determination		Determination	
	Recall	Recog.	Recall	Recog.
Total	126	318	155	326
Average	5.0	12.7	6.20	13.0
M.V.	1.65	1.26	2.06	.73

PHOTOGRAPHS AND NAMES

	No Determination			Determination		
	Recall	Recog. N.	Recog. Ph.	Recall	Recog. N.	Recog. Ph.
Total	77	391	461	134.5	435	415
Average	3.08	15.6	18.4	5.36	17.4	16.6
M.V.	.90	1.53	1.48	1.57	1.44	1.90

TABLE III

SYLLABLES

Ratio of Determined Recall to Undetermined Recall	100: 81.3
Ratio of Determined Recog. to Undetermined Recog.	100: 97.5

PHOTOGRAPHS AND NAMES

Ratio of Determined Recall to Undetermined Recall	100: 57.2
Ratio of Determined Recog. to Undetermined Recog.	100: 100.2

TABLE IV

SYLLABLES

Ratio of Determined Recog. to Determined Recall	100: 47.55
Ratio of Undetermined Recog. to Undetermined Recall	100: 39.62

PHOTOGRAPHS AND NAMES

Ratio of Determined Recog. to Determined Recall	100: 15.82
Ratio of Undetermined Recog. to Undetermined Recall	100: 9.03

Since some subjects might recognize more names than photographs and others more photographs than names, the score for recognition includes the number of photographs and names recognized.

TABLE V

Ratio of Det. Rc. of SYLLABLES to Det. Rc. of NAMES	100: 86.8
Ratio of Undet. Rc. of SYLLABLES to Undet. Rc. of NAMES	100: 61.1
Ratio of Det. Rg. of SYLLABLES to Det. Rg. of NAMES and PHOTOGRAPHS	100: 260.6
Ratio of Undet. Rg. of SYLLABLES to Undet. Rg. of NAMES and PHOTOGRAPHS	100: 267.9

In Table V. the materials are compared. The ratios of recall of syllables to recall of names are presented with no determination and with determination to remember; the ratios of recognition of syllables to recognition of photographs and names with and without determination to remember are also stated.

It will be noticed that fewer photographs were recognized when there was a determination to remember present, but many more names. This does not mean that determination to remember decreases one's ability to recognize faces. The larger number of names recognized with determination to remember seems to indicate that the observers considered names more difficult to remember and spent most of the time allotted to them to learn names. The total number of photographs and names recognized with determination to remember was 850 and without 852, or an average in each case of 34 items. Throughout, the calculations involving the recognition of the material with greater meaning the totals 850 and 852 have been used. Frequently a subject could recall the first or last name but not both; in each case the score of one half was given (Table VI.).

TABLE VI

RECALL OF NAMES

	No Determination			Determination		
	1st Only	2d Only	Both	1st Only	2d Only	Both
Total	40	50	32	49	62	79
Total first names		72			128	
Total last names		82			141	

The data indicate the following: (1) the factor of determination to remember influences recall memory, but its effect on recognition is little, if any; (2) the difference between recall and recognition is less when there is a determination to remember the material than when there is no intention to remember; (3) the influence of determination for the recall of names is greater than for the recall of nonsense syllables; there is little, if any effect of determination to remember on the total number of items recognized of either material.

A further study of the data for the recall of names shows interesting results. When no determining factor was present 72 first names were recalled and with determination 128, or an increase of 77 per cent.; with no determination to remember 82 last names were recalled, with intention present 141, or an increase of 72 per cent.; with no determination total number recalled 77 and with determination 134.5, or an increase of 75 per cent. Determination to remember seems to influence the recall of first and last names to the same degree. Determination to remember influences greatly the number of first and last names correctly connected, as there were only 32 with no determining factor and 79 with one, or an increase of 147 per cent. Moreover, the determining tendency influences the number of photographs which may be correctly named. Table VII. shows

TABLE VII

NAMES CORRECTLY ASSOCIATED WITH PHOTOGRAPHS

	No Determination				Determination			
	1st	2d	Both	Total	1st	2d	Both	Total
Total	2	9	13	18.5	10	24	48	65.0
Highest score	1	3	4	4	3	4	7	8
Lowest score	0	0	0	0	0	0	0	0

the number of first, last, and whole names which were correctly assigned to the photographs for both groups of subjects. According to the method of scoring adopted, 18.5 names were correctly given to the photographs by the first group and 65 by the second.

The data seem to indicate that

Determined recall differs from undetermined recall more than determined recognition differs from undetermined recognition.

The difference between determined recall and determined recognition is less than that between undetermined recall and undetermined recognition.

The influence of a determining factor is greater for recall of material rich with associations than for material devoid of them.

The determining factor influences the amount of material remembered which can be correctly associated with other material remembered.

3. *Primacy and Recency*

The third part of the investigation was concerned with determining the influence of primacy and recency on recall and recognition.

Ninety-one subjects were tested for two materials, nonsense syllables and photographs. To nineteen, a list of twenty-four nonsense syllables was presented at the rate of one item every two seconds in a given order (nos. 1–24). Immediately after the presentation they were asked to recall all they could in three minutes and then were given a list of forty-eight from which to select 24 which they thought had been previously presented. The subjects were then shown twenty-four pictures at the rate of one every two seconds. As each was presented a name was pronounced by the experimenter. Immediately after the presentation, they were given three minutes in which to recall the names. A set of forty-eight photographs, containing the original twenty-four, was used for the recognition test. Each subject was then provided with a list of forty-eight names from which to select twenty-four which he thought had been pronounced

TABLE VIII

PRIMACY AND REGENCY

Material: Syllables. *Process:* Recall

Order of Presentation	Total 1–24	Total 13–24, 1–12	Total 7–24, 1–6	Total 19–2 1–18	Grand Total	Per Cent. of Av.
1	11	12	10	22	55	206.2
2	8	8	8	8	32	119.9
3	9	10	3	8	30	112.5
4	10	4	2	12	28	104.9
5	2	7	3	3	15	56.2
6	6	4	11	8	29	108.3
7	1	6	8	3	18	67.5
8	2	0	5	5	12	44.9
9	6	4	10	7	27	101.2
10	3	5	1	11	20	74.9
11	4	4	10	4	22	82.5
12	3	3	2	5	13	48.7
13	9	2	10	6	27	101.2
14	1	10	7	3	21	78.7
15	5	4	3	4	16	59.9
16	3	7	7	15	32	119.9
17	5	1	2	8	16	59.9
18	3	2	6	6	17	60.4
19	8	2	4	15	29	108.3
20	4	6	10	8	28	104.9
21	1	5	3	12	21	78.7
22	8	11	7	7	33	120.3
23	3	8	12	17	40	149.9
24	9	12	21	16	58	217.5
No. of Subjects	19	19	25	28	91	
Total					639	
Average per position				26.67	100	

TABLE VIII (*Continued*)

PRIMACY AND RECENCY

Material: Syllables. *Process:* Recognition.

Order of Presentation	Total 1–24	Total 13–24, 1–12	Total 7–24, 1–6	Total 19–24 1–18	Grand Total	Per Cent. of Av.
1	17	19	20	27	83	140.0
2	17	17	18	23	75	126.5
3	15	19	15	22	71	119.7
4	16	11	16	26	69	116.4
5	14	18	20	19	71	119.7
6	10	9	18	25	62	104.6
7	15	15	20	23	73	123.1
8	16	12	12	23	63	106.2
9	13	16	20	17	66	111.3
10	13	13	10	20	56	94.4
11	13	15	25	19	72	121.4
12	11	16	7	11	45	75.9
13	15	15	17	16	63	106.2
14	11	15	16	15	57	96.1
15	10	9	16	22	57	96.1
16	9	14	19	23	65	109.6
17	17	11	13	24	65	109.6
18	12	8	17	16	53	89.4
19	14	13	17	23	67	112.9
20	9	11	17	14	51	86.0
21	13	9	16	23	61	102.9
22	16	16	15	12	59	99.5
23	7	15	17	23	62	104.6
24	18	13	13	14	58	97.8
Total	321	329	394	480	1424	
No. of Subjects	19	19	25	28	91	
Average per position				59.33	100	

as the pictures were presented. The experiment was repeated on the second group of nineteen, but the order of each material was 13 through 24, 1 through 12; and on a third group of twenty-five and on a fourth group of twenty-eight subjects, where the orders were 7–24, 1–6 and 19–24, 1–18 respectively.

The results were calculated according to the position of the item, *i.e.,* for position one the record is given for the number of times item 1 was remembered by members of the first group, item 13 by those in the second group, item 7 by those in the third group, and item 19 by those in the fourth. For position two the record is given for item 2 by the first group, item 14 by the second group, item 8 by the third, and item 20 by the fourth; and so on until position twenty-four shows the number of items 24, 12, 6, 18, were remembered by groups one, two, three, and four respectively.

TABLE IX

Material: Names. *Process:* Recall.

Order of Presentation	Total 1–24	Total 13–24, 1–12	Total 7–24, 1–6	Total 19–24, 1–18	Total	Per Cent. of Av.
1	2.7	1.5	4.5	3.5	11.5	55.0
2	0	0	2.0	9.0	11.0	52.6
3	2.5	3.5	5.5	9.0	20.5	98.0
4	1.0	2.0	5.5	7.5	16.0	76.5
5	.5	3.0	6.5	10.5	20.5	98.0
6	2.5	1.5	9.0	5.5	18.5	88.5
7	1.0	1.5	6.5	5.0	14.0	67.0
8	.5	4.0	1.5	4.0	10.0	47.8
9	1.0	1.5	13.0	13.5	29.0	138.7
10	1.5	2.0	3.0	7.0	13.5	64.6
11	2.0	2.5	6.0	4.0	14.5	69.3
12	.5	1.5	5.5	5.0	12.5	59.8
13	5.5	2.0	2.0	2.5	12.0	57.9
14	.5	0	10.5	2.0	13.0	62.2
15	4.5	4.5	6.5	4.0	19.5	93.3
16	2.0	2.0	8.0	2.5	14.5	69.3
17	2.5	1.5	14.5	3.0	21.5	102.9
18	1.5	2.0	7.5	9.5	20.5	98.0
19	1.0	1.0	7.0	7.5	16.5	78.9
20	6.0	.5	2.5	5.0	14.0	67.0
21	2.5	2.5	7.5	9.5	22.0	105.2
22	7.5	3.5	8.0	11.5	30.5	145.9
23	7.0	13.0	10.5	14.5	45.0	215.3
24	11.5	19.0	23.0	28.0	81.5	390.0
Total	67.0	76.0	176.0	183.0	502.0	
No. of Subjects	19	19	25	28	91	

Average per position 20.9 100

TABLE IX (*Continued*)

Material: Photographs. *Process:* Recognition

Order of Presentation	Total 1-24	Total 13-24 1-12	Total 7-24 1-6	Total 19-24 1-18	Total	Per Cent. of Av.
1	11	16w	12	14	53	98.9
2	11	8	16	19	54	100.7
3	14w	14w	17	14	59	110.0
4	3	15	11	22w	51	95.2
5	9	11	15	25	60	111.8
6	11w	11	12	21	55	102.6
7	10	10	11w	17	48	89.6
8	15	17	15	11	58	108.2
9	13	7	17w	25w	62	115.7
10	4	12w	16	17	49	91.4
11	14	13	10	13	50	93.3
12	4	13	17	23w	57	106.3
13	16w	12	14	18	60	111.8
14	7	8	17	17	49	91.4
15	10w	16w	15	18	59	110.0
16	10	3	18w	8	39	72.7
17	15	10	19	19	63	117.5
18	12	13w	13	10	48	89.6
19	10	14	15	26w	65	121.3
20	15	14	9	10	48	89.6
21	4	11	21w	20w	56	104.5
22	13w	5	10	19	47	87.7
23	17	15	9	10	51	95.2
24	10	3	11w	21	45	83.9
Total	258	271	340	417	1286	
No. of Subjects	19	19	25	28	91	
Average per position					53.6	100

(w indicates photograph of a woman.)

TABLE IX (*Continued*)

Material: Names. *Process:* Recognition.

Order of Presentation	Total 1-24	Total 13-24 1-12	Total 7-24 1-6	Total 19-24 1-18	Total	Per Cent. of Av.
1	18	13	20	21	72	105.4
2	6	8	18	28	60	87.8
3	16	18	17	24	75	109.8
4	15	15	22	21	73	106.9
5	12	17	21	27	77	112.7
6	14	15	23	22	74	108.3
7	18	12	16	25	71	103.9
8	6	18	9	26	59	86.4
9	12	12	24	28	76	111.1
10	9	13	23	26	71	103.9
11	13	11	23	19	66	96.6
12	12	12	20	20	64	93.7
13	14	12	16	21	63	92.2
14	7	12	24	11	54	79.0
15	16	15	22	20	73	106.9
16	10	13	17	20	60	87.8
17	11	11	23	21	66	96.6
18	11	11	22	22	66	96.6
19	9	12	22	21	64	93.7
20	18	11	17	18	64	93.7
21	9	11	21	23	64	93.7
22	14	10	22	27	73	106.9
23	19	16	21	23	79	115.7
24	15	18	18	24	75	109.8
Total	304	316	481	538	1639	
No. of Subjects	19	19	25	28	91	
Average per position				68.20	100	

Primacy and recency both influence recall memory. The influence of each on recognition is less than on recall, but is greater for material devoid of associations and less for material rich with associations.

4. *Variations in Size and Color*

Peterson (15) studied the influence of complexity and dissimilarity on memory. Among other interesting facts, he determined the effect of variations in size and in color for recall. His results suggest the questions:

4. What is the influence of variation in color on recall and on recognition?

5. What is the influence of variation in size on recall and on recognition?

Peterson and Gordon experimented on series of nine nonsense syllables. Although it was desirable to make the conditions of the present experiments as nearly the same as possible as in these authors', longer series were necessary to insure exceeding the subjects' recognition ability for a perfect score. Sixteen syllables were used in each series. Four colors (red, green, blue, yellow) were presented each four times in the series to give the variation in color. The letters were cut from Hering papers the size of Willson's Gummed Letters No. 21, one half inch high; and one syllable of three letters of the same color was pasted on cards 4×6 inches. For the size variation the black Willson Gummed Letters 1/8, 3/8, 1/2, 1, 1 1/2 inches high were used. On each card was pasted one syllable, each letter in the syllable being the same size. The size of syllable varied in the series in the following order: 1/8 in., 3/8, 1/2, 1, 1 1/2, 1, 1/2, 3/8, 1/8, 3/8, 1/2, 1, 1 1/2, 1, 1/2, 3/8 inches. The standard series for both color and size were in black letters one half inch high. Four series were presented to each subject, one varied in color, one in size, two were in black.

Each subject was tested separately. One item in the series of sixteen was shown every two seconds; recall and recognition tests were made as in the previous experiments. The forty subjects were divided into four groups of ten each, designated by the letters *A, B, C, D*. The different series may be represented by the Roman figures I., II., III., IV. The addition of "color" or "size" to a numeral indicates that the series varies in color or size respectively. To Group *A* the material was presented in the following order: I. Color, II., III. Size, IV., *i.e.*, for Group *A* the first series was varied in color, the second was all black, the third varied in size, and the fourth was in size one half inch high. To Group *B* the material was presented in the order I., II. color, III., IV. size. Thus what was "standard" for Group *A* was "varied" for Group *B* and what was "varied" for Group *A* was "standard" for Group *B*. The score of series I. color for Group *A* and series II. color for Group *B* compared with the score of II. for Group *A* and I. for Group *B* shows the influence of color, as such, on the memory of these nonsense syllables. The score of III. size for *A* and IV. size for *B* compared with IV. for *A* and III. for *B* shows the influence of size variation for these series. In these two groups, *A* and *B*, the recognition set was similar to the presentation series, *i.e.*, if a presentation series varied in color, the recognition set had thirty-two items, eight red, eight yellow, eight blue, eight green; if the presentation series varied in size, the recognition set contained thirty-two items, four of size 1/8 inch, eight 3/8

[1] *Journal of Abnormal Psychology,* Vol. XI., No. 1, p. 2.

inch, eight ½ inch, eight 1-inch, four 1½ inch. It is difficult to secure material absolutely devoid of meaning, for, as Morton Prince has said, "Even nonsense syllables have the meaning nonsense."[1] The materials used in this part of the experiment were almost entirely devoid of associations. No syllable was repeated; the series were so devised that no syllable in red had the letter R in it, no syllable in yellow had a Y, none in blue had a B, nor had any in green a G. None of the letters in one syllable appeared in the next syllable in the series. All the letters were capitals; the use of capitals made the syllables more nonsensical and less easy to associate.

The presentations to Groups *C* and *D* resembled those made to Groups *A* and *B* respectively. The recognition series differed; to Groups *C* and *D* the recognition sets were always standard, *i.e.*, black size ½ inch. The records for all groups are given in Table X. It will be noticed that the recall records for Groups *A* and *C* and for Groups *B* and *D* respectively can be combined.

Peterson writes "In immediate recall the series varying in size gave slightly better results than the standard—a gain of 9 per cent., but that color variation was of no aid." His table shows:

	Immediate Recall	
Average for standard	6.31	M.V. 1.15
Average for color variation	6.35	M.V. .74
Average for size variation	6.89	M.V. .68

The results of the present experiment show:

	Immediate Recall (40 Subjects)	
Average for standard	2.70	M.V. 1.22
Average for color variation	2.58	M.V. 1.02
Average for standard	2.63	M.V. 1.25
Average for size variation	2.75	M.V. 1.23

Tables X., XI. show the results for both recall and recognition. Color variation does not increase the amount recalled or recognized so far as these experiments are concerned. The variation in size of the syllables does not affect recall or recognition. Color and size variation, as such, present in this meaningless material, did not influence either recall or recognition.

In the study of the possible influences which these factors might have upon recall and recognition no characteristic differences between the processes are noticed. A factor which may have shown a tendency to influence recall in one way showed the tendency in the same direction for recognition, although the extent or degree of the influence may have differed. To study the relation of the processes in different materials and among different groups of subjects was the purpose of the experiments discussed in the following chapters.

TABLE X

RECALL

A II.	19		*A* I. color	25	
B I.	27	46	*B* II. color	23	48
C II.	31		*C* I. color	26	
D I.	31	62	*D* II. color	29	55
Total (40 subjects)		108			103
Average		2.70			2.58
M.V.		1.22			1.02

RECOGNITION

A II.	100		*A* I. (color)	113	
B I.	114	214	*B* II. (color)	103	216
			Average (20 subjects)		10.80
C II.	104		M.V.		1.34
D I.	119	223	*C* I. (pre. color)	108	
Total (40 subjects)		437	*D* II. (pre. color)	105	213
Average		10.93	Average		10.65
M.V.		1.34	M.V.		1.07

Presented in black, recognition in black ...... Av. 10.93 (40 subjects)
Presented in color, recognition in color ...... Av. 10.80 (20 subjects)
Presented in color, recognition in black ...... Av. 10.65 (20 subjects)

TABLE XI

RECALL

A III. (size)	28		*A* IV.	29	
B IV. (size)	25	53	*B* III.	23	52
C III. (size)	28		*C* IV.	21	
D IV. (size)	29	57	*D* III.	32	53
Total (40 subjects)		110			105
Average		2.75			2.63
M.V.		1.23			1.25

RECOGNITION

A III. (size)	110		*A* IV.	106	
B IV. (size)	109		*B* III.	102	208
Total (20 subjects)		219			
Average		10.95			
M.V.		.97			
C III. (pre. size)	107		*C* IV.	109	
D IV. (pre. size)	110		*D* III.	110	219
Total (20 subjects)		217	Total (40 subjects)		427
Average		10.85	Average		10.68
M.V.		1.30			

Presented in standard, recognition set standard .... Av. 10.68
Presented in variation, recognition set standard .... Av. 10.85
Presented in variation, recognition set variation .... Av. 10.95

CHAPTER IV

Methods and Procedure

1. *Materials*

The foregoing experiments seem to indicate that the degree of difference between recall and recognition, when different factors are present, varies with the material used. In general, if there is any noticeable difference in the influence of a factor on recall and recognition, it is greater for material rich with association. In the present study four kinds of materials are used; words, geometrical forms, proverbs, and nonsense syllables. There are twenty-five items of each material in the presentation series and fifty in the recognition series. The words are all simple nouns—hand, chicken, letter, coal, umbrella, kite, etc. The forms are similar to those employed by Whitley and later by Simpson. Proverbs were collected for several months before the experiment was begun, by noting all the proverbs seen or heard and by reference to collections of proverbs. It was desired to have all the proverbs "reasonably short" and "about the same length"—*i.e.*, the proverb must take only one line on the card and must not be so much longer than the one above or below it to be easily recognized on the recognition blank on account of size. The attempt was made to have as many proverbs that might be familiar as those that might not be so familiar in the presentation and control series. Among the proverbs used are: "Enough is as good as a feast," "No road is long with good company," "Pleasing everybody is pleasing nobody." The nonsense syllables used have three letters, the first and the last being consonants and the middle one a vowel—zof, dej, zaf, etc. The series are printed in small letters but made as devoid of associations as possible—none of the letters in a syllable appeared in the syllables above or below it in the list.

There are two presentation series and two recognition series for each material, except words for which there are four series. The items for each material, except forms, are printed in a single column on a separate card in type, Eight Point Scotch; each card is headed by the word Words or Proverbs or Syllables according to the material. The twenty-five forms are printed in five rows of five items each on a separate sheet. For the recognition tests fifty items of each material, twenty-five being the same as in the presentation series, are printed in two columns (one column for proverbs) in the same type on separate blanks.

In the tests for children the only materials used are words, forms and syllables. Proverbs are included in the adult tests as it is desired to have a material for which the subjects might have a wealth of associations. This material is not used for children as it is probably not familiar to them. Some of the younger children read slowly and could not finish reading the list of proverbs in the given time.

The selection of the number of items and the time allowed for study are based on previous experiments. It is desired to have the series of all the materials the same length, and long enough that none or few persons tested would have a perfect score for recognition. Of the 734 persons tested one had a perfect score.

For clearness in the later discussion we shall call the first set presented *A* and the second *B*. The first items in the series for words, proverbs, syllables, words, for set *A* are flag, A rolling stone gathers no moss, qoh, glove respectively and for Set *B* hand, Make hay while the sun shines, zof, steeple, respectively.

2. *Methods of Scoring*

When recall memory has been tested by other investigators the method of scoring by items has been to credit the subject with a unit for each item remembered—*i.e.*, if he recalls five words or syllables his score is 5, if six times, 6, and so on. Washburn has scored ⅔ when two out of three letters in a nonsense syllable are recalled. One may question if it is not a more difficult task to remember one more item after having learned ten, than one more after having learned five, but it is customary in experimental studies in recall when scoring in terms of items (not time) to call the first record 11 and the second 6, thus giving the extra item in each case the score of 1. This method of scoring does not need empirical formulæ and gives a simple system. The use of percents in scoring recall is often misleading. If *X* has Recall Score 10, 25 words having been presented, he has recalled 40 per cent., *Y* has Recall Score 5 and has recalled 20 per cent., but is *X*'s memory twice as good as *Y*'s? We should not say it necessarily is. Again, *M* has score of 11 and *N* a score of 6—*M*'s score in terms of percents is 44 per cent. of the total, *N*'s score is 24 per cent.; *M*'s score 4 per cent. greater than *X*'s score, and *N*'s score 4 per cent. greater than *Y*'s score. Is the difference between *X*'s and *M*'s memory the same as that between *Y*'s and *N*'s? We should not say it necessarily is. Percents in the scoring of recall memory are often misinterpreted. Throughout this report the meanings of "Recall Score 5" and "Recall Score 10," etc., are "5 items recalled correctly out of 25" and "10 items recalled correctly out of 25," etc., respectively.

In the experiments reported in the foregoing chapters the recognition score is the number recognized correctly when the subjects are required to select a certain number (number originally presented). This method has been criticized adversely because it is thought that the "guessing" factor is not sufficiently controlled. Strong (16) has devised a formula,

$$\frac{\text{Correct recognitions}}{\text{Total number presented}} \times \frac{\text{Correct recog.} - \text{Incorrect recog.}}{\text{Correct recog.} + \text{Incorrect recog.}} \times 100$$

which he has used extensively in his work on recognition memory of advertisements. Myers has also employed Strong's formula. In explaining his formula Strong writes: "There are three factors which must be taken into account in obtaining a fair summary. There is first the number of advertisements that are correctly recognized—relationship between the number recognized and the total number that should be recognized. There is second the accuracy of the recognitions—the relationship between the number of correct and the number of incorrect recognitions. And there is third the general scheme of the experiment. In this experiment the subject had to select from an equal number of right and wrong advertisements. . . . Turning now to the first factor we see at once that by reducing the total number recognized to per cent. of all that should have been seen we can compare directly the results from, say, a series of 5 advertisements with results from other series. Such comparison is expressed by the formula

$$\frac{\text{Correct recognitions}}{\text{Total number presented}}$$

In presenting the second factor we must recognize that when there is an equal chance of selecting a right or wrong advertisement (when an equal number of each are presented as in this experiment) a record of 50 per cent. correct recognitions means nothing but pure chance. This 50 per cent. recognition really means nothing else than *zero memory* for although the subject has picked out x advertisements correctly from the n advertisements presented originally yet he has picked out an equal number incorrectly from those advertisements which had not been presented to him. . . . *Perfect memory,* on the other hand, would be, of course, where the n advertisements presented were all recognized and none of the wrong advertisements were selected. The following formula,

$$\frac{\text{Correct recognitions} - \text{Incorrect recognitions}}{\text{Correct recognitions} + \text{Incorrect recognitions}}$$

will give 100 under these conditions corresponding to perfect memory as defined above, 0 under those conditions corresponding to zero memory as defined above, and equal steps between the two extremes as the factors vary successively.

"By combining the two formulæ and multiplying the results by 100 to have it read in terms of per cent. instead of a decimal we have

$$\frac{\text{Correct recognitions}}{\text{Total number presented}} \times \frac{\text{Correct recognitions} - \text{Incorrect recognitions}}{\text{Correct recognitions} + \text{Incorrect recognitions}} \times 100$$

This formula combines the per cent. of correct recognitions among the possible recognitions with the accuracy of the recognitions."

Woodworth has suggested a simple scoring method for recognition: Total number—two times errors—omissions=Recognition score. He writes: "The subject's score should be a measure of the memory factor in determining his reactions and is found according to the following principles: evidently the cases where the subject says 'I don't remember' give no evidence of the memory factor. The cases where he says, 'It was shown before' or 'It was not shown before' (Yes or No), on the other hand, purport to represent the subject's memory and if they contain no errors they can be accepted as due wholly to memory and their number as giving a measure of memory. But if there are errors the number of Yeses and Nos can not be accepted at full value. Nor can the number of correct Yeses and Nos be taken as the measure since the presence of errors shows that something like guessing came in. Now, under the conditions of this experiment a guess is equally likely to turn out right as wrong and accordingly we assume that as many are found among correct answers as among the incorrect and find the number of guesses by taking twice the number of errors. Adding twice the number of errors to the 'Don't remember' cases we obtain the total number of cases in which no memory is in evidence and subtracting this from the whole number of answers, or 50, we obtain the subject's score.

"A rather better defence of the above method of scoring would substitute for the word guess some such expression as 'tendencies to react that are not derived from memory of the presentation.' There may be present in the subject numerous tendencies leading towards an answer Yes or No but all these tendencies can be divided into two classes. Those derived from the presentation or from the 25 stimuli originally shown and those derived from other sources. The tendencies of the first class tend always toward right answers, while those of the second class tend equally toward right and wrong answers. Any given reaction may be the resultant of both sorts of tendencies, but in the long run the effect of the tendencies of the second class must show itself by producing an equal number of right and wrong answers and therefore a measure of these tendencies can be obtained by taking twice the number of errors. This procedure, like all others found on probability, may do violence to the fact in the single case but comes out right in the total.''

The formula cos π U may also be employed to determine the score for recognition.

Strong's scores are presented, on a per cent. basis, a perfect score being represented by 100 and the lowest score by 0. For Woodworth the scores run theoretically from the total number of possible recognitions to 0. With the cos π U formula the scores run from 100 to 0. Thus all give similar scores for "all right" and "all wrong." The following examples will illustrate the differences in scoring.

EXAMPLE I

Total Number of Items in Recognition Series 10

Total Number of Recog.	Correct Recog.	Incorrect Recog.	Strong's Score	Step	Wood-worth's Score	Step	cos πU	Step
10	10	0	100	28	10	2	1.0000	.0490
10	9	1	72	24	8	2	.9510	.1421
10	8	2	48	20	6	2	.8089	.2212
10	7	3	28	16	4	2	.5877	.2788
10	6	4	12	12	2	2	.3089	.3089
10	5	5	0		0		.0000	

Strong's penalties decrease as the number of wrong recognitions increases, the penalties in the cos π U formula increase as the number of wrong recognitions increases, while Woodworth's penalties remain the same. Thus when 50 items are given, suppose X gets 45 right and 5 wrong and Y has 46 right and 4 wrong, the difference between X's and Y's scores is 5.3, but if M gets 35 right, 15 wrong and N 34 right and 16 wrong the difference in their score is 3.5. According to Woodworth's formula X's score is 40, Y's 42, M's 20, N's 18; the difference in X's and Y's score is the same as the difference between M's and N's scores, 2.

Strong's formula is based on the argument that to recognize one more after having recognized 45 correctly is more difficult than one

EXAMPLE II

Total Number Recog.	Correct Recog.	Incorrect Recog.	Strong's Score	Step	Woodworth's Score	Step	cos πU	Step
50	50	0	100.0	5.9	50	2	1.0000	.0018
	49	1	94.1	5.8	48	2	.9982	.0058
	48	2	88.3	5.4	46	2	.9924	.0098
	47	3	82.7	5.4	44	2	.9826	.0138
	46	4	77.3	5.3	42	2	.9688	.0178
	45	5	72.0	5.1	40	2	.9510	.0215
	44	6	66.9	5.0	38	2	.9295	.0251
	43	7	61.9	4.8	36	2	.9044	.0287
	42	8	57.1	4.6	34	2	.8757	.0318
	41	9	52.5	4.5	32	2	.8439	.0350
	40	10	48.0	4.3	30	2	.8089	.0382
	39	11	43.7	4.1	28	2	.7707	.0414
	38	12	39.6	4.1	26	2	.7293	.0445
	37	13	35.5	3.9	24	2	.6848	.0473
	36	14	31.6	3.6	22	2	.6375	.0498
	35	15	28.0	3.5	20	2	.5877	.0419
	34	16	24.5	3.4	18	2	.5358	.0539
	33	17	21.1	3.2	16	2	.4819	.0559
	32	18	17.9	3.0	14	2	.4260	.0578
	31	19	14.9	2.9	12	2	.3682	.0593
	30	20	12.9	2.7	10	2	.3089	.0604
	29	21	9.3	2.6	8	2	.2485	.0612
	28	22	6.7	2.4	6	2	.1873	.0620
	27	23	4.3	2.1	4	2	.1253	.0625
	24	24	2.1	2.1	2	2	.0628	.0628
	25	25	0		0		.0000	

more after 15. The argument may be right, but Strong is not convincing in showing why the step should be 5.9 between perfect score and score of one wrong when 50 items are in the series, 3.5 between 15 wrong and 16 wrong when 50 items are in the series, 28 between perfect score and one wrong when there are 10 items in the series.

For purposes of comparing recall and recognition the formula for recognition more nearly comparable with that used for recall is that of Woodworth. To employ Strong's formula or the cos π U formula would necessitate an empirical formula for recall. The chances of a guess in recall are not nearly so great, but the question as to the greater difficulty of recalling one more after having recalled 10 arises. The results of Strong's formula for recognition are not sufficiently convincing of its worth to warrant an attempt to devise an empirical formula for recall.

The scoring "Total Correct" will be used throughout the following chapters for recall. For recognition the score will be "Total number—twice number of errors—omissions (if there are any)."

The possible criticism for each and the full meaning of the figures must not be forgotten.

The coefficients of correlation in the following chapters were obtained by the statistical method explained by Woodworth in the *Psychological Review,* March, 1912, Vol. XVI., pages 97–123.

3. *Groups Tested and Procedure*

Two groups of adults were tested, in one there were twenty-eight women, twenty-four men; in the other twenty-four women and twenty men, making ninety-six persons.

The material was presented in the following order: (Set *A*) words, geometrical forms, proverbs, nonsense syllables, words; (Set *B*) words, geometrical forms, proverbs, nonsense syllables, words. To the subjects in the group of fifty-two adults a test for recall memory was made after each of the last five series (Set *B*). The tests were given separately for each subject. Each subject was given the following written directions:

> You will be shown different kinds of material and afterwards an attempt will be made to determine how much you have remembered. The material will vary—there will be series of words, of forms, of proverbs, and of syllables. The series will always be the same length—25 items. Your memory will be tested by different methods—sometimes by asking you to write down what you remember, and sometimes by showing you the material again mixed with other items which have not been seen and requesting you to state which you had seen and which you had not seen.
>
> The time given to read the presented series will always be the same—50 seconds. In order that you may have some idea how long that is, before the experiment begins a picture will be shown for 50 seconds.

Teachers, business men, students, housekeepers, etc., were among the subjects, most of them were undergraduate or graduate students in Columbia University.

Tests for recall of words, recall of forms, recall of syllables, recognition of words, recognition of forms, recognition of syllables were made on six hundred thirty-eight children, two hundred eighty-five boys and three hundred fifty-three girls in a large city public school. The children were in twenty-two classes in the ten grades *4A, 4B, 5A, 5B, 6A, 6B, 7A, 7B, 8A, 8B*. Each class was tested as a group; the classes were mixed so that the boys and girls were tested under the same conditions. The following table shows the number of girls and boys in each grade.

	Girls	Boys
4*A*	50	41
4*B*	39	29
5*A*	47	39
5*B*	30	41
6*A*	36	26
6*B*	37	22
7*A*	11	25
7*B*	31	22
8*A*	34	21
8*B*	38	19
	353	285

The following directions were given orally to each group tested:

I am going to show you some things and afterwards find out how much you can remember. Sometimes I shall show you words, sometimes forms, such as circle or a square, and sometimes nonsense syllables—nonsense syllables always have three letters put together in such a way that they do *not* make sense. Sometimes I shall find out how much you can remember by asking you to write down afterwards how many you remember and sometimes by showing you them again mixed with some that you have not seen before. There will always be 25 things to study and you will always be given 50 seconds in which to study them. In order that you may have some idea how long that time is I shall show you a picture for 50 seconds before we begin.

It was desired to give the children some idea of how long the period of fifty seconds is. To have timed empty space would have caused an illusion of time to enter. The gazing at a picture served also to get the children's attention and interest. To keep the conditions the same for adults a picture was shown to them too. Each subject had his or her material so that any advantage or disadvantage due to his or her position in the room might be avoided. The materials were kept in different colored envelopes.

Nineteen insane persons were tested each separately. The results of these subjects will be considered in Chapter VI.

CHAPTER V[1]

Relation between Memory Processes; Recall and Recognition

1. *Comparison of Different Series of the Same Material*

(See charts on following pages)

It has been stated in the preceding chapter that there are two sets of material. Set *A* and Set *B* complete each contain a series of twenty-five words, twenty-five forms, twenty-five proverbs, twenty-five syllables and a second series of twenty-five words. In every case

TABLE XII

Set *A*

52 Subjects	Recall Words	Recall Forms	Recall Proverbs	Recall Syllables	Recall Words
Average	9.23	6.17	5.52	3.44	11.54
A.D.	2.80	2.26	1.39	1.09	3.31
P.E.	.33	.27	.16	.13	.39

Set *B*

44 Subjects	Recall Words	Recall Forms	Recall Proverbs	Recall Syllables	Recall Words
Average	8.27	6.77	5.52	2.73	9.59
A.D.	2.41	2.02	1.64	1.67	2.66
P.E.	.31	.26	.21	.21	.34

Set *B*

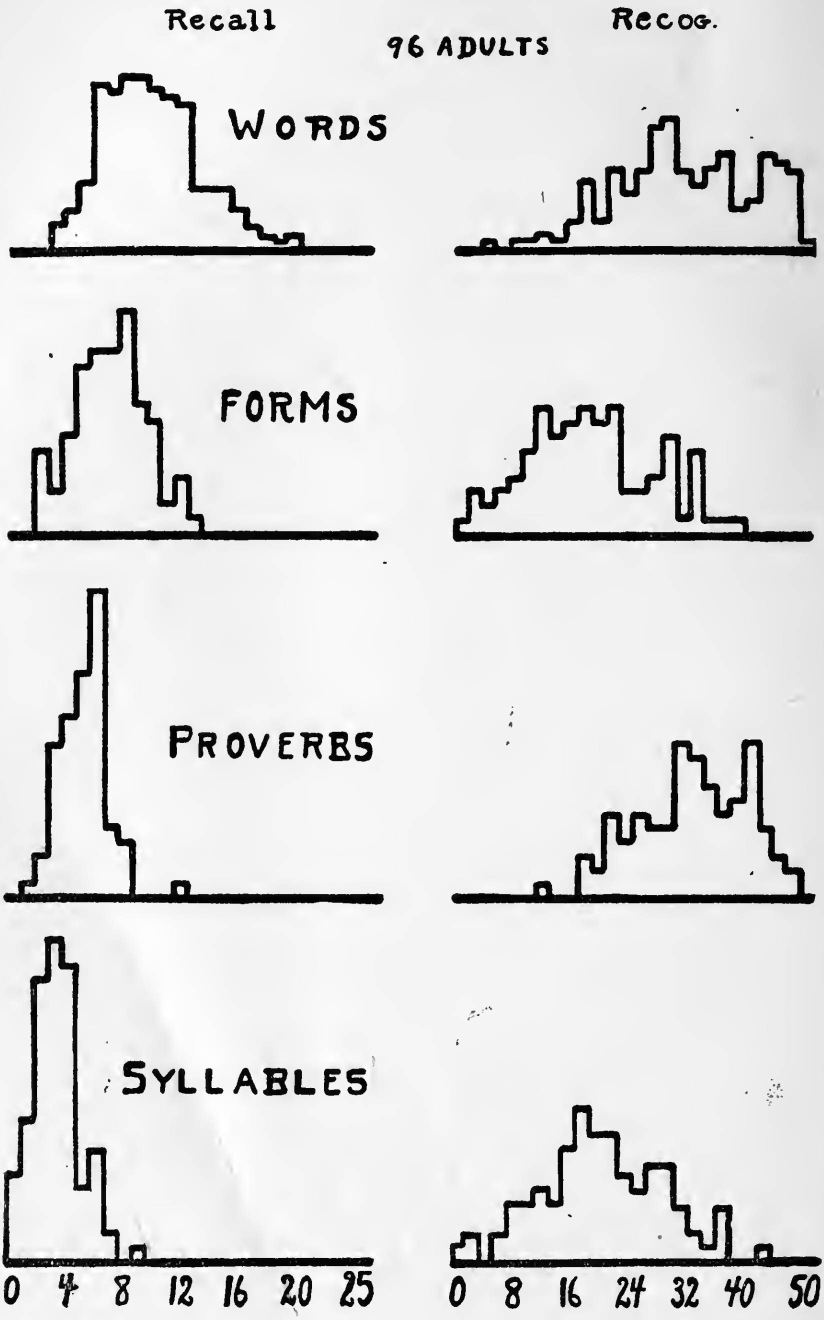
52 Subjects	Recog. Words	Recog. Forms	Recog. Proverbs	Recog. Syllables	Recog. Words
Average	32.52	13.52	33.12	18.65	33.46
A.D.	8.79	5.21	6.72	6.56	8.18
P.E.	1.03	.61	.79	.77	.96

Set *A*

44 Subjects	Recog. Words	Recog. Forms	Recog. Proverbs	Recog. Syllables	Recog. Words
Average	29.68	25.54	34.14	23.82	33.59
A.D.	7.30	6.68	6.47	6.64	6.97
P.E.	.93	.85	.82	.85	.89

Set *A* was presented to the subject first and Set *B* second. For the group of fifty-two adults the test for recall was made for Set *A*, for recognition for Set *B*. In the group of forty-four adults the recognition test was made for Set *A* and the recall test for Set *B*.

[1] Some of the data in this chapter have been published by the author in "Tests of the Memories of School Children," Edith F. Mulhall, *Journal of Educational Psychology*, May, 1917.


Recall
Recog.
96 ADULTS
WORDS
FORMS
PROVERBS
SYLLABLES
0 4 8 12 16 20 25
0 8 16 24 32 40 50

The materials in Set *A* were recognized by forty-four adults and recalled by fifty-two adults. The materials in Set *B* were recalled by forty-four adults and recognized by fifty-two adults. Thus one group recalled the material which the other group recognized. In this way, one might see whether the series in the two sets were about equally difficult. Table XII. shows the average, the average deviation, and the P.E. of the average for these groups.

The different series of the same material appear to be about equally difficult. In each case, except for the recognition of forms, the difference between the scores is small and the deviation fairly large. More words in Set *A* were recalled by the group of fifty-two than in Set *B* by the group of forty-four. The average score of the second series of words is slightly higher in both recall and recognition for both groups. The forms in each set seem to be equally difficult to reproduce, but those in Set *A* harder to recognize. This is probably due to the control figures being more confusing and more like the presentation series in Set *A* than the control figures in Set *B* are like the presentation figures in Set *B*.

The scores for recall for the two groups have been combined and the scores for recognition for the two groups have been combined giving the following results for ninety-six subjects.

	Words	Forms	Proverbs	Syllables	Words
Recall	8.75	6.47	5.55	3.09	10.57
Recog.	31.10	...	33.63	21.24	33.53

2. *Recalling and Recognizing of the Same Material*

The complex question concerning the relation between the results obtained by testing memory by recall and by recognition is reduced to several simple ones which are stated in Chapter II.

The first inquires about the relation between the recalling and recognizing of the same material. Does the person who recalls one kind of material well also recognize that kind of material well?

The coefficients of correlation between recall of words and recognition of words, between the recall of forms and the recognition of forms, between the recall of proverbs and the recognition of proverbs, between the recall of syllables and the recognition of syllables, were calculated for the two groups of adults. The coefficients of correlation between the recall of words and the recognition of words, between the recall of forms and the recognition of forms, and between the recall of syllables and the recognition of syllables by the children were computed for each grade. Table XIII. gives these coefficients of correlation.

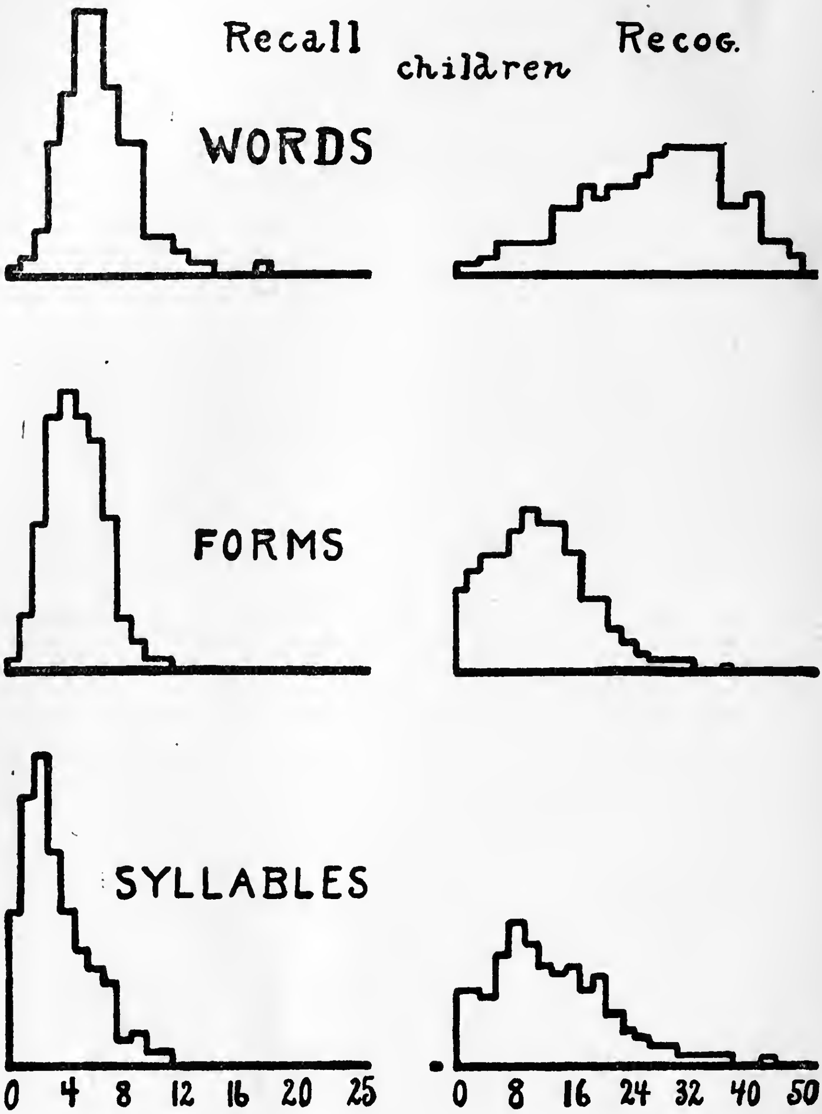

Recall
children
Recog.
WORDS
FORMS
SYLLABLES
0 4 8 12 16 20 25
0 8 16 24 32 40 50

TABLE XIII

Adults	Rc. Set *A* Rg. Set *B*	Rg. Set *A* Rc. Set *B*
Recall Words Recog. Words	.24	.18
Recall Forms Recog. Forms	—.18	—.16
Recall Proverbs Recog. Proverbs	.42	—.08
Recall Syllables Recog. Syllables	.32	.74
Recall Words Recog. Words	.24	.34
Average	.21	.25
Recall Words Recog. Words, corrected by attenuation	.75	.52

Children	4*A*	4*B*	5*A*	5*B*	6*A*	6*B*	7*A*	7*B*	8*A*	8*B*	Av.
Recall Words Recog. Words	.12	.18	.38	.52	.23	.06	.10	.22	.44	.44	.25
Recall Forms Recog. Forms	.34	.28	.00	.06	.42	—.16	.32	.12	.22	.34	.21
Recall Syl. Recog. Syl.	.34	.18	.34	.00	.22	.14	.12	.16		.06	.17

If the coefficients had been near —1.00 it would have indicated that those who recalled well recognized poorly and if the coefficients had been 1.00 it would have indicated that those who recalled well recognized well. The coefficients are low; the average (without correcting for attenuations) for the adult group being .23, for children .21 with a large P.E. Coefficients corrected for attenuation are higher. Before drawing any conclusions from these data, an examination of the following tables may be made.

TABLE XIV

Recall Words

Recognition Words	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
50									1								
48					1				2		1	1	4		1		
46									2		4	3	2	1			
44										3	3	4	1	1			1
42																	
40			1	2		1	1	5	7	2	2	3	5	1	1	1	1
38				2					1	3	7	8	7	2			
36						3		1	7	6	3	6	4	8			
34			2				2		7	6	7	5	8	3	3		
32							1	2	4	4	1	12	4	5	1	1	
30						1	1		2	4	8	9	6	5	2		
28									1	4	6	7	8	8	3	2	
26								2	3	5	6	8	3	3	8		
24	1						1	2	4	2	4	2	8	5	4	2	1
22								2	6	5	2	7	6	4	3	3	
20								1	1	2	3	9	5	5	1	2	
18								1	1	2	3	11	9	7	4		
16								1	1	1	3	4	3	6	5	6	1
14									1	1	3	4	8	3	3	1	1
12										2	1	3	8	1	2		
10											1		4	2	1	3	
8										1	1		2	2	1	1	
6									1	1	2	1		4	4	1	1
4										3	1		1	2			
2												1	1	2	1		1
0													1			2	
— 2														1	1		
—12										2							
—18														1			
—20															1		

RECALL FORMS

Recognition Forms	11	10	9	8	7	6	5	4	3	2	1	0
38						1						
32					1			1				
30					1	1	1					
28					1		3				1	
26					2	3		1				
24				1	2	2	2	2	2	1	2	
22			2	2	5	2	2	2	5	1		
20			2		2	1	5	6	6	5		
18			1		5	9	7	8	1	4		
16				4	9	10	14	10	4	5		
14			2	3	8	8	9	13	5	6	5	
12		1		5	5	9	10	10	11	7	4	
10	1	1		2	8	14	12	7	14	12	1	
8			2	3	5	8	8	9	12	4	5	
6		1	2	1	3	7	9	9	12	2	1	
4				1	2	5	7	11	10	9	1	1
2					4	6	11	12	8	5	3	
0			1			4	1	10	5	1		1
— 2			1	1	1		3	2	5	2		1
— 4						1	5	2	2			
— 6						1	1	3		2		
— 8						1		2	4	1		
—10							1					

RECALL SYLLABLES

Recognized Syllables	9	8	7	6	5	4	3	2	1	0
44										1
42										
40		1					1			
38					1			1		
36								1		
34								1		1
32			1				1	1	1	
30					1	1		4	3	
28							1		1	
26						2	4	3	1	
24			2		1	3	4	3	3	
22			1	4		2	5	5	6	3
20			2	1		9	7	10	7	2
18					3		8	10	5	6
16					4	4	13	10	11	1
14						7	3	10	10	8
12				1	2	4	12	9	9	7
10	1		1		3	7	5	21	10	6
8			1	1	2	7	8	20	14	12
6				1	1	3	7	14	16	7
4						2	2	11	9	6
2	1			1		4	5	7	7	5
0						1	3	10	8	12
— 2			1		1	1	2	7	7	2
— 4						1	1	5	2	2
— 6								2	1	1
— 8						1				1
—10										
—12									1	

Along the horizontal line are represented the scores for recall, along the vertical are the scores for recognition. The numbers in the diagram represent the number of children who had the different score, thus four persons with the score of 5 for recall of words had the score of 48 for recognition, two persons with the recall score of 5 had recognition score for words of 46.

The tables show the recall and recognition scores for all the children. If the correlation between recall and recognition were great—*i.e.*, if the children who recalled well recognized well, most of the cases would lie along the diagonal from the upper left side to the lower right side, and if there is a negative correlation, most of the cases would lie along the other diagonal, from upper right side to lower left side. Neither of these situations occurs in any cases—there is no high positive nor negative correlation. In the recall of words twenty children have a score of over eleven and in recognition thirty-five had a score of forty-two or over, but only one of the twenty is in the group of thirty-five. The best scores for recall are 17, 15, 15, 14 and the recognition scores for these subjects are respectively 24, 34, 34, 40. The best scores for recognition are 50, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, and the recall scores for the subjects having these scores or respectively 9, 13, 9, 9, 7, 6, 5, 5, 5, 5, 3. The seven worst scores for recall are all one (1) and the recognition scores of these subjects 44, 40, 24, 16, 14, 6, 2. In the test for forms four children had a recall score of 10 or over and six a recognition score of 30 or over. None of the four is among the six. The best scores for recall are 11, 10, 10, 10, and the subjects with these recall scores had the following scores for recognition, 10, 12, 10, 8 respectively. The best scores for recognition of forms are 38, 32, 32, 30, 30, 30 and the recall scores of the same subjects are respectively 6, 7, 4, 7, 6, 5. The three subjects with zero scores for recall of forms were among the 113 subjects who had 4 or less for their recognition scores.

In the test for syllables three children have the recall scores 8, 9, 9; their recognition scores are 40, 10, 2 respectively. The three best recognition scores are 40, 40, 44 and the recall scores of the same subjects 8, 3, 0, respectively. Those with recognition scores below 4 have recall scores ranging from 9 to 0; recall scores of zero have high and low recognition scores.

From the coefficients of correlation and from the tables one finds no very high negative nor positive correlation. These data would indicate that a person who recalls a certain material well may recognize that kind well, fairly well, or poorly—we know little about one's recognition memory from a test of recall. There is a strong tendency, in general, for the correlation to be positive rather than negative. The grade-groups of children on the average do not differ from the groups of adults in degree of correlation any more than the different grade-groups of children differ among themselves. There is no marked increase in degree of correlation from grade to grade. Except for the low correlation among adults for forms, there is no consistent difference in degree of correlation between recall and recognition material.

3. *Recall of Different Materials*

Earlier investigators have shown that those who recall one kind of material well may not recall another kind well. The present results may be examined for a confirmation or contradiction of their conclusion. The coefficients of correlation between the recall of words and recall of forms, between the recall of words and the recall of proverbs, between the recall of words and the recall of syllables, between the recall of forms and the recall of proverbs, between the recall of forms and the recall of syllables, between the recall of proverbs and the recall of syllables for both groups of adults and the coefficients of correlation between the recall of words and recall of forms, between the recall of words and syllables and between the recall of forms and syllables for each of the ten grades of school children are given in Table XV.

TABLE XV

GROUP OF 52 ADULTS—CORRELATION BETWEEN THE RECALL OF DIFFERENT MATERIALS IN SET *A*

	Words	Forms	Proverbs	Syllables	Words
Words		.84	.52	.48	.38
Forms	.84		.12	—.02	
Proverbs	.52	.12		.12	
Syllables	.48	—.02	.12		
Words	.38				

GROUP OF 44 ADULTS—CORRELATION BETWEEN THE RECALL OF DIFFERENT MATERIALS IN SET *B*

	Words	Forms	Proverbs	Syllables	Words
Words		.16	.18	.16	.52
Forms	.16		—.12	—.04	
Proverbs	.18	—.12		—.04	
Syllables	.16	—.04	—.04		
Words	.52				

GROUP OF 638 SCHOOL CHILDREN—CORRELATION BETWEEN RECALL OF DIFFERENT MATERIALS

Number in each grade	91	68	86	71	62	59	36	53	55	57
Grade	4*A*	4*B*	5*A*	5*B*	6*A*	6*B*	7*A*	7*B*	8*A*	8*B*
Rc Words .... Rc. Forms....	—.06	—.10	—.04	.18	—.10	.04	—.28	.42	—.06	.32
Rc. Words.... Rc. Syl.	.18	—.02	.16	.28	.18	.30	.00	.24		.12
Rc. Forms.... Rc. Syl.	.12	—.08	—.12	—.10	.22	.18	.10	.10		.44

The coefficients average about zero. The data confirm the results of previous investigations that a person who recalls one material well may not recall another material well. This appears to be true for children and for adults.

The coefficients vary, but there is a tendency for them to be positive rather than negative. Five of the six coefficients between the recall of different materials in Set *A* are positive and the total of the positive coefficients is 1.84 and of the negative .02. Three of the six coefficients for recall between different materials in Set *B* are positive and the total of positive coefficients is .52 and the negative .28. Eighteen of the twenty-eight coefficients among the groups of children for recall of different materials are positive; the total of positive coefficients is 3.58 and negative .96.

The coefficients of correlation for recall of words with other materials average for Set *A* .61, for Set *B* .13; the coefficients of correlation for recall of forms with recall of other material is for Set *A* .31, for Set *B* .01; the coefficients for recall of proverbs with the recall of the other materials for Set *A* .25, for Set *B* .01; the coefficients for recall of syllables with recall of other materials is for Set *A* .19, for Set *B* .03. The average of the coefficients of recall of words with recall of forms and recall of syllables among children is .095, recall of forms with recall of words and recall of syllables is .07, recall of syllables with recall of words and recall of forms .12.

There is little difference in materials—the coefficients for all are low and the P.E. large, with the exception of recall of words with recall of other materials in Set *A* by the adults. From the results among the adults it might appear that words are a better index of one's recall memory than the other materials, but the children's records do not indicate this.

4. *Recognition of Different Materials*

Does the person who recognizes one material well recognize another well?

The coefficients of correlation between recognition of different materials for both groups of adults and for each grade of school children are given in Table XVI.

The average of the coefficients of correlation between the recognition of different materials in Set *A* is .37 and for Set *B* .28. The average of the coefficients of the grade-groups of children is .18.

The coefficients vary, but there is a tendency for them to be positive rather than negative. Five of the six coefficients between recall of the different materials in Set *A* and in Set *B* are positive. In Set *A* the total of the positive coefficients is 2.26, the negative .04; in Set *B* the total of the positive is .78, negative .12. Twenty-three of the twenty-eight coefficients of correlation between recognition of different materials are positive; the total of the positive coefficients is 5.32 and the negative .16.

The average of the coefficients of correlation for recognition of words with other material in Set *A* is .33, in Set *B* .42; for recognition of forms with other material in Set *A* .25, in Set *B* .15; for recognition of proverbs with other material in Set *A* .39, in Set *B* .44; for recognition of syllables with other material in Set *A* .51, in Set *B* .39. The average of the coefficients for recall of words with recall of forms and recall of syllables among children is .20, for recall of forms with recall of words and recall of syllables is .17, for recall of syllables with recall of words and recall of forms is .17.

The correlation between recognition of different materials is low. The person who recognizes one material well may or may not recognize another material well. The correlation between recognition of different materials is about the same as the correlation between recall of different materials. Between recall of different materials the coefficients of correlation for the adult groups are .35 and .27 and between the recognition of different materials .37 and .28.

TABLE XVI

GROUP OF 52 ADULTS—COEFFICIENTS OF CORRELATION BETWEEN RECOGNITION OF DIFFERENT MATERIALS IN SET *B*

	Words	Forms	Proverbs	Syllables	Words
Words		—.04	.48	.54	.42
Forms	—.04		.24	.56	
Proverbs	.48	.24		.44	
Syllables	.54	.56	.44		
Words	.42				

GROUP OF 44 ADULTS—COEFFICIENTS OF CORRELATION BETWEEN RECOGNITION OF DIFFERENT MATERIALS IN SET *A*

	Words	Forms	Proverbs	Syllables	Words
Words		.32	.28	.66	.38
Forms	.32		—.12	.24	
Proverbs	.28	—.12		.28	
Syllables	.66	.24	.28		
Words	.38				

GROUP OF 638 CHILDREN—COEFFICIENTS OF CORRELATION BETWEEN RECOGNITION OF DIFFERENT MATERIALS

Grade	4*A*	4*B*	5*A*	5*B*	6*A*	6*B*	7*A*	7*B*	8*A*	8*B*	Av.
Rg. Words... / Rg. Forms...	.52	—.04	.10	.32	.30	.14	—.04	.04	.42	.22	.20
Rg. Words... / Rg. Syl......	.34	.06	—.02	.52	.04	.20	.04	.38		.28	.20
Rg. Forms... / Rg. Syl......	.66	.04	.04	—.02	.28	—.04	.12	.22		.04	.15

5. *Differences in Material*

Are some materials more easily recalled than others? Are some more easily recognized? Are those recalled well also recognized?

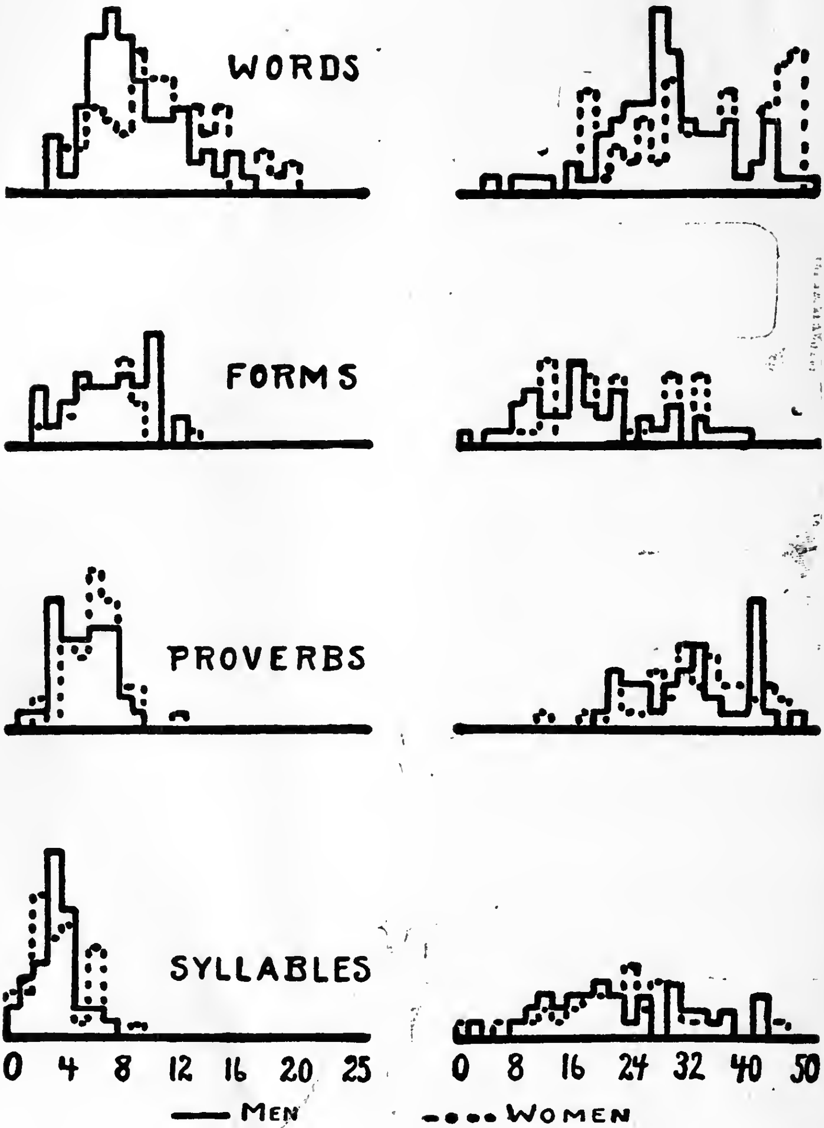
From Table XII. it is seen that the material with the best score for recall is words, then forms and proverbs, and the worst score is for syllables, for both groups of adults. For recognition the order of the scores from highest to lowest is proverbs, words, forms and syllables, and is the same for both groups. Syllables are the hardest material to recall and to recognize. Words are recalled well and also recognized well, but proverbs are the easiest recognized.

Some of the adults tested were asked to give their introspections. These show that for all the materials the subjects tried to associate the items, sometimes one item with another and sometimes with an experience outside of the presented series. Many of the subjects visualized the items or tried to remember the sound of them. Recency, primacy, familiarity, likes and dislikes all served as aids. The introspective accounts are similar for the different materials, and for recall and recognition, but different among different subjects.

The degree to which the method or means of remembering is successful depends upon the material—it is more difficult to form associations for syllables than for words. It appears that the material with the greatest number of associations is best remembered. In the tests for words one can form associations with other words or with outside experiences, visualize, or imagine the sound; the word can be remembered as a single item, not each letter separately. In some cases, the forms might be named but often by a phrase rather than a word, they could be visualized, but few associations, either between them or with outside experiences can be made, and often separate lines in the form have to be remembered instead of the form as a single item. The proverbs are familiar material, but are remembered for their "idea" more than word for word and hence are easily recognized but not so easy to recall. The syllables have the fewest associations and are most difficult to recall or recognize.

6. *Sex Differences in Achievement*

In examining sex differences every attempt was made to exclude other factors, or to keep them constant, which might influence the scores. It has been suggested that in many experiments on sex differences unequal numbers of each sex have been subjects and comparisons might be fairer if the same number of each sex were tested. Equal numbers of men and women are, therefore, in each of the adult groups compared. They were selected at random from the subjects which were graduate and undergraduate students in Columbia University. There are not equal numbers of boys and of girls in the grade-groups of children. To have taken equal numbers of boys and of girls in each grade would have reduced the number of subjects in


WORDS
FORMS
PROVERBS
SYLLABLES
0 4 8 12 16 20 25
0 8 16 24 32 40 50
MEN
WOMEN

each grade. The number of girls and of boys in each grade may be seen by referring to Table XV. on page 42.

TABLE XVII

SEX DIFFERENCE IN RECALL AND RECOGNITION

Group of 20 Women—Recognition Set A, Recall B

See chart on opposite page.

	Recall					Recognition				
	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	Words
Total ...	180	132	123	53	208	660	536	683	493	732
Av.	9.00	6.60	6.15	2.65	10.40	33.00	26.80	34.15	24.65	36.60
M.V	2.30	1.99	1.45	1.74	2.61	8.30	5.39	6.28	6.65	7.14
P.E.	.44	.37	.27	.33	.50	1.37	1.02	1.18	.26	1.35
Range ...	4–19	3–13	2–9	0–7	5–20	18–48	16–34	12–48	0–38	22–48

Group of 20 Men—Recognition Set A, Recall Set B

	Recall					Recognition				
	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	Words
Total ...	150	129	102	61	176	550	493	684	467	616
Av.	7.50	6.46	5.10	3.05	8.80	27.50	24.65	34.20	23.35	30.80
M.V. ...	2.25	2.05	1.07	1.32	2.29	4.25	7.70	6.45	6.78	6.28
P.E.	.43	.39	.32	.27	.43	.80	1.46	1.22	1.28	1.19
Range ...	3–15	2–10	1–4	0–7	5–17	10–44	6–40	22–42	8–38	16–50

TABLE XVIII

Group of 20 Women—Recall Set A, Recognition Set B

	Recall					Recognition				
	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	Words
Total ...	217	137	122	80	274	712	295	675	406	738
Av.	10.35	6.85	6.10	4.00	13.70	35.60	14.75	33.75	20.30	36.90
M.V. ...	3.33	1.95	1.34	1.20	2.81	9.74	5.05	6.35	8.08	7.01
P.E.	.64	.37	.25	.23	.53	1.84	.95	1.20	1.51	1.33
Range ...	4–18	4–12	4–9	2–9	8–20	17–48	4–24	22–44	6–44	24–48

Group of 20 Men—Recall Set A, Recognition Set B

	Recall					Recognition				
	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	Words
Total ...	158	149	98	65	204	589	258	642	340	598
Av.	7.90	7.45	4.90	3.25	10.20	29.45	12.90	32.10	17.00	29.90
M.V. ...	2.41	2.36	1.41	.83	2.94	7.50	5.10	6.10	5.60	7.94
P.E.	.46	.45	.25	.16	.56	1.42	.96	1.20	1.59	1.50
Range ...	3–14	2–12	3–7	1–6	3–16	12–48	0–26	22–48	2–32	22–40

The scores of the boys and of the girls are given in the following table.

TABLE XIX

	Recall Words						Recognition Words					
	Boys			Girls			Boys			Girls		
	Av.	M.V.	P.E.	Av.	M.V.	P.E.	Av.	M.V.	P.E.	Av.	M.V.	P.E.
4A	4.12	1.36	.18	4.30	1.41	.17	20.15	8.92	1.18	23.36	8.14	.97
4B	4.41	1.29	.20	5.33	1.73	.23	14.79	8.64	1.35	23.32	9.91	1.34
5A	5.18	1.58	.21	5.64	1.75	.22	23.74	9.17	1.24	28.19	9.31	1.15
5B	5.66	1.44	.19	5.60	1.54	.24	23.88	8.94	1.18	26.73	7.64	1.18
6A	5.69	1.82	.30	6.08	1.87	.26	29.38	9.58	1.27	28.22	9.86	1.39
6B	6.36	1.28	.23	6.49	1.39	.19	24.25	8.05	1.45	35.04	7.80	1.05
7A	6.48	2.02	.34	7.36	1.67	.42	27.12	8.46	1.43	32.73	7.25	1.87
7B	8.18	1.88	.34	7.45	2.12	.32	24.04	6.15	1.11	31.81	8.86	1.34
8A	6.86	1.56	.29	8.09	1.95	.28	29.62	7.01	1.29	28.94	9.89	1.43
8B	6.63	1.55	.35	7.76	1.81	.30	27.68	9.84	1.91	32.39	6.94	.95

	Recall Forms						Recognition Forms					
	Av.	M.V.	P.E.	Av.	M.V.	P.E.	Av.	M.V.	P.E.	Av.	M.V.	P.E.
4A	3.32	1.30	.17	4.30	1.41	.19	7.56	6.28	.83	8.16	5.98	.71
4B	3.69	1.49	.23	3.33	1.09	.15	8.10	6.59	1.03	5.36	6.20	.83
5A	4.26	1.47	.20	4.15	1.28	.16	11.15	6.02	.81	9.21	5.47	.67
5B	4.59	1.61	.21	4.00	1.30	.21	9.73	6.19	.82	8.23	4.73	.83
6A	4.42	1.46	.24	4.95	1.40	.20	11.88	4.98	.83	13.47	6.73	.95
6B	5.27	1.55	.22	5.22	1.58	.22	10.36	6.29	1.13	14.92	5.64	.78
7A	5.08	.85	.14	4.55	1.80	.46	8.36	3.43	.58	8.09	4.26	1.09
7B	6.23	1.29	.21	5.77	1.67	.25	9.77	5.16	.93	10.48	5.50	.83
8A	5.43	1.33	.24	5.38	1.57	.23	12.14	6.63	1.01	11.71	6.71	.97
8B	5.63	1.18	.23	5.24	1.50	.21	14.61	5.89	1.14	11.48	6.34	.87

	Recall Syllables						Recognition Syllables					
	Boys			Girls			Boys			Girls		
	Av.	M.V.	P.E.	Av.	M.V.	P.E.	Av.	M.V.	P.E.	Av.	M.V.	P.E.
4A	1.54	.94	.12	2.00	1.24	.15	6.95	6.10	.81	8.62	7.12	.85
4B	1.51	.71	.11	1.82	1.09	.15	2.45	6.19	.97	10.26	6.57	.89
5A	1.74	1.43	.19	1.83	1.17	.14	7.34	6.21	.84	10.68	6.29	.78
5B	2.87	1.39	.18	2.10	1.13	.17	9.98	7.39	.98	9.60	6.99	10.08
6A	1.81	.76	.17	2.41	1.34	.19	15.81	6.45	1.07	15.83	6.44	11.16
6B	1.86	.93	.17	2.21	1.13	.16	11.41	4.90	.88	13.19	6.33	1.88
7A	2.24	.60	.10	2.64	1.03	.26	11.24	6.87	1.16	12.55	6.96	1.77
7B	3.00	1.18	.23	2.55	1.34	.22	10.32	6.17	1.11	14.06	8.33	1.26
8B	2.84	1.22	.24	5.37	2.70	.37	10.79	7.04	1.36	16.71	7.12	.98

For the recall records the men's scores are 7.50, 8.80, 7.90, 10.20 and the women's 9.00, 10.40, 10.35, 13.70. The women appear to recall words better than the men. For the recognition of words the men's scores are 27.50, 30.80, 29.45, 29.90 and the women's 33.00, 36.60, 35.60, 36.90, the scores of the women are higher. In the tests for recall of forms the average scores for the men are 6.45, 7.45, and for the women 6.60, and 6.85; for recognition the men's average scores are 24.65 and 12.90 and the women's 26.80 and 14.75. For recall of forms men in one group on the average are superior and women in the other group, for recognition of forms women in both groups are superior. The men's average scores for recall of proverbs are 5.10 and 4.90 and the women's 6.15 and 6.10; for recognition of proverbs, men 34.20, 32.10, women 34.15, 33.75—a superiority for recall of proverbs among women but little difference for recognition. In the tests for recall of syllables men's average score are 1.32 and 3.25 and the women's 1.74 and 4.00; for recognition of syllables for men 23.35 and 17.00, for women 24.65 and 20.30.

Among the adults tested the women appear to be superior to the men, on the average. The tendency for the women's scores to be higher is greatest in the scores for recall and recognition of words and in the recall of proverbs. The P.E. of each average is given in the table.

In the table for children, girls are superior to boys six times out of ten for the recall of words. If one averages the averages, boys have 5.96 and the girls 6.41 or there is a difference of .45 in favor of the girls. Girls are superior to boys eight out of ten times for recognition of words, and the average of the averages for boys is 24.49 and for girls 29.07 or a difference in favor of the girls of 4.58. For recall of forms, girls are superior to boys two times out of ten and averaging the average, boys 4.79, girls 4.69, the boys are superior by .10. For recognition of forms girls are superior four times out of ten and averaging the averages, boys 10.37, girls 10.11, the boys are superior by .26. Girls are superior to boys seven times out of nine for recall of syllables and averaging the averages, boys 2.09, girls 2.55, the girls are superior by .46. In the recognition of syllables girls are superior to boys eight times out of nine and in the average of the averages, boys 9.59, girls 12.61, girls are superior by 3.02. To summarize, in the twenty-nine recall tests girls are superior seventeen times and in the twenty-nine recognition tests girls are superior twelve times. Little or no sex difference is present for recall of forms. These data show a tendency for girls to be superior to boys in recall and in recognition of words and of syllables.

Other investigators have reported superiority of girls in recall memory. Chamberlain (6), however, did not find any confirmation of the statement in his results. Any difference between the sexes which exists in the data in the preceding tables is present for recall and recognition; the difference between the sexes varies with the material. No apparent sex difference seems to be present either in achievement in recall or in achievement in recognition.

Some investigators who find women superior to men offer as an explanation, that women remember more details. The following experiment was performed on twenty men and twenty women each separately to see if there was any sex difference in the number of details remembered.

To each subject a colored picture seven and three eighth inches by eight and a half inches (Kaffeebesuch, P. Philippi) was shown with the following printed instructions:

"You will be shown a picture for ten seconds. Look at it so that you will be able to describe it afterwards."

After the subject had seen the picture for ten seconds, he was given thirty-six cards three by five inches on each of which was pasted a piece of a picture. Eighteen of the pieces were from the picture previously presented, almost all of the picture being included among the eighteen pieces. The other eighteen pieces were from another picture but one with somewhat similar setting. The subjects were asked to place each card in one of four piles—(1) Sure you have seen, (2) fairly sure you have seen, (3) fairly sure you have not seen, (4) sure you have not seen. In piles 1 and 2 are Yes answers and in 3 and 4 are No answers. The number of right Yeses among the women averages 11.35 and among the men 10.00. The women remembered on the average more than the men. The number of right Nos by women averages 14.80 and for men 14.60. There is practically no difference between the sexes in recognizing that a certain item has not been seen before, but the women are a little superior to the men in recognizing a thing which has been seen before.

TABLE XX

	Right Yeses			Right Nos		
	Average	M.V.	P.E.	Average	M.V.	P.E.
Woman	11.35	2.14	.40	14.80	2.06	.39
Men	10.00	2.70	.51	14.60	2.32	.49

The subjects were also shown fifteen pictures taken from the *National Geographic Magazine*. They were about eight by five inches, each pasted on paper ten by eight inches. They were black and white or brown and white. The pictures were selected on the basis of there being another picture available similar to it. The original fifteen pictures were shown successively, one each two seconds. They were mixed with the fifteen other pictures and the subject asked to arrange them in four piles as described above. To be able to distinguish which of the two pictures was the one seen one would often have to remember details of the picture. For example, two pictures were of monuments. It was not enough to remember "monument." There were two interiors of churches, two walls with Egyptian writing, two windmills, two bridges, two groups of people, etc. The average score of right responses in piles 1 and 2 is 12.30 for women and 12.60 for men; the range for women being 8 through 14 and for men 10 through 15. The average number of wrong responses in these piles for women is 1.35 and for men .70. The subjects were asked to introspect, telling why they choose each picture. Both the men and the women mentioned details. The following are a few of the ones given: (men) "Trend of the grass," "Little girl standing by river," "Side view of this bridge different," "Cloud effects," "Flag near top of this monument," "Light and dark trees"; (women) "Dark trees," "Grass, twigs," "Figures on base," "Sky," "Long aisle," "Sand dunes," "Telegraph pole in the back." The men and the women recognize about the same number of pictures correctly; the women incorrectly select more than the men, this point will be discussed in Chapter VII. The data do not suggest that women remember more details.

Women appear to remember more items than do the men in the memory tests, but there seems no evidence that this is because they remember more *details.*

7. *Sex Difference in Variability*

Are there any sex differences in variability among the adults or among the children in the recall of different materials? in the recognition of different materials?

In Table XXI. the average, the probable error of the average (P.E.), the Pearson Coefficient of Variability (P.C.) and the range are given for the two groups of men and for the two groups of women. The measure of variability known as the Pearson coefficient is the gross variability divided by the average.

Among the twenty Pearson coefficients seven are larger for the women. Eight of the twenty P.E. measures are higher for the women, two are equal for both sexes and ten are higher for men. The range indicates the lowest and the highest scores in the group. Among the twenty scores representing the poorest scores in the group for each test, two are lower for women than for men and three are the same for both men and women. Eleven of the twenty best scores in each group for each test are higher for women, in five cases the highest scores are the same for men and for women, and in four cases the men's best scores are higher than the women's best scores.

No sex difference in variability is apparent among the adult

TABLE XXI

Group of 20 Women—Recognition Set *A*, Recall Set *B*

	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	W
Av.	9.00	6.60	6.15	2.65	1.04	33.00	26.80	34.15	24.65	36.6
P.E.	.44	.37	.27	.33	.50	1.57	1.02	1.18	1.26	1.3
P.C.	.2556	.3015	.2358	.6566	.2510	.2515	.2011	.1839	.2698	.1
Range ...	4–19	3–13	2–9	0–7	5–20	18–48	16–34	12–48	0–38	22

Group of 20 Men—Recognition Set *A*, Recall Set *B*

	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	W
Av.	7.50	6.45	6.10	3.05	8.80	27.50	24.65	34.20	23.35	30.8
P.E.	.43	.39	.37	.27	.43	.80	1.46	1.22	1.28	1.1
P.C.	.3000	.3178	.3333	.4328	.2602	.1545	.3119	.1889	.2903	.2
Range ...	3–51	2–10	1–9	0–7	5–17	10–44	6–40	22–42	8–38	16

Group of 20 Women—Recall Set *A*, Recognition Set *B*

	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	W
Av.	10.35	6.85	6.10	4.00	13.70	35.60	14.75	33.75	20.30	36.9
P.E.	.64	.37	.25	.23	.53	1.84	.95	1.20	1.51	1.3
P.C.	.3217	.2847	.2197	.3000	.2051	.2736	.3424	.1882	.3980	.3
Range ...	4–18	4–12	4–9	2–9	8–20	17–48	4–24	22–44	6–44	24

Group of 20 Men—Recall Set *A*, Recognition Set *B*

	Words	Forms	Proverbs	Syllables	Words	Words	Forms	Proverbs	Syllables	W
Av.	7.90	7.45	4.90	3.25	10.20	29.45	12.90	32.10	17.00	29.9
P.E.	.46	.45	.25	.16	.56	1.42	.96	1.20	1.59	1.5
P.C.	.3051	.3168	.2878	.2554	.2882	.2547	.3954	.1900	.3296	.2
Range ...	3–14	2–12	3–7	1–6	3–16	12–48	0–26	22–48	2–32	22

groups when the Pearson coefficient or the P.E. or the range are used as measures of variability.

In the tests for recall of words among the children, the Pearson coefficient of variability is greater for the girls in six of the ten cases; for recall of forms six of the ten coefficients are greater for girls; for recall of syllables seven of the ten coefficients are greater for girls. In the recognition of words the Pearson coefficient of variability for girls is greater than that of the boys' scores two of the ten times; in recognition of forms five of the ten times; in recognition of syllables once in the nine times. Of the fifty-eight coefficients of variability, twenty-seven show greater variability for boys.

Among the P.E. measures the girls' are higher than the boys' four of the ten times in the test for recall of words, three times out of ten in test for recall of forms and twice the scores are the same, eight times out of nine for the test in recall of syllables the girls' scores are higher; four of the ten times in recognition of words the girls are higher, and once the same for both, in test for recognition of forms three times out of the ten and in test for recognition of

TABLE XXII

CHILDREN

		Recall Words			Recall Forms			Recall Syllables		
		Av.	P.E.	P.C.	Av.	P.E.	P.C.	Av.	P.E.	P.C.
4A	B.	4.12	.18	.3301	3.32	.17	.3916	1.54	.12	.6101
	G.	4.30	.17	.3279	4.30	.19	.3279	2.00	.15	.6200
4B	B.	4.41	.20	.2925	3.69	.23	.3225	1.51	.11	.4702
	G.	5.33	.23	.3246	3.33	.15	.3273	1.82	.15	.5789
5A	B.	5.18	.21	.3050	4.26	.20	.3451	1.74	.19	.8218
	G.	5.64	.22	.3103	4.15	.16	.3084	1.83	.14	.6393
5B	B.	5.66	.19	.2544	4.59	.21	.3508	2.27	.18	.6123
	G.	5.60	.24	.2750	4.00	.21	.3250	2.10	.17	.5650
6A	B.	5.69	.30	.3198	4.42	.24	.3303	1.81	.17	.4144
	G.	6.08	.26	.3076	4.95	.20	.2869	2.41	.19	.5560
6B	B.	6.36	.23	.2013	5.27	.22	.2941	1.86	.17	.5000
	G.	6.49	.19	.2182	5.22	.22	.3027	2.21	.16	.5112
7A	B.	6.48	.34	.3114	5.08	.14	.1673	2.24	.10	.2679
	G.	7.36	.42	.2269	4.55	.46	.3956	2.64	.26	.3902
7B	B.	8.18	.34	.2298	6.23	.21	.2071	3.00	.23	.3933
	G.	7.45	.32	.2847	5.77	.25	.2894	2.55	.22	.5255
8A	B.	6.78	.29	.2274	5.43	.24	.2431			
	G.	8.09	.28	.2410	5.38	.23	.2918			
8B	B.	6.63	.35	.2338	5.63	.23	.2096	2.84	.24	.4331
	G.	7.76	.30	.2332	5.24	.21	.2863	5.37	.37	.5028

		Recognition Words			Recognition Forms			Recognition Syllables		
		Av.	P.E.	P.C.	Av.	P.E.	P.C.	Av.	P.E.	P.C.
4A	B.	20.15	1.18	.4427	7.56	.83	.8307	6.95	.81	.8777
	G.	23.36	.97	.3485	8.16	.71	.7328	8.62	.85	.8260
4G	B.	14.79	1.35	.5842	8.10	1.03	.8136	2.45	.97	2.5266
	G.	23.32	1.34	.4250	5.36	.83	1.1567	10.26	.89	.6404
5A	B.	23.74	1.24	.3863	11.15	.81	.5399	7.34	.84	.8465
	G.	28.19	1.15	.3303	9.21	.67	.5094	10.68	.78	.6403
5B	B.	23.88	1.18	.3744	9.73	.82	.6362	9.98	.98	.7405
	G.	26.73	1.18	.2858	8.23	.83	.5747	9.60	1.08	.7281
6A	B.	29.38	1.27	.3261	11.88	.83	.4192	15.81	1.07	.4080
	G.	28.22	1.39	.3250	13.47	.95	.4996	15.83	1.16	.4068
6B	B.	24.45	1.45	.3292	10.36	1.13	.6071	11.41	.88	.4294
	G.	35.04	1.05	.2226	14.92	.78	.3807	13.19	.88	.4784
7A	B.	27.12	1.43	.3119	8.36	.58	.4103	11.24	1.16	.6112
	G.	32.73	1.87	.2215	8.09	1.09	.5266	12.55	1.77	.5546
7B	B.	24.04	1.11	.2558	9.77	.93	.5281	10.32	1.11	.5978
	G.	31.81	1.34	.2785	10.48	.83	.5248	14.06	1.26	.5925
8A	B.	29.62	1.29	.2366	12.14	1.01	.5461			
	G.	28.94	1.43	.3417	11.71	.97	.5730			
8B	B.	27.68	1.91	.3555	14.61	1.14	.4031	10.79	1.36	.6525
	G.	32.39	.95	.2146	11.48	.87	.5523	16.71	.98	.4261

syllables five of the nine times the girls are higher and once they are the same for both sexes.

Of the fifty-eight measures of the P.E. twenty-seven are greater for the girls' scores, four are equal for both boys and girls and twenty-seven are greater for the boys' scores.

No difference in variability between the boys and girls is apparent when the Pearson coefficient or the P.E. are used as measures of variability.

No sex difference in variability in these tests for recall and for recognition is present among adult nor among children.

8. *Age Differences in Recall and Recognition*

Is there any difference in recall among children of different ages regardless of class in school?—in recognition?

In Table XXIII. are the averages for each material, words, forms, and syllables, by each method of testing, recall and recognition, for each age. The number of children in each age-group is also given.

Each child was asked to state how old he or she was on the last birthday and when this birthday was. Since all children eight on their last birthday, nine on their last birthday, etc., are here considered "eight-year-olds," "nine-year-olds," etc., the meaning of eight-year-olds, nine-year-olds, is 8.5, 9.5, etc., in this report. Furthermore, all eight-year-olds, nine-year-olds have not been tested, so that the averages are for the eight-year-olds, nine-year-olds, etc., who were tested in the grades *4A*, *4B*, *5A*, *5B*, *6A*, *6B*, *7A*, *7B*, *8A*, *8B*. There is a marked selection in the case of eight-year-olds, sixteen-year-olds, and seventeen-year-olds; the results for them are therefore given in parenthesis and are not included in the diagram. Only fif-

TABLE XXIII

AGE DIFFERENCES

No. of Children	Age	Recall Words	Recall Forms	Recall Syl.	Recog. Words	Recog. Forms	Recog. Syl.
15	8.5	(4.60)	(3.27)	(1.00)	(18.13)	(5.40)	(9.80)
82	9.5	4.70	3.36	1.68	21.49	8.70	7.49
120	10.5	5.41	4.10	1.86	24.49	9.65	9.56
130	11.5	5.56	4.54	2.03	25.87	11.08	11.17
110	12.5	5.85	4.64	2.28	26.08	9.55	11.17
91	13.5	7.16	5.33	2.38	29.54	9.56	11.77
62	14.5	7.19	5.51	2.86	31.24	12.58	13.23
27	15.5	8.33	5.41	1.91	35.39	10.26	14.32
9	16.5	(6.11)	(5.11)	(1.43)	(24.55)	(9.11)	(6.67)
2	17.5	(10.00)	(6.00)	(5.00)	(39.00)	(11.00)	(15.00)

teen of the one hundred and fifty-nine in the fourth year (4*A* and 4*B*) are eight years old, most of the children of that grade being at least nine years old. Due to the laws permitting children of fourteen who have reached a certain grade to obtain "working papers," to the fact that most children of sixteen and seventeen have completed grade 8*B*, and to other causes, only eleven children of sixteen and seventeen are among the children in the seventh and eighth grades (7*A*, 7*B*, 8*A*, 8*B*).

The scores for recall and for recognition of words increase gradually from the age of 8.5 through 15.5. In general, there is a tendency for improvement with age for recall and for recognition of forms and syllables also. The improvement with age may be seen by comparing the average of the averages for ages 9.5, 10.5, 11.5 with the average of the averages for ages 13.5, 14.5, 15.5. The average of the average scores for ages 9.5, 10.5, 11.5 for recall of words is 5.22 and the average of the average scores for ages 13.5, 14.5, 15.5 for recall of words is 7.22. For recognition of words the first average is 23.95 and the second 28.95. For the recall of forms the scores are 3.97 and 5.42 respectively, and for recognition of forms 9.61 and 10.8. In the tests for syllables recall averages 1.86 and 2.38 respectively and for recognition 9.41 and 13.11 respectively. For all materials and for recall and recognition the scores are higher for the older ages than for the younger ones. Both recall and recognition scores seem to improve with age.

9. *Grade Differences in Recall and Recognition*

Is there any difference in recall among children of different grade regardless of age?—in recognition?

In Table XXIV. the averages are given for each material in each grade.

TABLE XXIV

	Recall Words		Recall Forms		Recall Syllables	
	Av.	M.V.	Av.	M.V.	Av.	M.V.
4*A*	4.22	1.44	3.40	1.32	1.79	1.12
4*B*	4.94	1.60	3.48	1.30	1.69	.89
5*A*	5.43	1.67	4.15	1.28	1.79	1.28
5*B*	5.63	1.49	4.33	1.53	2.20	1.28
6*A*	5.92	1.87	4.73	1.46	2.16	1.23
6*B*	6.44	1.35	5.24	1.55	2.08	1.03
7*A*	7.03	1.92	4.97	1.69	2.36	1.19
7*B*	7.75	2.00	5.96	1.48	2.74	1.30
8*A*	7.62	1.93	5.40	1.48		
8*B*	7.39	1.77	5.37	1.41	3.07	1.50

TABLE XXIV.—*Continued.*

	Recognition Words		Recognition Forms		Recognition Syllables	
	Av.	M.V.	Av.	M.V.	Av.	M.V.
4A	21.91	8.76	7.98	6.09	7.87	6.01
4B	21.09	9.65	7.72	6.07	6.87	6.49
5A	26.19	9.90	10.09	5.68	10.51	6.37
5B	25.08	7.94	9.10	5.42	9.81	7.13
6A	28.71	10.17	12.77	6.05	15.82	6.41
6B	31.10	7.80	13.22	6.02	12.86	5.76
7A	28.83	8.22	8.28	5.68	11.64	6.91
7B	28.59	9.03	10.19	5.50	12.51	7.65
8A	29.20	8.37	11.87	6.61		
8B	30.82	8.48	11.48	6.34	14.74	7.41

There is a tendency for the scores to improve gradually from grade *4A* through *8B* for recall and for recognition of words, forms, and syllables. This tendency is more evident when the average scores for *4A, 4B, 5A, 5B, 6A* are compared with the average of the scores for *6B, 7A, 7B, 8A, 8B.* For recall of words these scores are 5.23 and 7.25 respectively, for recognition of words 24.59 and 29.71 respectively, for recall of forms 9.53 and 11.01 respectively, for recall of syllables 1.93 and 2.05 respectively, for recognition of syllables 10.18 and 12.94 respectively.

The tables also show the general tendency of improvement from *4A* through *8B* in both recall and recognition for words, forms, and syllables.

10. *Age and Grade Differences*

Since the tables and curves for grade tend to show an improvement with advance in grade and the tables and curves for age tend to show an improvement with age, it is possible that both grade and age have their influence. This leads to the question: Do the youngest children in each grade have the best or worst scores? Do the oldest children in each grade do best or worst in memory tests? Do the children whose age is the same as most of the other children in their grade have the best, medium, or worst scores?

The following method of scoring was adopted. One step in age is one year, 8 to 9, 9 to 10, etc., and for grade the step used is one year, 4 to 5 (*4A, 4B*), 5 to 6 (*5A, 5B*), 6 to 7 (*6A, 6B*), 7 to 8 (*7A, 7B*), 8 to high school (*8A, 8B*). The following table (XXV.) shows the number of children of each age in each grade.

The medium for grade 4 is at age 9.5, for grade 5 at 11.5, for grade 6 at 11.5, for grade 7 at 13.5, for grade 8 at 14.5. By adding

TABLE XXV

Age	4	5	6	7	8	Total
8.5	15					15
9.5	*68*	14				82
10.5	52	55	12	1		120
11.5	16	*58*	*50*	6		130
12.5	5	23	33	25	14	100
13.5	3	4	18	*32*	34	91
14.5		1	6	16	*39*	62
15.5		2	2	8	15	27
16.5				1	8	9
17.5					2	2
	159	157	121	89	112	638

the scores for the recall of words of the nine year olds for grade 4, of the eleven year olds for grade 5, of the eleven year olds for grade 6, of the thirteen year olds for grade 7, and of the fourteen year olds for grade 8, the total score of those who are in a grade where the median age is their age, was obtained. By dividing this total by the total number of persons whose scores have been added, the average score for the children who are in a grade where the median age is their age is found. Then the scores of those who are one year, two years, etc., older and one year, two years, etc., younger than the median age for their grade may be computed. Thus the tabulation for the recall of words is:

Scores

−3	−2	−1	0	1	2	3	4
		69	320	225	74	23	9
	65	324	304	132	22	7	13
		79	305	194	115	39	15
11	40	181	252	115	60	8	
	114	257	285	127	47	20	
11	219	910	1466	793	318	97	37

Subjects

−3	−2	−1	0	1	2	3	4
		15	68	52	16	5	3
	14	55	58	23	4	1	2
		12	50	33	18	6	2
1	6	25	32	16	6	1	
	114	34	39	15	8	2	
1	34	141	247	139	54	15	7
Av.. 11.00	6.44	5.45	5.93	5.70	5.88	5.47	5.30

The averages for the six tests are given in Table XXVI.

TABLE XXVI

	−3	−2	−1	0	1	2	3	4
Recall Words		6.44	6.45	5.93	5.70	6.47	6.47	5.30
Recall Forms	6.00	5.36	4.87	4.51	4.45	4.60	4.60	3.57
Recall Syl.	0.00	2.65	2.49	2.26	2.19	2.87	2.87	2.43
Recog. Words	44.00	25.38	26.71	26.55	24.76	29.13	29.13	31.85
Recog. Forms	16.00	10.50	10.58	10.53	8.61	11.67	11.67	7.71
Recog. Syl.	12.00	13.74	12.66	10.73	10.17	9.69	6.07	9.57

The score under zero is that of the children who are in a grade where the median age is their age, under —1 is the score of the children who are one year younger than the median, under 1 is that of the children who are one year older than the median for their grade, etc.

It is evident that the oldest children in a grade, on the average, do not have the best scores, but they do not always have the worst scores. The children at the median age for their grade do not have the best nor the worst scores. There is a tendency for the scores of those who are younger than their classmates to be higher, except in the case of syllables, but the number of subjects is small in the group under —3. The curves and the tables suggest that there is a tendency toward improvement by age and by grade.

11. *Group Differences in Recall and Recognition*

TABLE XXVII

	Recall					*Recognition*				
	Words	Forms	Prov.	Syl.	Wds.	Wds.	Forms	Prov.	Syl.	W
Grad. Stud. Av.	8.53	7.00	5.13	3.33	10.40	31.60	18.55	32.58	11.90	33.
Undergrad. Av.	9.10	6.63	6.00	3.13	11.15	31.18	21.00	34.53	21.55	33.

In Table XXVII. the results of the graduate and undergraduate students are given. Graduate students are a selected group of college graduates, but no marked difference in achievement is noticed in any of the memory tests between them and the undergraduate students. There is a marked difference between the college and university students and the school children. The children's group is not so selected as the adults' group.

The differences are present or absent in the same way for recall and recognition.

CHAPTER VI

Memory Tests of Insane Patients

Books on psychiatry often mention memory defects among insane patients. Little is said whether the defect is for both recall and recognition.

To attempt to see if any differences in recall and in recognition could be observed among insane patients, the tests used on the normal subjects were given to several patients in a hospital for the insane. The cases will first be considered separately.

Several patients known as Korsakoff's were tested.

M–1 male, adult, exact age unknown. Korsakoff Disease.

Recall scores: Words 12, Forms 6, Proverbs 2, Syllables 0, Words 1.

Recognition scores: Words 12, Forms 6, Proverbs 14, Syllables 6, Words 6.

The patient was born in this country and formerly had been in business. He was pleasant and coöperated well in the tests, for he was anxious to leave the hospital and go to work. The picture had been shown to him for fifty seconds before the tests were made so that he might have some idea of the length of the period. After the ten memory tests were made, the examiner asked him if he remembred her showing him a picture. He did not, but when shown the picture again he said he remembered having seen it when he first came into the room.

The examiner showed him six cards (5 by 8 inches) on each of which was pasted an object (big blue *B*, a red postage stamp, a dish of cereal, man's head, baby and dog, watch). Each card was shown for two seconds. Immediately afterwards they were mixed with six more cards, each having a picture. He was asked to select all which he had seen. He recognized the first, second, third, and sixth, a total of four. He asked if they were right.

M–2 male, age 62.

Recall scores: Words 3, Forms 1, Proverbs 1, Syllables 1, Words 3.

Recognition scores: "I don't remember ever having seen them."

This patient when examined (1915) had been in the hospital for seven years but did not know how long. The records show that in 1913 he knew he had come in 1908. The patient in his youth, had attended a well known American university; later he had been a ticket seller. The hospital records call his case "residual of alcoholic Korsarkoff psychosis."

After studying the presented series for fifty seconds he could not recognize anything—this was true for all the materials. He would say, "I don't remember ever having seen any. Isn't that extraordinary?" In the test for recall he remembered three words, one form, one proverb, one syllable, three words.

He was shown the six cards mentioned above, one every two seconds. They were mixed with six other cards on each of which was pasted the picture of objects and from the twelve he was asked to select those that he had seen. He said that he did not remember having seen any. This test was extremely easy as the six objects seen, including a red postage stamp and a large blue *B*, could easily be distinguished by most persons.

F–1 Female, age 66, religion: Roman Catholic. Occupation: cook. Born in the U. S.; single. Korsakoff disease.

Recall scores: Words 4, Forms 0, Proverbs 2, Syllables 1, Words 3.

Recognition scores: Words 16, Forms 6, Proverbs 16, Syllables 2, Words 14.

The patient had been addicted to use of alcohol for a great many years and used it to excess between 55 to 65 years of age. She fabricates and has very defective judgment, according to other examiners. Physically some evidence of an arterio-sclerotic condition had been found.

She could write her name, but did it very poorly. The material had to be presented auditorially. The six cards mentioned above were shown and the first, second, fifth, and sixth correctly selected from the twelve afterwards, *i.e.*, four out of six were recognized.

F–2 Female, age 36 or 39. Born in Ireland; in U. S. about 14 years; religion: Roman Catholic. Occupation of husband: general laborer. Korsakoff disease.

Recall scores: Words 2, Forms 0, Syllables 0, Proverbs 2, Words 1.

Recognition scores: Words 6, Forms 6, Syllables 0, Proverbs 0, Words 0.

The patient probably had very little schooling; she can not write her own name. She has indulged excessively in alcoholics and has physically signs of neuritis.

The patient could read, but wrote poorly and slowly. The experiment was therefore conducted by letting the patient study the presented lists, but recall orally. She recalled two words from the first series, the last two on the list. She attempted to draw the forms but only made a mark resembling the letter H. She understood the proverbs and recalled two. In trying to recall the syllables she mentioned several combinations of letters, but none was correct. After the second series of words, she mentioned five words, but only one from the second list, one from the first list and three which had appeared on no list so far shown; thus her response was: carriage, stove, white, pipe, mirror. For the recognition tests the patient saw the material and answered whether or not she had seen it on the other list, the examiner recording the reply.

The patient coöperated well. She was sad, however, and moaned. Before the test materials were presented she was shown a picture for fifty seconds. Forty minutes later she was asked if she had been shown a picture. When presented again, she said she had not seen it before. She pointed out several objects on the picture indicating that she could see them.

F–3 Female, age 38. Born in Ireland; in U. S. about 16 years. Occupation of husband: coachman. Korsakoff.

Recall scores: Words 5, Forms 0, Proverbs 0, Syllables 0, Words 3.

Recognition scores: "No."

During the examination she remarked several times that she had never had any schooling because she had to work out, adding, "It's a sad thing indeed to have no schooling." She cried frequently because she was so happy, saying, "I thought I was in the bad house but I came to myself to-day and know it's not the bad house." She could neither read nor write so that the material was presented auditorially. She recalled three words. Although she could not write she drew two figures, but they were not like those shown. The proverbs seemed familiar, she often would finish one in chorus with the examiner. She could recall only two. The syllables could not be presented in a satisfactory manner. She recalled three words from the second set. In the recognition tests she would say "Yes" or "I'm not sure Ma'am."

M–3 Male, age 55. Born in Germany, in U. S. 20 years. Occupation: baker. Alcoholic psychosis.

Recall scores: Words 4, Forms 0, Proverbs 0, Syllables 0, Words 3.

Recognition scores: Words 22, Forms 0, Proverbs 0, Words 15.

The patient could not remember the doctor's name ten minutes after having said it three times. He said the forms were French.

M–4 Male, general paralysis.

Recall scores: Words 1, Forms 2, Proverbs 0, Words 1.

Recognition scores: Words 2, Forms 2, Proverbs 2, Words 2.
The patient could not do the syllable test.

M–5 Male. General paralysis with slight intellectual defect.
Recall scores: Words 1, Forms 4, Proverbs 1, Syllables 0, Words 4.
Recognition scores: Words 8, Forms 6, Proverbs 26, Syllables 8, Words 20.

The patient coöperated *very* well. He was pleasant, willing and tried to do his best.

Thirty minutes after having seen the picture which had been shown for fifty seconds, he remembered having seen it and named the objects in it. After having seen the six cards mentioned above he was able to select five from the series of twelve.

M–6 Male. General paralysis.
Recall scores: Words 6, Forms 0, Proverbs 0, Syllables 0, Words 5.
Recognition scores: Words 6, Forms 8, Proverbs 12, Syllables 8, Words 8.

He selected the first, second, third, fifth, and sixth items—*i.e.*, five of the six presented on the cards when they were mixed with the six others.

Thirty minutes after having seen the picture shown for fifty seconds he remembered and described it.

M–7 Male.
Recall scores: Words 6, Forms 4, Proverbs 1, Syllables 1, Words 4.
Recognition scores: Words 12, Forms 8, Proverbs 0, Syllables 0, Words 10.

He described the picture thirty minutes after having seen it for fifty seconds. He also selected six cards correctly from the set of twelve. His coöperation was good.

M–8 Male, brain syphilis.
Recall scores: Words 5, Forms 3, Proverbs 2, Syllables 2, Words 7.
Recognition scores: Words 6, Forms 6, Proverbs 10, Syllables 4, Words 10.

The patient coöperated well.

M–9 Male, age 60. Arterio-sclerosis. Occupation: singer.
Recall scores: Words 6, Forms 1, Proverbs 3, Syllables 2, Words 6.
Recognition scores: Words 16, Forms 6, Proverbs 32, Syllables 24, Words 32.

The patient was pleasant and tried to do his best. He remembered having seen the picture thirty minutes afterwards. He thought it was a picture of a little girl but could not say what was beside her. He recognized all six of the six cards in the series of twelve.

M–10 Male, age 63. Arterio-sclerosis. Born in U. S. Common school education; left school at age of 14. Occupation had been machinist, grocer, and salesman.

Recall scores: Words 3, Forms 0, Proverbs 4, Syllables 0, Words 10.

Recognition scores: Words 16, Forms 8, Proverbs, Syllables 0, Words 10.

He remembered the picture and could describe it twenty-five minutes after having seen it. He recognized five of the six cards correctly in the series of twelve.

M–11 Male, age 53. Arterio-sclerosis. Born in U. S.; high school education. Occupation: advertising agent, said to have earned $5,000 a year as secretary to a publishing house at one time.

Recall scores: Words 4, Forms 2, Proverbs 0, Syllables 0, Words 2.

Recognition scores: Words 10, Forms 4, Proverbs 6, Syllables 0, Words 12.

He remembered and could describe the picture thirty minutes after having seen it for forty seconds. He also selected the six cards correctly from the series of twelve.

M–12 Male, age 56. Arterio-sclerosis. Common school education. Occupation: carpenter.

Recall scores: Words 3, Forms 0, Proverbs 0, Words 1.

Recognition scores: Words 23, Forms 4, Proverbs 0, Words 4.

The materials, except forms, were presented verbally. The tests for syllables were not given.

He could select the six cards correctly from the series of twelve.

A few cases of senile dementia are described below:

F–5 Female, age 79 or 84 (she does not know). Senile. Born in Ireland; in U. S. 60 years.

The material was presented auditorially. She could not recall any of the material nor could she recognize any, as may be seen by the scores for recognition—Words 6, Forms 2, Proverbs 0, Words 6.

She could remember nothing in the picture a half hour after it had been shown to her. After seeing the six cards mentioned above she selected ten among the twelve as those previously seen.

F–6 Female, age 81. Occupation: housework. Religion: Methodist. A typical "dear old lady."

She could not see well enough to read the print nor hear. The examiner was unable to conduct the experiment, except the six cards were shown and afterwards she selected four from the series of twelve.

A few other cases not easily diagnosed are given:

M–13 Male, age 54.

Recall scores: Words 3, Forms 0, Proverbs 0, Syllables 1, Words 2.

Recognition scores: Words 10, Forms 0, Proverbs 6, Syllables 6, Words 16.

When attempting to recall words the patient mentioned several that had not been presented.

M–14 Male, age 54.

Recall scores: Words 2, Forms 0, Proverbs 0, Words 1.

Recognition scores: Words 12, Forms 4, Proverbs 12, Words 14.

The data of all the cases are not comparable for the conditions of the experiment were not the same for each patient. Some of the patients were unable to see well and for them it was necessary to vary the conditions and present the material auditorially. When it was possible the conditions were kept like those for the normal subjects to whom the materials were always presented visually. The patients differ in age, in amount of education which they have had, in previous experience and environmental conditions. They all have not the same mental disorder, nor are those whose disorders are similar all in the same stage of the disease. M–1, for example, has been diagnosed alcoholic psychosis-Korsakoff disease. He was greatly improved and allowed to leave the hospital shortly after these memory tests had been made. F–2 and F–3 were in more advanced stages of the Korsakoff disease and in each case the prognosis was poor.

No attempt will be made, therefore, to average the results of the subjects in this group of insane patients, but from the following table one can compare the results of those having Korsakoff disease, general paralysis, and arterio-sclerosis. All show a memory defect and the defect is present in both recall and recognition.

In general, from these data, there is little difference in recall among the patients suffering from general paralysis and arterio-sclerosis. In recognition, there is no difference except in the case of words for which the arterio-sclerosis patients score higher. The scores of the Korsakoffs are increased when those of patient M–1 are included. Omitting the records of M–1 who was recovering and about to leave the hospital, one finds the scores among the Korsakoffs lower than those among the general paralysis and arterio-sclerosis cases. These patients were less able to attempt the tasks. There is no striking difference between the way the diseases affect the recall and recognition.

TABLE XXVIII

Korsakoff

	M-1	M-2	M-3	F-1	F-2	F-3	F-4
Recall							
Words	12	2	4	4	2	5	3
Forms	6	1	0		0	0	0
Proverbs	2	1	1	2	2	0	
Syllables	0	1	0	1	0	0	3
Words	1	3	3	3	1		
Recognition							
Words	12		22	16	6		
Forms	6		0		6		
Proverbs	14		0	16	0		
Syllables	6			2	0		
Words	10		15	14	0		
Recog. of 6 out of 12 pictures	4	0	0	4			

	General Paralysis				Brain Syphilis	Arterio-Sclerosis			
	M-4	M-5	M-6	M-7	M-8	M-9	M-10	M-11	M-12
Recall									
	1	1	6	6	5	6	3	4	3
	2	4	0	4	3	1	0	2	0
	0	1	0	1	2	3	4	4	0
			0	1	2	2	0	0	
	1	2	5	4	7	6	5	2	1
Recognition									
	2	8	6	12	6	16	16	10	23
	2	6	8	8	6	6	8	4	4
	2	26	12	0	10	32	8	6	0
		8	8	0	4	24	0	0	
	2	20	8	10	10	32	10	12	4
Recog. of 6 out of 12 pictures		5	5	6		6	5	6	6

CHAPTER VII

RECOGNITION

INTEREST in the process of recognition has been shown recently. In the present study of recall and recognition the data on recognition may be examined with interest apart from its relation to recall. Each subject was shown twenty-five items and in the recognition test was given fifty and asked to indicate which he thought had been shown before by writing "YES" before them and before those which he thought had not been shown, "No." Twenty-five of the fifty items had been on the list previously seen and twenty-five had not been seen before.

An examination of the data will show

1. The comparison between the number of the YESES and NOS used for each material.

2. The per cent. of wrong YESES and right YESES among all the YESES used; the per cent. of wrong NOS and right NOS among all the NOS used.

3. The comparison of the per cent. of wrong YESES among all responses with the per cent. of wrong NOS to all responses.

4. The per cent. of all the responses correct for each material.

5. The comparison of the per cent. of wrong YESES among the men and among the women; the comparison of the per cent. of wrong NOS among the men and women.

6. The comparison between men and women of the number of wrong YESES among the total number of YESES used, the comparison between men and women of the wrong NOS among the total number of NOS used.

7. The comparison of *old* and *new* judged correctly.

1. *The Comparison of the Number of Yeses and Nos Used for Each Material*

More NOS are used than YESES in all of the tests. The difference is less for words than for the other materials. There is no consistency between the two sets of the same material, however. Proverbs have differences between YESES and NOS used of 399 and 385 in the first and second sets respectively; syllables have differences 490 and 299 in the first and second sets respectively when the total number of responses is 2,000.

2. *The Per Cent. of Wrong Yeses and Right Yeses Among All the Yeses Used; the Per Cent. of Wrong Nos and Right Nos Among All the Nos Used*

A larger number of Nos are wrong than Yeses. This might be expected since more Nos are used than Yeses. This is true for each of the ten tests. The amount of difference varies in the first group from 130 to 230 and in the other from 86 to 177. The differences are not consistently large or small for any material. The greatest difference, 230, between the wrong Nos and wrong Yeses in the first group and the least difference, 86, in the second group are in the tests for syllables.

The per cent. of wrong Yeses among all Yeses used and the per cent. of wrong Nos among all Nos used are given in the table. In each of the ten cases the per cent. of wrong Yeses is less than the per cent. of wrong Nos. The difference is not great; it varies in the second group from .5 per cent. to 15 per cent. and in the first group from 4.5 per cent. to 15 per cent. The difference in the first group for words is 8.5 per cent., for forms 4.5 per cent., for proverbs 15.5 per cent. and for the second series of words 11 per cent. In the second group the difference for words is 4.5 per cent., for forms 9 per cent., for proverbs 15 per cent., for syllables .5 per cent. and for the second series of words 9 per cent. The difference is greatest for proverbs in each case. In the first group the per cent. of wrong Yeses is greatest for proverbs, next for forms, then for words, and least for syllables. In the second group the order is proverbs, words, syllables, forms. In an earlier chapter it was noticed that the difficulty of the two series of forms is not the same for recognition. If the forms are omitted from our lists here, and only the materials involving letters or words considered, the two lists have the same order—proverbs, words, syllables. Thus, from the data of this experiment it appears that among materials containing letters or words, the greatest per cent. of false recognition is found in that material where the greatest meaning is present. The per cent. of false recognitions decreases when the meaningful associations in the material decrease.

The range of the per cents of wrong Nos used in the first group is from 22 per cent to 28 per cent. and in the second group from 21.5 per cent. to 37.5 per cent. The range is smaller and therefore the difference between the materials is less for wrong Nos than for wrong

Yeses. Forms in both groups lead the list in having the highest per cent. of incorrect Nos. Among the materials containing letters and words, syllables have the greatest number of wrong Nos and words and proverbs least. There is little or no difference in the materials in regard to per cent. of wrong Nos, except a possible tendency for the per cent. of wrong Nos to decrease with the increase of meaningful associations in the materials. These data show a tendency for increase in the per cent. of wrong Yeses and a decrease in the per cent. of wrong Nos to occur with the amount of associations present, but the former is more evident than the latter.

TABLE XXIX

Yes Used	No Used	Wrong Yeses	Wrong Nos	Wrong Yes All Yeses	Wrong No All Nos	Wrong Nos All Resp.	Wrong Nos All Resp.	% Correct Responses
910	1,074	158	235	17.5%	22.0%	8.0%	12.0%	84.0%
837	1,154	162	327	19.5%	28.5%	8.0%	16.5%	75.5%
804	1,189	68	244	8.0%	23.0%	3.5%	12.0%	84.5%
836	1,155	2,216	302	25.5%	26.0%	11.0%	15.5%	73.5%
886	1,008	112	226	13.0%	22.0%	5.7%	11.0%	73.5%
865	1,132	112.5	242.5	13.0%	21.5%	6.0%	12.5%	82.0%
844	1,147	280.5	428.5	33.0%	37.5%	14.0%	21.5%	64.0%
799	1,198	65.5	208.5	8.0%	23.5%	3.0%	14.0%	83.0%
754	1,244	196	426	26.0%	34.5%	10.0%	21.0%	69.0%
818	1,182	78	254	10.5%	21.5%	3.9%	12.9%	68.0%

The preceding table has compared the wrong Yeses among all Yeses used with wrong Nos among all Nos. The following table (XXX.) summarizes the data on the number of Yeses and Nos used and expresses the relation of *right* Yeses to all Yeses and right Nos to all Nos. In eight of the ten cases the per cent. of right Yeses among all Yeses is higher than the per cent. of right Nos among all Nos, in one case they are equal, and in one case less. These tables also show the per cent. of right responses among the old and new items. "Old" items are those which have been shown before and "New" items are the control items which had not been seen before. The correct responses to "old" items is Yes and to "new" items No. In each of the ten cases the per cent. of right responses among the new items is greater than among the old. The difference between the per cents varies from 5 to 25 per cent. The subjects appear to be more often correct in judging a thing as not seen before than as seen before.

TABLE XXX

RECOGNITION WORDS

	First Set, First Series				First Set, Second Series			
	Old	New	Total	% Right	Old	New	Total	% Right
Yes	740	78	818	90	752	158	910	83
No	254	928	1,182	79	235	839	1,074	78
Total	994	1,006	2,000		987	997	1,984	
% right ..	76	92			76	84		

	Second Set, First Series				Second Set Second Series			
	Old	New	Total	% Right	Old	New	Total	% Right
Yes	852.5	112.5	765	87	774	112	886	87
No	242.5	889.5	1,132	79	226	882	1,008	87
Total	995	1,002	1,997		1,000	994	1,894	
% right ..	76	89			77	89		

RECOGNITION FORMS

	First Set				Second Set			
	Old	New	Total	% Right	Old	New	Total	% Right
Yes	563.5	280.5	844	66	672	162	834	86
No	428.5	718.5	1,147	63	327	817	1,144	71
Total	992	999	1,991		999	989	1,978	
% right ..	57	72			67	83		

RECOGNITION PROVERBS

	First Set				Second Set			
	Old	New	Total	% Right	Old	New	Total	% Right
Yes	733.5	65.5	799	97	736	68	804	91
No	208.5	989.5	1,198	83	244	945	1,189	79
Total	942	1,055	1,997		980	1,013	1,993	
% right ..	78	94			75	93		

RECOGNITION SYLLABLES

	First Set				Second Set			
	Old	New	Total	% Right	Old	New	Total	% Right
Yes	558	196	754	74	534	302	836	64
No	426	818	1,244	66	216	939	1,155	81
Total	984	1,014	1,998		750	1,241	1,991	
% right ..	56	81			71	76		

3. *The Comparison of the Per Cent. of Wrong Yeses Among All Responses with the Per Cent. of Wrong Nos to All Responses*

Table XXIX. shows the per cent. of wrong YESES among all YESES and the per cent. of wrong NOS among all NOS. The per cent. of wrong YESES among all YESES varies in the first group from 3 to 14 per cent. and in the second group from 3.5 to 11 per cent. From the data of this experiment for both groups it appears that among materials containing letters or words, the highest per cent. of correct responses is for proverbs, and the lowest per cent. for syllables. Proverbs offer the greatest number of associations, then words, and syllables the least. The per cent. of wrong Nos among all responses varies in the first group from 12.7 per cent. to 21.5 per cent. and in the second from 11.9 per cent. to 16.5 per cent. The variation between the materials is less than it is for the per cent. of wrong Yeses among all responses. Among the materials, containing letters and words, proverbs have the highest per cent. of errors among the Nos and syllables the least. The order from highest to least per cent. of errors is the same for both sets—proverbs, words, syllables.

The per cent. of wrong Nos among all the responses is greater in each of the ten cases than the per cent. of wrong Yeses.

4. *The Per Cent. of All the Responses Correct for Each Material*

The per cent. of all the correct responses among all the responses is given in the last column of Table XXIX. According to the formula used, 50 per cent. would indicate a zero score and 100 per cent. a perfect score. The range in the first group is from 73.5 per cent. to 84.5 per cent. and in the second group is 64.5 per cent. to 83 per cent.—the order from highest to lowest for materials containing words and letters is proverbs, words, and syllables. The materials rich with associations are better recognized than materials where associations are few.

5. *The Comparison of the Per Cent. of Wrong Yeses Among the Men and Among the Women*

The data mentioned above have been for the two groups of adults. In each group are twenty men and twenty women. The scores for the men and for the women are given separately in Table XXXI., M. indicating male and F. female.

In the ten cases stating the per cent. of wrong Yeses among all the Yeses, the men have a higher per cent. than the women, in one case they are the same and in one case the per cent. of wrong Yeses for women is higher. Thus in eight out of the ten tests the women made few mistakes in selecting an item as one seen before.

The per cent. of wrong Nos among all the Nos used is also given in the table. In seven cases the per cent. of wrong Nos among all Nos is greater for men, in three cases it is greater for women. Thus in seven of the ten tests women make fewer mistakes than the men in selecting an item as one not seen before. In general the women are

TABLE XXXI

	Yeses Used	Nos Used	Wrong Yeses	Wrong Nos	Wrong Yes All Yeses	Wrong No All Nos	Wrong Yeses All Resp.	Wrong Nos All Resp.	%
F.	430	568	37	107	9%	19%	4.0%	11.0%	
M.	435	564	75.5	125.5	17%	24%	8.0%	14.0%	82.0%
F.	440	555	139.5	206.5	32%	37%	14.0%	21.0%	
M.	413	592	141	222	34%	38%	14.0%	22.0%	64.5%
F.	424	575	34.5	131.5	8%	23%	3.0%	13.0%	
M.	375	623	31	149	8%	24%	3.0%	15.0%	83.0%
F.	389	609	105	193	27%	32%	11.0%	19.0%	
M.	365	635	91	233	25%	37%	9.0%	23.0%	69.0%
F.	405	595	18	113	4%	19%	1.8%	11.3%	
M.	413	587	60	141	17%	24%	6.0%	14.1%	68.0%
F.	477	519	72	93	15%	18%	7.2%	9.3%	84.0%
M.	433	555	86	143	20%	26%	8.6%	14.3%	76.0%
F.	436	551	82	151	19%	27%	8.0%	15.0%	77.0%
M.	401	593	80	176	20%	30%	8.0%	17.6%	74.0%
F.	414	583	43	110	10%	26%	4.0%	11.0%	85.0%
M.	390	606	25	134	6%	22%	3.0%	13.0%	84.0%
F.	435	557	115.5	135.5	27%	24%	12.0%	14.0%	74.0%
M.	401	598	100.5	166.5	24%	28%	10.0%	17.0%	73.0%
F.	445	453	48	98	11%	21%	4.8%	10.9%	84.3%
M.	441	555	64	128	15%	23%	6.6%	12.9%	80.5%

superior to the men in the recognition of old items as old and of new items as new. By calculating the average per cent. of the wrong Yeses among all Yeses in the ten tests by women and by men the scores are 16.2 for the former and 18.6 for the latter or a difference of 2.4. The average per cent. of wrong Nos among all the Nos in the ten tests by women and by men are 24.6 for former and 27.6 for latter or a difference of 3.0.

6. *Comparison Between Men and Women*

In each of the ten tests the women have a higher per cent. of correct responses than the men. The difference between the sexes varies from 1 per cent. to 6.4 per cent. in the second group and from 2 per cent. to 7 per cent. in the first group. The women recognize better than the men do as it has been seen in Chapter V.

7. *The Comparison of Old and New Judged Correctly.*

How many old are judged correctly? How many new are judged correctly? The correct response to the Old is Yes and to the New is No. The data show that more Yeses for old occur than Nos for new—that is, the new are right oftener than the old. The new make a distinct impression and the subject responds with more certainty.

Among the old ones are failures of recognition but among the new are few false recognitions. Nos are used oftener than Yeses and more Nos are wrong. When Yes is used the subject has a fairly clear recognition. The Nos include errors due to failure to recognize but the Yeses do not include many new not recognized.

Methology of Scoring

In scoring for recognition the score of the positive and the score of the negative should be considered. The right Yeses minus the wrong Yeses equals the score of the positive; the right Nos minus the wrong Nos gives the score of the negative. The sum of the score of the positive and the score of the negative equals the Recognition Score.

"Negative" Recognition

In positive recognition the answer Yes is given to an item which is *old*. We recognize it as something familiar. In so-called negative recognition the answer is No to an item which seems new. There is a "newness for the new" which leads to the conclusion that the item has not been seen before. This strangeness or newness appears to be a positive thing. "Negative" recognition is not always the mere casting aside of something which lacks familiarity but rather something which possesses a "newness" or strangeness.

CHAPTER VIII

Conclusions

The present study has been interested in the two methods of testing memory, recall and recognition. To reproduce or recall what one has seen or heard is different from recognizing it as something previously seen or heard when it is presented again. To the writer both experiences equally seem to deserve the term memory, but the terms recall or reproduction and recognition should be used to distinguish them.

The results of the foregoing experiments on recall and recognition may be summarized as follows:

1. More items are recognized than recalled. The difference in recall and recognition memory is in part dependent on the richness of associations present.

2. Determined recall differs from undetermined recall more than determined recognition differs from undetermined recognition. In the tests where the subjects did not know that their memory would be measured the records are for undetermined recall and undetermined recognition, and in the tests where the subjects were aware that their memory was to be tested, the records are for determined recall and determined recognition. The difference between the records when the subjects knew and when they did not know their memory was being tested is greater in the tests for recall than in the tests for recognition.

3. The difference between recall and recognition is greater when the subjects did not know that their memory was to be tested than when they did know.

4. The influence of a determining factor is greater for the recall of material rich with associations than for material devoid of them. The advantage of making the observer determine to remember the material presented is greater when the material is meaningful and rich with associations than when it is nonsense. There is more difference in the scores for photographs than for nonsense syllables.

5. The determining factor influences the amount of the material remembered which can be correctly associated with other material remembered. The subjects could name the photographs better when they had tried to remember names and faces. This might have interesting applications in daily life. People who say that they can

not remember names and faces might be able to if they added more determination to remembering them and established associations between names and faces.

6. Primacy and recency both influence recall memory. The influence of each on recognition is less than on recall, but greater for material devoid of associations and less for material rich with associations.

7. Color variation does not increase the number of syllables recalled nor recognized so far as the results of this experiment show.

8. The variation in size of the syllables gave a higher score on the average than when the syllables were all the same size, but the difference is hardly significant.

9. The data indicate that a person who recalls a material well may recognize that kind well, fairly well, or poorly—we know little about one's recognition memory from a test of recall. There is a tendency, in general, for the correlation to be positive rather than negative.

10. The coefficients of correlation between recall of different materials show that a person who recalls one material well may or may not recall another material well. There is a tendency for the correlation to be positive rather than negative. There is little difference in materials, but among the adults tested, words are a better index of recall memory than the other materials.

11. The correlation between recognition of different materials is low. The person who recognizes one material well may or may not recognize another material well. There is a tendency for the coefficients of correlation to be positive rather than regative.

The correlation between recognition of different materials is about the same as the correlation between recall of different materials.

12. Some materials are more easily recalled than others. Some materials are more easily recognized than others. The number recognized is somewhat dependent upon how much confusion arises between the items previously seen and the new or control items. In general, for both recall and recognition the material with the greatest number of associations is remembered most easily and that with few or no associations is remembered poorly.

13. Among the adults tested the women appear to be superior to the men, on the average. The tendency for the women's scores to be higher is greatest in the scores for recall and recognition of words and in the recall of proverbs. The data show a tendency for girls to be superior to boys. This is in accordance with Pyle's experiments for recall of words where the norms for boys between the ages of

eight and eighteen are higher than those for the girls at the ages sixteen and seventeen only.

In general, there is no marked difference in achievement between recall and recognition for the sexes.

14. No difference in variability between the boys and girls is apparent when the Pearson Coefficient or the P.E. are used as the measures of variability. The coefficients of variability for adults are slightly higher for men but there is no striking sex difference.

15. For words, forms, and syllables for both recall and recognition the scores are higher for older ages than for younger ones. Both recall and recognition seem to improve with age.

16. There is a tendency for the scores to improve gradually from 4*A* through 4*B* for recall and for recognition of words, forms, and syllables, regardless of the age of the child.

17. The oldest children in a grade, on the average, do not have the best scores, but they do not always have the worst scores. The children at the median age for their grade do not have the best nor the worst scores. There is a tendency for the scores of those who are younger than their classmates to be higher, with the possible exception of syllables.

18. Among the insane patients tested there is no evidence of the disease influencing recall and recognition in different ways.

19. When the subject is asked to respond Yes to items which he recognizes as seen before and No to those not seen before, he responds No more often than Yes. The difference between the number of Yeses and Nos used is less for words than for other materials. There is no consistency between the two sets of material, however. For example, the difference between Yeses and Nos used in the tests for recognition of syllables are 399 and 385 and in the tests for recognition of syllables are 490 and 299.

20. A larger number of Nos are wrong than Yeses. From the data of this experiment it appears that among materials containing letters or words, the greatest per cent. of false recognitions is found in that material where the greatest meaning is present. The per cent. of false recognitions decreases when the meaningful associations in the material decrease.

In each of the ten tests the per cent. of right responses among the new items is greater than among the old. The subjects appear to be more often correct in judging a thing as not seen before than as seen before.

21. The per cent. of wrong Nos among all the responses is greater in each of the ten cases than the per cent. of wrong Yeses. Among

the materials containing letters and words, proverbs have the highest per cent. of errors among the Nos and syllables the least.

22. The new items are judged rightly oftener than the old are.

23. Women, in general, are superior to the men in recognition of old items as old and of new items as new in this experiment.

24. In each of the ten tests for recognition women have a higher per cent. of correct responses than men.

The writer has stated that both recall and recognition deserve to be included under the word memory. The field of memory might be illustrated by the crude simile of a bottle of milk—the method of recall measures those items which perhaps are the "cream." The threshold for recall is high, but items which can not be recalled may be above the threshold for recognition, its threshold being far lower. Those items which we can not quite recall are easily recognized and lie just below the threshold for recall. Thus the two processes, recall and recognition, should not be thought of as in opposition or methods whose scores differ by a gap. The difference in our memory is one of degree—the item may be easily recalled, recalled with difficulty, easily recognized, recognized with difficulty. The difference may be expressed as a difference in the distance above the lower threshold of memory. How short the distance seems when a word "we can not just recall" is mentioned and we instantly say "That is it" when we hear it!

Both methods test our memory. We should be aware that the threshold for each is at a different level and choose the method of testing according to which measure we wish to obtain.

References

1. Baldwin, J. M., and Shaw, W. J. Memory for Square Size. Fragments in Philosophy and Science.
2. Bean, C. H. Curve of Forgetting. Archives of Psychology.
3. Binet, A. Notes on Experimental Study of Memory. *Amer. Nat.*, XXXI.
   La Memorie des Mots. *L'Année Psychologique*, Vol. I.
   La Memorie des Phrases. *Année Psychologique*, Vol. II.
4. Bolton, F. L. Growth of Memory in School Children. *Amer. J. of Psych.*, Vol. IV., 1892.
5. Burnham, W. Memory. *Amer. J. of Psych.*, Vol. II.
6. Chamberlain, A.H. A Memory Test with School Children. *Psych. Rev.*, Vol. 22, 1915.
7. Ebbinghaus, H. Über das gedächtnis untersuchungen zur experimentallen psychologie, 1885.
8. Gates, Arthur I. Correlations and Sex Differences in Memory and Substitution. U. of Cal. publications, Vol. I. No. 6, 1916.
9. Gregor, A. Leitfaden in der experimentalle Psychpatholique, Chaper 2.
10. Heine, Rosa. *Zeitschr. f. Psychol.*, 1914.

11. HOLLINGWORTH, H. L. Characteristic Differences between Recall and Recognition. *Amer. J. of Psych.*, 1913.
12. KENNEDY, F. On the Experimental Investigation of Memory. *Psych. Rev.*, 1898.
13. McDOUGALL, ROBERT. Recognition and Recall. *J. of Phil.*, 1904.
14. MYERS, G. A. Incidental Memory. ARCHIVES OF PSYCHOLOGY.
A Comparative Study of Recognition and Recall. *Psych. Rev.*, Vol. XXI., No. 6, 1914.
Affective Factors in Recall. *J. of Phil.*, Vol. XII., No. 4, 1915.
15. PETERSON, H. A. Influence of Complexity and Dissimilarity on Memory.
Recall of Words, Objects and Movements. *Psych. Rev. Mon. Sup.* No. 4.
16. STRONG, E. K., JR. Effect of Length of Series upon Recognition Memory. *Psych. Rev.*, Vol. XIX., 1912.
Effect of Time Interval upon Recognition Memory. *Psych. Rev.*, Vol. XIX., 1912.
Two Factors which Influence Economical Learning. *J. of Phil.*, Vol. XI., 1914.
An Interesting Sex Difference. *Ped. Sem.*, Vol. XXII., 1915.
17. WOLFE, H. K. Untersuch. über d. Tongedächtniss, *Phil. Stud.*, Vol. III., 1886.

APPENDIX

The materials used which are reported in Chapter IV. were words, forms, syllables and proverbs. The words were simple nouns such as *flag, stove, hand,* etc. The forms were those used by Miss Whitley and printed in the Archives of Psychology No. 19. The syllables each had three letters, the first and third consonants, the middle one a vowel. The one hundred proverbs in the two sets of recognition series are appended.

Go farther and fare worse.
Brave actions never want a trumpet.
Too much rest itself becomes a pain.
A broken sack will hold no corn.
Feast to-day and fast to-morrow.
Justice delayed is justice denied.
Pleasing everybody is pleasing nobody.
Genius is nothing but an especial talent for patience.
Better go to bed supperless than rise in debt.
All is not gold that glitters.
A prophet hath no honor in his own country.
Noblest minds are easiest bent.
Happy is he who can live in peace.
Smooth runs the water where the brook is deep.
He who begins many things finishes few.
Idle folks have the least leisure.
One hour to-day is worth two to-morrow.
Better twice remembered than once forgotten.
Live to learn and learn to live.
It is better to be sure than sorry.
Pride joined with many virtues chokes them all.
Keep your shop and your shop keeps you.
It's an ill wind that does not blow someone good.
Enough is as good as a feast.
Great hopes make great men.
An honest countenance is the best passport.
Wilful waste makes woeful want.
New occasions teach new duties.
A stumble may prevent a fall.
Caution is the parent of safety.
The sweetest grapes hang highest.
Haste trips up both its heels.
Good counsel has no price.
Experience is the best teacher.
All is well that ends well.
A stitch in time saves nine.

Better late than never.
Gifts make beggars bold.
Too far east is west.
No news is good news.
Offenders never pardon.
Make hay while the sun shines.
Look before you leap.
The early bird catches the worm.
Beggars cannot be choosers
Many hands make light work.
Easy come easy go.
Guilt is always jealous.
A bird in the hand is worth two in the bush.
Better ask than go astray.
He who gives quickly, gives doubly.
No road is long with good company.
He who is well paid is well satisfied.
You can force an ox to water but you can't make him drink.
It never rains but it pours.
A burnt child dreads the fire.
A good hope is better than a bad possession.
Company in distress makes trouble less.
A full cup must be carried steadily.
A friend in need is a friend in deed.
Many cooks spoil the broth.
Abundant caution does no harm.
Practice makes perfect.
Every hill has a valley.
A small gift is better than a great promise.
A golden bit makes none the better horse.
A thing too much seen is little prized.
Necessity is the mother of invention.
One good turn deserves another.
Don't cross the bridge until you come to it.
A fog cannot be dispelled by a fan.
One must cut his coat according to his cloth.
All comes right to him who can wait.
Those who climb high often have a fall.
An empty bag cannot stand upright.
Lazy folks take the most pains.
A little spark kindles a great fire.
A handful of common sense is worth a bushel of learning.
Expect not at another's hand what you can do by your own.
Spin not too fine a thread lest it break in weaving it.
People who live in glass houses should not throw stones.
Opportunity knocks but once, for the world hates a knocker.
Birds of a feather flock together.
A penny saved is a penny gained.
Better cut the shoe than pinch the foot.
Where everyone goes the grass never grows.
There is no bush so small but casts its shadow.

You can't have the cake and eat it too.
Prevention is better than cure.
Faint heart never won fair lady.
A willing helper does not wait until he is asked.
Better twice measured than once wrong.
The boughs that bear most hang lowest.
The proof of the pudding is the eating.
Practice what you preach.
Imitation is the sincerest flattery.
Good coral needs no coloring.
A rolling stone gathers no moss.
Where there's fire there's smoke.
The better the day the better the deed.

ol.

THE MORPHOLOGIC ASPECT OF INTELLIGENCE

BY

SANTE NACCARATI, M.D., Sc.D., Ph.D.

ARCHIVES OF PSYCHOLOGY

EDITED BY

R. S. WOODWORTH

No. 45, August, 1921

Columbia University Contributions to Philosophy and Psychology
Volume XXVII, No. 2

NEW YORK

1921

Agents: G. E. STECHERT & CO., London (2 Star Yard, Carey St., W. C.);
Paris (16, rue de Condé).

THE MORPHOLOGIC ASPECT OF INTELLIGENCE

The problem of the correlation between bodily and mental traits has attracted the attention of educators, psychologists, physicians and sociologists during the last few decades.

For a great many years the problem has been laid merely on hypothesis and studied empirically; but only in recent years, namely after the introduction of the mental tests, has it been possible to approach the same with rational methods and to put it on a scientific basis.

In the matter of correlation with intelligence it must be admitted however that, while in the study of the tests for physical and motor capacity, different investigators have reported satisfactory and usually rather concordant results, the so-called anthropometric tests are still the subject of discrepancies and controversies.

Of course today we possess better scales for measuring intelligence; but the use of less accurate scales for intelligence on which former investigators had to rely alone does not justify the contrasting results obtained in the study of correlation between intelligence and anthropometric traits, such as height, weight, skull diameter, cephalic index, etc.

If we take height for instance, we find that Kline [1] reports that boys in public schools are taller than boys in truant schools. Smedley [2] reports that boys in the schools for incorrigibles and truants are shorter than normal boys, and that bright children are taller than dull children. These conclusions of Smedley's agree with those of Sack,[3] Gratianoff,[4] Porter,[5] Mac Donald,[6] De Busk [7] and others; while West [8] found the opposite to be true and Gilbert [9] failed to find any definite correlation between height and mental ability.

Again if we take weight, we find that Porter,[5] Smedley [2] and De Busk [7] reported that bright children are heavier than dull children of the same age; while for West [8] and Gilbert [9] the reverse is true.

The same conflicting reports were given by investigators who studied cephalic index, lung capacity, facial measurements, color of eyes, etc.

No wonder the results are so diverse. A single anthropometric measurement cannot constitute the characteristic of such a complex mental trait as intelligence, to which so many factors contribute.

Moreover height, weight, and cephalic index are traits much more constant than intelligence in the different races.

I cannot enter here in the discussion of Boas' views regarding the instability of the cephalic index. Probably Boas' contention has a scientific basis, if we consider that there is a certain correlation between the cephalic index and the morphologic type of the individual; therefore a modification of the cephalic index as a consequence of the transformation of the morphologic types produced by environmental factors through generations may be expected.

In the matter of height it may be said that, aside from external factors (environmental, social, political), an ethnic group tends to keep its average stature.

If we make a study of the population of the globe, we find people having tall, medium and short stature scattered all over from the very tall (m. 1.999) to the very short (m. 1.209) pygmies of Central Africa, not speaking of the pathological statures which are found beyond these extremes.

Now to admit that there exists a constant, definite correlation of height-intelligence would mean that all the people having a short stature, such as Lapps, Eskimos, Japanese, Hottentots, Negritos, Senois, etc. are not intelligent; and that people having tall stature such as Curds, Malays, Patagonians, Dinkas, inhabitants of some islands of the Pacific, etc. are intelligent, if we put them together, regardless of their respective race.

Everyone can see how absurd such an assumption would be, as any ethnic group includes in its community intelligent and unintelligent individuals. Of course if races could be kept pure, a physical trait such as height or weight may have a significance amongst the individuals of the same ethnic group; but nowadays with the continuous intermixture of races and with the great difficulty in differentiating the stocks from which the individuals spring, in cosmopolitan countries, one would do injustice to all the short individuals who originate from races having short statures, by regarding them less intelligent than the tall individuals of the same community, who owe their tall stature to hereditary factors.

The same may be said when weight or cephalic index are considered in connection with intelligence.

MORPHOLOGIC TYPES

After this consideration, it is obvious that none of the anthropologic traits alone could solve the problem of correlation with intelligence. As I said above intelligence is a most complex trait, therefore, I believe that any physical trait in order to be a correlative of intelligence must be a compound one, namely it must be a trait made up of many elementary traits.

Starting from this point of view I have made an anthropologic study of groups of individuals, aiming at the research of the morphologic characteristics of the intelligent type. This study has led me to the introduction of the morphologic index, following the criterion of the morphologic types individualized by De Giovanni [10] and Viola [11] for clinical purposes.

Viola, starting from the anthropometric studies of Broca, Bouchard, Manouvrier, Benecke and especially of his teacher De Giovanni, the founder of the clinical anthropology, and following the lines traced by the "Biometrika" in the study of the problems of evolution, after a diligent anthropometric study of 400 subjects, formulated his "law of deformation of the ethnic type," which reads as follows: "Individuals having a small trunk tend to assume a longilinear body which corresponds to the phtisic habitus; individuals having a large trunk tend to assume a short body which corresponds to the apopleptic habitus; individuals having a normal trunk tend to maintain normal proportions of the body." The so-called phtisic and apopleptic habitus are old denominations used by the ancient physicians to designate respectively a long thin and a short broad physical constitution.

According to the volume of the trunk in relation to the other portions of the body, Viola differentiated three morphologic types, the microsplanchnic, the macrosplanchnic and the normosplanchnic.

Microsplanchnics are individuals possessing a small trunk so that the development of the limbs is in excess over it, that is the vertical diameters predominate over the horizontal diameters in the body as a whole and in its constituents, trunk, extremities and portions of the extremities.

Macrosplanchnics or *Megalosplanchnics* are individuals possessing a large trunk which is excessively developed in comparison with the limbs; that is the horizontal diameters are prominent in comparison with the vertical diameters in the body as a whole and in its constituents, trunk, extremities and portions of the extremities.

Between these two opposite types are the *normosplanchnics* who represent individuals in which trunk and limbs show a harmonious development, in as much as neither one, when the numerical value of each is taken, predominates over the other; that is there exists a constant proportional relation between the horizontal and the vertical diameters of the body.

Of course it is difficult to draw a line of demarcation between the microsplanchnic and the normosplanchnic on one side, and between the macrosplanchnic and the normosplanchnic on the other. There is a great deal of overlapping. Viola among 400 subjects representing an ethnic group of Northern Italy found that 47.7% were normosplanchnic, 28% were megalosplanchnic and 24.3% were microsplanchnic. In making this classification he proceeds by finding the middle normal ethnic type and then calculates in degrees the deviations above and below the normal.

The limits in which this dissertation must be kept does not permit me to enter into the details of Viola's work. Students of Medicine and of Anthropology, who may be interested in it, are referred to the original publications given in the bibliography. My present problem is concerned with the ranking of groups of individuals when intelligence and morphologic aspect are taken as scales. Once we find a criterion for comparing the individuals of a given group, we will be able to rank them.

In order to put in a numerical form the morphologic characteristics of a group of individuals, one has to find the measure-value of the trunk and the measure-value of the limbs. The trunk, as Viola observes, contains the organs of the vegetative life, which represent the nutritional system. These organs fulfill a task different from the muscular and nervous systems and skeleton, which constitute the animal system or a system that mediates contact with the external world. These two systems show a certain degree of independence and even antagonism during the development; in the sense that they do not grow simultaneously, but in alternate phases; and the more an organism develops the animal system, the less it develops the vegetative system when considered in relation of their reciprocal dependence.

The difficult task confronted in this study is to find the value of the trunk, namely the volume of the abdominal and thoracic cavities.

Viola takes 11 measurements, namely:

1. *Height.*

2. *Length of sternum:* (A B)—from the jugular incisure to the point of insertion of the ensiform appendix.

3. *Length xipho-epigastric:* (B C)—from the point of insertion of the xiphoid appendix to the epigastric point. This point (C) is at the crossing of the middle vertical line of the trunk with the horizontal line passing through the lower margin of the tenth rib (Z W).

4. *Length epigastric-pubic:* (C D)—from the epigastric point to the upper margin of the pubis.

5. *Length of the lower extremities:* (V T)—from the upper margin of the pubic bone to the external malleous of the foot.

6. *Length of the upper extremities:* (R S)—from the margin of the acromion process to the wrist-joint while the arms hang down; (I have preferred the stiloid process of the radius as point of repere).

7. *Transverse thoracic diameter* or breadth diameter taken at the level of the 4th rib (E F).

8. *Antero-posterior thoracic diameter* or depth diameter taken also at the level of the 4th rib (M N).

9. *Transverse epigastric diameter* taken at the mid-point of the xipho-epigastric line (G H).

10. *Antero-posterior epigastric diameter* taken as the same level of the preceding diameter (P Q).

11. *Transverse pelvic diameter* taken between the iliac crestae at the point of the maximum breadth (I L).

(For the explanation of letters in parentheses see figure 1 and the annexed anthropometric blank.)

Viola has devised special instruments for the morphologic measurements, but one who is familiar with anthropometry can obtain practically the same results with a little more time and patience, by using an anthropometric tape, a chest depth caliper, a chest breadth caliper and a height stand.

For the treatment of the anthropometric data Viola proceeds in the following way:

He obtains a *thoracic index* or a *thoracic value* by multiplying the length of the sternum by the transverse thoracic diameter and by the antero-posterior thoracic diameter (AB × EF × MN). By multiplying the length xipho-epigastric by the transverse epigastric diameter and by the antero-posterior epigastric diameter he obtains the *index of the upper abdomen* (BC × GH × PQ). *The index of the lower abdomen* is obtained by multiplying the length pubo-epigastric by the transverse pelvic diameter and by the antero-posterior epigastric diameter (CD × IL × PQ), (the antero-posterior pelvic diameter is not taken).

Sum of the indices of the upper and lower abdomen gives the *total abdominal value.*

Sum of the thoracic value with the total abdominal value gives the *value of the trunk.*

The value of the limbs is obtained by adding the length of one of the upper limbs with that of one lower limb (RS+TV).

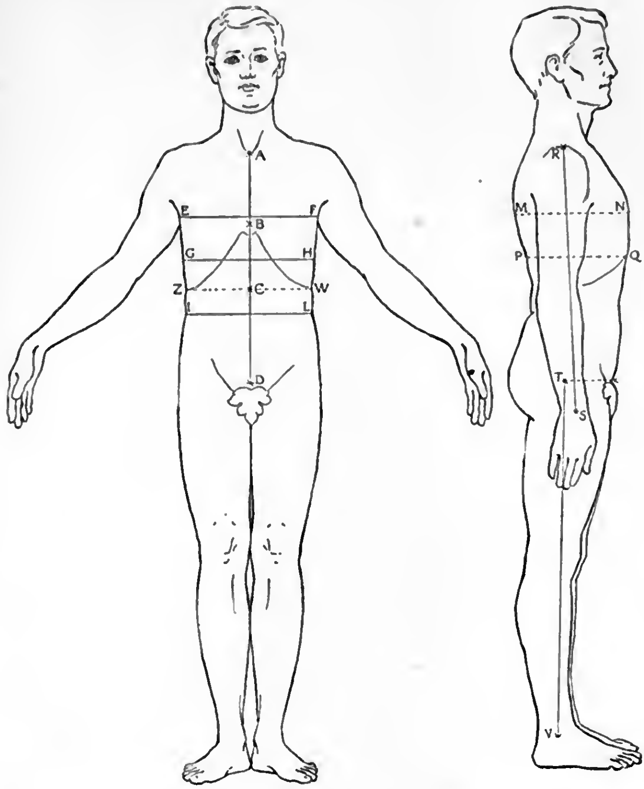


FIG. 1. Showing how measurements are taken. For the explanation of the symbols see below.

ANTHROPOMETRIC BLANK

No. . . . . .Name . . . . Age . . . .
Address . . . .
1. Length of sternum . . . . AB . . .
2. Xipho-epigastric line . . . . BC . . . AD
3. Pubo-epigastric line . . . . CD . . .
4. Transverse thoracic diameter . . . . EF . . . .
5. Antero-posterior thoracic diameter . . . . . MN . . . .
6. Transverse epigastric diameter . . . . . . GH . . . .
7. Antero-posterior epigastric diameter . . . PQ . . . .
8. Transverse pelvic diameter . . . . IL . . . .
9. Length of upper extremity . . . . RS . . . .
10. Length of lower extremity . . . . TV . . . .
11. Height . . . .
12. Weight . . . .

MORPHOLOGIC INDEX AS AN INDICATOR OF INTELLIGENCE

For a long time I have observed that bright individuals during the period of growth in length tend to maintain the vertical diameters of the body and of the components of the body, trunk and extremities, relatively in excess over the respective horizontal diameters, when compared with less bright or dull individuals. In other words bright boys tend to grow in length rather than in width in a relatively greater proportion than dull boys of the same chronological age: that is what made me think that bright boys tend toward *microsplanchny*. Therefore I came to the conclusion that the very bright child is likely to be microsplanchnic, not merely feeble in health, as it was the pretension of a common belief, now pretty thoroughly exploded.

In order to demonstrate my theory, I have first tried to find an anthropometric index which could take in account the length of the extremities and the volume of the trunk with the purpose of correlating it with the intelligence score.

The index which I first used, and which I have named "*morphologic index*"[1] is given by the ratio of length of the limbs to value of the trunk. The higher indices will represent the microsplanchnics, the lower indices the macrosplanchnics. In other words if we distribute a group of individuals on a frequency curve, according to their morphologic index, we should find the microsplanchnics and macrosplanchnics occupying the ends of the curve and the normosplanchnics the center. As I said above, I have not tried to fix the limits between the three types: there is much overlapping.

After an anthropometric study of two groups of 50 and 75 students respectively, I found that there exists a positive correlation of +.7 and even more between the ratio of height to weight, and the ratio of limbs value to trunk value: therefore, I assumed the ratio of height to weight as an approximate indicator of the morphologic index, and took that ratio as a simpler method

[1] Bean has presented a paper on "The Morphologic Index" at the 37th session of the American Association of Anatomists, March 24–26 of this year. For this author the morphologic index represents the percentage above or below the world average of any anthropometric character, such as stature, cephalic index, nasal index, etc.

I like to call the attention of the reader to the fact that Bean's morphologic index is entirely different from mine, and that his conception on the morphologic types does not correspond to that of the Italian school.

of study for several other groups, whose morphologic measurements could not be taken, without pretending that the ratio of height to weight may substitute the morphologic index all the time when correlating mental traits.

The advantages of using the morphologic index or the ratio of height to weight over either height or weight alone, in the correlation with mental traits, are obvious.

One does not need to be tall in order to be microsplanchnic nor does the macrosplanchnic need to be short. In my groups one can find short subjects among the microsplanchnics and tall subjects among the macrosplanchnics. Microsplanchnics, macrosplanchnics and normosplanchnics are found almost in the same proportions in all the ethnic groups, and one does not need to look into remote ancestry of the individuals in a cosmopolitan community. Without denying that an ethnic group may give more individuals of a definite morphologic type in the same way as it may give more intelligent or less intelligent types, it must be admitted though that such influence of a predominant morphologic type in a stock is never so great as stature. Predominancy of a given type is undoubtedly well pronounced in the three different human races, white, yellow and black. In this respect Stratz's [12] distinction of the human types in (A) *Leukederm* (white races) in which limbs and trunk are proportionately developed, (B) *Melanoderm* (Negro races) in which exists an excess of the limbs over the trunk, (C) *Xantoderm* (yellow races) in which exists an excess of the trunk over the limbs, has to be kept in mind when studying groups of subjects, in order that the groups may be kept homegeneous.

Ranke [13] divided the human races into races of culture and races of nature including in the first group the white and the yellows who tend to brachyskely (short limbs) and in the second group the negroes who tend to dolichoskely (long limbs). Of course in the mind of the German anthropologist no hint existed as to the conception of hyperevolution and hypoevolution of the individuals brought forward in the differentiation of the two opposite types the microsplanchnic and the macrosplanchnic. His study was racial and his distinction after all was theoretical, because there are people of culture showing both dolichoskely and brachyskely, as *e.g.* Mediterranean and Baltic races respectively brachyskele and dolichoskele, and people of nature such as Bushmen and Australians respectively brachyskele and dolichoskele. From the individual point of view we consider the dolichoskele type of culture and the brachyskele type of nature, whether macrosomatic or microsomatic.

Recent attempts are being made to correlate facial measurements with intelligence, but the method by which the problem has been approached is not free from phrenological influence. I think though that combined head and face measurements may lead to some practical results if they are directed to the discovery of fœtal characteristics in adults which would indicate hypoevolution of the individual. The same may be said of the intermembrae index, of the vital index of Goldstein and of any other indices intended to bring forward morphologic signs suggestive of physical hypoevolution. However it must be remembered that in the matter of head and face measurements we have to deal with small differences, which increase the sources of error and give very little chance for comparison.

The assumption I made that intelligent subjects are more likely to be found among microsplanchnics than among macrosplanchnics is based on these three physiological facts:

1. The relative independence existing in the growth of the two great systems, differentiated by Bichat, namely the nutritional or vegetative system, in which energy is stored up, and the animal system (mainly organs of locomotion) by which energy is transformed and utilized or wasted.

2. The physical hyperevolution of the microsplanchnic type which hyperevolution I think must exist also for the mental characteristics.

3. The microsplanchnics correspond to the hyperthyroid types.

Regarding the first fact, I have to add that by the term animal system is meant the nervous system, the muscles and the skeleton, which are systems of relation in as much as they mediate contact between the individual and the external world. They constitute about 60% of the total weight of the body.

In general there is a positive correlation between the growth of the internal organs and all the other systems: but the rate of growth of the internal organs and of the other tissues which constitute the system of nutrition or vegetative life, when considered in relation to the other organs and apparatus constituting the animal system, is not the same in microsplanchnics and macrosplanchnics. This rate during the period of development is subject to individual variations which have not been well estimated so far. Generally speaking it may be said that in normosplanchnic subjects it is kept relatively constant and proportionate, but in macrosplanchnics and microsplanchnics the rate of growth of one system (nutritional) does not correspond, in relation of reciprocal dependence, to the rate of the other system (animal) and vice versa. Therefore the microsplanchnic is apt to develop relatively more in the animal system, the macrosplanchnic more in the vegetative system.

One may observe that muscles and bones constitute the greater part of the weight of the body, whereas the viscera represent a small fraction. Judging from this point of view weight should constitute a positive and not a negative factor as it appears from the correlation between ratio height to weight and intelligence. While this cannot be denied, it must be admitted though that in true macrosplanchnic types the weight of the viscera, fat and cutaneous annexes, which in normosplanchnic individuals represent about 40% of the body weight, go beyond the 40% at the expenses of the systems of the animal life. In some individuals having tendency to obesity the muscular fibers are usually infiltrated with adipose tissue; therefore even a method which could give the absolute weight of the striated muscles, such as the volumetric estimation of a limb by displacement of fluid, would not be free from error because it would not take into account the quality of the muscular system.

But there is another element to be considered in the larger volumetric or ponderal mass of the muscles of the macrosplanchnic. The macrosplanchnics, having short limbs, have also short muscles and smaller portion of attachment on the bones: this, Viola points out, means that the muscular system was *primarily* deficient and only *secondarily* became excessive, owing to the increase in size and number of muscular fibers due to the excessive nutritional activity of the organism. Besides a short large muscle may have advantage over a long thin muscle in what concerns amount or quantity of energy, not in what concerns quality of achievement; a thin long muscle may give poorer but more highly specialized movements.

The substitution of the ratio of height to weight for the morphologic index practically eliminates the error due to muscle influence, in as much as the muscles of the limbs are neither added nor subtracted in our calculation. For this reason probably the correlation morphologic index-intelligence was found larger than the correlation ratio height to weight-intelligence. At any rate even the morphologic index is not free from errors of computation.

Since an ideal method which could give the exact value of both the animal and the vegetative systems, taking into account quantity and quality, in living subjects is not possible, we have to be satisfied with methods which give us approximate values.

Regarding the second fact, here is the summary of the proofs brought by Viola to demonstrate that the microsplanchnic is a hyperevolute type and the macrosplanchnic is hypoevolute from the point of view of physical development.

All the characteristics which differentiate the newborn from the adult are found in the macrosplanchnics; all the characteristics of the adult are found in the microsplanchnics in an exaggerated form. The microsplanchnics in comparison with the normosplanchnics show: thorax flatter *in toto*, narrower at the basis and more predominant in volume over the abdomen; umbilicus more distant from the pubis; ribs more inclined; costo-vertebral angles smaller; heart more vertical; the thorax has an expiratory shape; diaphragmatic vault is pronouncedly convex toward the thorax; lungs longer; Luska's incisure well accentuated; lower extremities more developed in comparison with the upper extremities; hands and feet longer; arch of foot more pronounced.

In macrosplanchnics, as in newborn, the system of vegetative life prevails over the system of animal life. During the period of growth the visceral system gradually decreases, passing from the infantile megalosplanchnic to the normosplanchnic or to the microsplanchnic. Thus the macrosplanchnic adult has been left behind in its development in comparison with the normosplanchnic, while the microsplanchnic has gone beyond.

The characteristics which liken the macrosplanchnic to the newborn and differentiate him from the normosplanchnic are: relatively larger size of liver, spleen and other abdominal organs; abdomen predominant in volume over the thorax; thorax larger at the base; antero-posterior thoracic diameter larger and prominent over the transverse diameter; (the antero-posterior thoracic diameter is gradually reduced during ontogenesis, while the transverse diameter is increased); less of forward inclination of the ribs; costo-vertebral angles larger; more forward projection of the sternum; lungs shorter; thorax has a general inspiratory shape; diaphragmatic vault is almost flat; lower extremities are relatively short in comparison with the upper extremities; flat foot; broad hands and feet.

There are other points of differentiation between the two opposite types macrosplanchnic and microsplanchnic, which bear no little weight on the general activity of the organism, particularly on the nervous reactions.

The megalosplanchnics have a rather small cutaneous surface in relation to their body volume, whereas microsplanchnics have a relatively larger cutaneous surface and a small body volume. The first have little dispersion of heat, the second have a great dispersion. When we come to consider that the 9/10 of the total energy of an organism is dispersed under form of heat, we must conclude that such a great dispersion is not a waste of energy to the organism but an employment somehow and somewhere beneficial to it, although we are still unable to discover how and where this employment takes place. If such enormous dispersion is necessary for biological potential activities yet unknown to us, we may suppose that microsplanchnics must gain some sort of compensatory advantages for their larger dispersion of heat.

External stimuli, which have so much influence on our nervous system and, therefore, on our mood and behavior must necessarily act in a different way on the relatively larger receptive surface of the microsplanchnic and on the relatively small cutaneous surface of the macrosplanchnic. Take for instance such cosmic stimuli as sun heat, sun light, barometric pressure, etc., we all know how much influence they have on our mood and consequently on our behavior. We usually give importance to the atmospheric conditions in a general sense, but we never take into consideration the important element of our somatic individuality. Now the macrosplanchnics having a relatively reduced receptive surface for the external stimuli must have retarded nervous reactions, comparatively little sensitive life, lessened psychic functions: whereas the opposite must be true of the microsplanchnics who will be quicker but more exhaustible in their nervous reactions, and will possess a greater degree of sensitiveness to pain, thermal and electric stimuli.

Regarding the third fact, it is nowadays accepted by endocrinologists that in physiological hyperthyroidism the organism tends to grow more along the vertical diameters, namely it tends to grow in length rather than in width. Thyroid hormones have some relationship with intelligence. Witness cretinism and the manifold indications which come to us from clinical cases. Schlesinger (13) has recently reported that in a region where goiter is endemic, the growth and development of the children with this hyperthyroidism are usually in advance of their years, both physically and mentally.

Now a few points must be made clear before concluding.

I do not say that intelligence can be measured with ratio of length of limbs to volume of the trunk, or with ratio of height to weight in the sense that the higher the morphologic index is, the more intelligent is the subject and vice versa. Probably the best intelligence is not found among the highest microsplanchnics, as these are likely to be borderline pathologic cases. I simply say that generally speaking individuals showing a microsplanchnic type in a given group have more likelihood to be intelligent, than those who have a macrosplanchnic tendency. Normosplanchnics show all degrees of intelligence. Usually they tend to normality also in the intellectual domain. If we wanted to calculate the degree of intelligence by the degree of *microsplanchny*, we would err, in as much as the ultramicrosplanchnic types are pathological subjects who have attained that high morphologic index in consequence of exhausting diseases, usually tuberculosis. The volume of the trunk in my 125 individuals ranges from a minimum of 17.57, found in a healthy boy 17 years old to a maximum of 40.68, found in a subject 30 years of age. Viola found in 400 Northern Italian subjects a minimum of 18.28 and a maximum of 58.96, having taken subjects much older than mine and concluded that below a value of 18.28 for the trunk, human existance becomes highly improbable beyond the age of 20. Undoubtedly Viola's conclusion is correct. Above the age of 20, a value of trunk below 18 can be found only in pathological cases, in the white race.

One must not think that *microsplanchny* is usually the outcome of exhaustion as in tuberculosis and of excess of mental work and reduced diet in students, as it may appear at first sight. The morphologic type is outlined during the period of growth as the exponent of internal factors which, we know now, are intimately connected with the function of the endocrines. The diet of a microsplanchnic may be increased and enriched; the subject will gain a few pounds but will not lose his morphologic individuality in the sense that he will become a macrosplanchnic, unless functional or pathological changes in the endocrine glands occur. The microsplanchnic being a hyperthyroid possesses a constitutionally weak digestive system and a poor assimilative power, therefore overfeeding if protracted will be injurious rather than beneficial. Conversely an elimination diet will not effect the transformation of a macrosplanchnic into a microsplanchnic, without danger to the organism.

Of course from the 25th to the 35th year of age the thorax expands laterally. Besides physiological factors, such as pregnancy and menopause in women, environmental factors such as change of climate, occupation, marriage, etc., and diseases may induce profound modifications within the internal secretory glands and thus the altered metabolism may cause great gain or loss of weight after the age of 25. In this way the morphologic type may be transformed from what it was primarily. But this transformation is unusual in normal individuals before the growth in length has ceased; therefore, when speaking of the morphologic type of an adult, one has to refer to measurements taken before the age of 25 years. For this important reason I have taken college students considering that at the age of 20 the individual has attained the 99% of his stature. Measurements taken in subjects older than 25 may give the fallacious impression that we have to deal with macrosplanchnics in individuals who were primarily normosplanchnic or even microsplanchnic. This fact must be kept in mind when taking morphologic indices for correlation with intelligence.

When I say that measurements should be taken between the ages of 20 and 25, I do not deny that the same, if taken in children and in adolescents and in older individuals may give good results. Children and youths may be suitable subjects for morphological experiments provided they be grouped according to their respective physiological ages. This fact is important because, as it is known from physiology, at about the seventh year of age and at the puberal period the organism shows a rapid growth in length and therefore it tends toward *microsplanchny:* whereas during infancy, after the 25th year of age, and a few years before pubescence it shows a definite tendency toward *macrosplanchny.* Thus in each normal individual the highest degree of *microsplanchny* is attained at puberty, while the highest degree of *macrosplanchny* exists at birth.

According to Pende (15), during the periods of greater growth in height, a physiologic hyperactivity of the hormones, which promote the development of the animal system, exists, i.e. some hormones of the thyroid, of the hypophysis, of the cromaffin tissue, of the endocrine tissue of the sex glands. These hormones promote and stimulate also the neuropsychic activity, and possess a certain degree of inhibition over the hormones which favor the nutritional or vegetative system.

I feel justified in making the foregoing statements, because I believe that some of the hormones, chiefly thyroid hormones, which during the period of growth regulate the morphology of the body, influence also the development of the mentality. Mind in the same way as the body is shaped before the subject has attained his full stature, and we may have some idea of the hormonic actions exerted upon the mind, looking through the morphologic type.

Summing up, in saying that the microsplanchnic type is the intelligent individual, I do not extend the assumption to pathological cases, nor do I affirm that the macrosplanchnic type cannot be intelligent. This statement must be made clear in order to avoid erroneous interpretations of my theory, which are likely to follow in the form of criticism, as was true in the case of Lombroso's theory on the somatic features of the criminal, so that every mark of degeneracy was taken to mean an indication of criminality or degeneration, an assertion which Lombroso never made.

HORMONES AND MORPHOLOGIC TYPES

Although we do not yet possess a full knowledge of the different hormones of each of the ductless glands and therefore their respective definite functions, physiological and clinical observation enable us to make a tentative classification of the different morphologic types studied from the endocrine point of view. In the *microsplanchnic* type we find characteristic features of hyperthyroidism of hyperpituitarism and hypogenitalism, while in the *macrosplanchnic* type we find characteristics of the hypothyroidism, hypopituitarism and to some an extent also of hyperadreninism (to limit ourselves to the study of those glands whose function is better known).

Of course what was said for the morphologic types namely that a sharp line of demarcation between the microsplanchnic type and the normosplanchnic type on one side and between the macrosplanchnic type and the normosplanchnic type on the other, cannot be drawn since both microsplanchnic and macrosplanchnic types merge into the normosplanchnic type when a group of individuals is recorded on a frequency curve, is also true for the hyperthyroids and hyperpituitarics on the one side and hypothyroids and hypopituitarics on the other, in the sense that a sharp distinction between each of the two groups and the normal individuals does not exist. When we speak of hyperthyroids and hypothyroids of hyperpituitarics and hypopituitarics, we do not mean the real pathological cases but we refer only to those individuals having a constitutional or a congenital hyperactivity and hypoactivity respectively speaking which is kept under physiological limits. Pende and Levi and Rothschild have respectively named "Ortoplastic Hyperthyroidism" and "Hyperthyroid Temperament" a condition of constitutional hyperthyroidism in which the toxic and dysharmonic hormonic actions characterizing pathologic hyperthyroidism do not exist.

If we examine the microsplanchnic type we find many of the subjective and objective signs of hyperthyroidism viz: wide palpebral fissures, large pupils, glistening eye, long eye lashes, thin moist shining hair; well developed and healthy teeth and nails; hands and fingers long and thin, this being a part of the general tendency of the hyperthyroid to grow in length rather than in width; skin usually moist, especially that of the hands and feet which are also warm, with tendency to show rapid vasomotor and secretory changes; digestion and assimilation irregular and defective, a condition which renders the hyperthyroids very cautious in the selection and use of food; circulatory system shows tendency to an inverted oculo-cardiac reflex (negative reflex index) (16), to tachicardia and to arterial hypertension, although pulse and blood pressure show large fluctuations during the same day; baldness and gray hair rare or late to appear.

Similarly as macrosplanchnics show somatic resemblance to infants, that is they keep the somatic characteristics of infants, the hyperthyroids (microsplanchnics) tend to preserve the somatic features of youth. For this reason the adult hyperthyroid looks younger and maintains a youthful complexion even in mature age. This is due to the fact that in hyperthyroids the hyperactivity of the hormones which promote the development of the animal system and which are very active throughout the period of growth in length, predominates even after the end of said period: while in the macrosplanchnics, the antagonistic group of hormones, which promote the development of the visceral system, instead of giving way to the other group of hormones at the proper time, has persisted in its physiologic hyperactivity thus giving the macrosplanchnic the infantile trend of the body.

Not less striking are the psychic characteristics of the hyperthyroids. As Viola observes, the microsplanchnic, owing to the possession of a minimal organic mass and a maximal surface area of the body, is by nature endowed with a strong catalytic stimulation and with an eretistic nervous system. His large receptive surface area renders him excitable, and any form of external influence affects his type more than any other. On account of their limited muscular expansion, the hyperthyroids do not indulge in athletic exercises and take little interest in the practical side of life but conversely acquire a great transport for its aesthetic side. Therefore the hyperthyroids love indoor games, music, poetry, arts in general, theatre, readings and works of the nature of scientific research. Also their minds not unlike their bodies tend to preserve the characteristics of youth, so they are rather prone to day dreaming and to being absent-minded. Being intelligent they possess live ideation, prompt perception, vast imagination, strong memory, and shrewd critique; but lack of concentration and unsteady will power may hinder their learning capacity. Cœnæsthetic variations are at the bottom of their exaggerated emotional display and of their rapid changes in mood which they often show. Most of the manic-depressive characteristics are found amongst the hyperthyroids, who are likely to pass from one emotional state to the opposite one, from enthusiasm to pessimism for slight causes. Endowed with exquisite sensibility the hyperthyroid feels joys and griefs deeply; offenses and wrongs do not pass without a long and profound repercussion on his mood. Too much mental repression and the quick muscular exhaustion make the hyperthyroid a prospective candidate for the asthenic forms of psychoneurosis.

In general it must be said that the macrosplanchnics on account of their hyperactive vegetative functions show prominently acts and instinctive reactions, such as eating, sleeping, sexual impulse, etc., which satisfy the somatic self, whereas the microsplanchnics develop more the mechanisms intended to preserve the psychic self.

Intuition is one of the characteristics of hyperthyroids. When we consider this characteristic in connection with their quick physical and mental exhaustibility, a practical conclusion can be drawn from the point of view of vocational guidance. The intelligence of the hyperthyroid possesses more intensity than duration, it acts as a stored up energy which can be better employed as explosive material. Hyperthyroids should be directed toward those disciplines for which inventory capacity is needed; invention is in some respects intuition plus condensed mental energy which acts in an explosive manner. Occupations which require long, patient application and too much concentration are not fit for the bright but exhaustible hyperthyroid.

Conversely the macrosplanchnic, who possesses a vigorous *physique*, can persist longer than the microsplanchnic in any kind of physical and mental work. Therefore macrosplanchnics compensate with duration and amount of performances what they cannot accomplish by intensity and quality. The achievements of the macrosplanchnics are due to the determination and endurance they show in bringing about a given task. Rather by persistent application than by intuitive genius can they solve such difficult problems in which the microsplanchnics would hardly succeed, before reaching their limit of exhaustibility.

SURVEY OF EXPERIMENTAL WORK

This research work has been carried out on College and University students of the white race. The tests used correspond to the best hitherto devised for measuring intelligence, namely the "Alpha Army Test," the "Otis Intelligence Test," and "Thorndike Intelligence Examination Tests." The above mentioned tests, especially the ones devised by Thorndike, greatly reduce the chances of error in what concerns the measurement of one of the traits which I have correlated; *i.e.*, intelligence.

I omit the full description of these tests, as they can be easily obtained from the authors or publishers.

Tests embodying the principles of the "Alpha" and "Otis" scales were used on more than a million and a half men upon their entry into the United States army. They are designed to test general mental ability and are suitable for all literate persons. The scales of both Alpha and Otis tests consist of 8 to 10 tests respectively, each test consisting of a series of questions or problems. There are several forms of the Alpha test, and two forms of the Otis test, A and B; these forms, while different in substance are similar in structure; and the Total Point Scores of one Alpha Test form are equivalent to those of the other Alpha forms; likewise, the Total Point Scores of one Otis form are equivalent to those of the other Otis form. The purpose in constructing independent scales is to provide for reexamination after a short interval without the scores being influenced by memory of previous questions, and to prevent collusion between groups successively examined.

Without entering into the details of the scoring of the Alpha and the Otis tests, I simply mention that scores range from 0 to 212 for "Alpha" and from 0 to 230 for "Otis." In our groups scores ranged from 76 to 176 and from 121 to 220 respectively.

The Thorndike Intelligence Examination Tests were devised by Prof. Thorndike with the purpose of supplanting the old-fashioned examinations for students entering colleges, schools of engineering and Professional Schools. The good results obtained during the first two years these intelligence examinations were administered authorize us to consider their scores the best indicators of intelligence at the present time.

Here are a few points I take from the Standard Instructions:

The examination is composed of three parts. Part I comprises two forms of 13 tests each: Part II comprises 8 tests, and Part III is made up of three tests, a total of 37 tests.

For the purpose of giving the student a notion of the examination he is to take, a fore-exercise with 13 tests of part I precedes the examination; the student being informed that this trial is not a part of the examination.

Part I comprises two examinations of the type of the army Alpha, but extended and made harder; part II and III are made up of examinations of different sorts, including the tests which have been found by Haggerty and Thorndike to be specially indicative of ability to succeed with the work of a college or professional school.

A Thorndike Intelligence examination can be administered in about 3½ hours, time being distributed as follows:

10 minutes for fore-exercise with Part I, Practice Form.
30 " Trial I with Part I, Form E.
30 " Trial II with Part I, Form F.
60 " Part II.
40 " Part III.
From 20 to 35 minutes for explanation, distribution of papers and rest period.

One examiner can take care of 40 to 50 men, if he is quick in distributing the material.

The score is determined by quality of achievement more than by speed, except in so far as the latter is an essential consequence or accompaniment of quality.

Scoring does not require great pains or special ability. Standard instructions and keys for scoring the tests are so arranged that any college officer can give the examination and that at least 80% of the tests can be scored by any clerk. With proper supervision and enough clerks, any number of tests can be scored in a single day.

The total scores vary from about 40 to 100 with a few records above 100.

Individuals 17 years old or older who score less than 60 are as a rule unsuitable material for college education. If under 17 they may be submitted to a second examination after one year's study. Individuals scoring from 60 to 69 have, perhaps, enough intellect to attain a college degree, if they are specially earnest and industrious. Individuals scoring 70 or over possess intellect such as is adequate for college work and may be admitted with no risk of detriment to present standards. Individuals scoring 85 or over may be safely admitted regardless of even gross deficiencies in their preparations.

The scores of the subjects studied by me range from a maximum of 117 to a maximum of 54. Their ages vary from 17 to 22 years. Some were younger when they took the examination.

Aside from these college boys who constitute the greater number of my subjects, I have examined two other groups of 50 and 94 Summer Session students of both sexes, whose ages in many cases were above 25 years.

Except for the last two groups the anthropometric measurements were taken at the gymnasium in metric system units.

Regarding the reliability of the physical measurements a few points need to be explained.

Weight and height do not require any particular skill, with a good scale and a reliable stadiometer any intelligent examiner can claim to be accurate.

Some anatomical knowledge and anthropometric training is required before one can advance any claim to accuracy in the anthropometric measurements.

The anatomical points must be first established on the subject; it is better to mark them down with a dermographic pencil. The vertical measurements are taken between the marked points with an anthropometric tape. For the transverse and antero-posterior diameters of the trunk the *Seaver Rod Caliper,* which can be used for breadth as well as for depth measurements has served my purpose most satisfactorily.

Control measurements are advisable, as for instance in the case of the three first measurements Sternum, Xipho-epigastric line and Pubo-epigastric line, one should ascertain himself that the whole length Sternum-Pubis corresponds to the sum of the three different segments.

Measurements should be extended to millimeters. An error of 1 mm. will reduce or increase the index to about 10 counts. An error of ½ cm. plus or minus would cause a change of 50 counts in the index. This probable error in excess or in defect in all the measurements is of course rare, error usually occurring in one or two measurements with the probability of being compensated by errors in defects or if not compensated at all, with the modification of a few counts on the morphologic index. At any rate such errors will have no substantial influence on the correlation and in no instance will they effect the transformation of a microsplanchnic into a macrosplachnic and vice versa.

In the group of 75, the morphologic index ranges from 435 to 721. Now by disregarding the third digit the indices are computed from 43 to 72 making 30 steps, the correlation will persist.

In the group of 50 the morphologic index ranges from 360 to 671.

The reduction of the morphologic index in this group is due to age influence; after puberty the index diminishes as the age increases.

Complete anthropometric measurements were taken on:

1. A group of 50 students. Of these only a few returned for the Otis test, therefore, no correlation between morphologic index and intelligence could be studied. At any rate this group has served to demonstrate the correlation existing between the ratio height to weight and length of limbs to volume of the trunk, which had to be expected. The coefficient of this correlation has been found equal to + .75 with a P.E. of 0.05. The existence of a high positive correlation between morphologic index and ratio height to weight (as shown in this group and in the other group of 75) indicates without further explanation what is the factor which causes a positive correlation between ratio and intelligence.

2. A group of 75 students, who took the entrance examination test of Professor Thorndike. The coefficient of correlation was found equal to +.35, P.E., .068. The coefficient of correlation ratio height to weight-intelligence in this same group was only + .16, with a P.E. of .08. The co-efficient of correlation between ratio length of limbs to volume of trunk and ratio height to weight in this same group was found equal to + .70 with a P.E. of 0.04.

Ratio height to weight was correlated with intelligence in the other four groups, respectively of 50, 94, 80, 100 individuals. The group of 50 who were given the Alpha test and the group of 94 who were given the Otis test were studied in co-operation with Lewy Guinzburg. These tables will be given in another work when the measurements of a larger group, now under study, is completed. The coefficient of correlation were +. 44 and + .14 with P.E. equal to 0.06 and 0.07 respectively.

The groups of 80 and 100 as the other group of 75 were scored with the ratings obtained at the entrance examinations, given by Prof. Thorndike: they yielded coefficients of + .27, + .33, + .16 with P.E. of .07, .06, .08 respectively.

In order to reduce the P.E. I made a single group out of the three groups of 75, 80 and 100 subjects who took the same entrance examination test. As each of the three groups were taken at different times, it happened that a few students appeared in more than one group; therefore out of 245 students a group of 221 was obtained, whose correlation of the ratio of height to weight to intelligence gave a coefficient of + .228 with a P.E. of .044.

The ratio height to weight in the group of 221 ranges from 2.066 to 3.565.

I like to call the attention to the fact that the subjects used in my research, being usually intelligent individuals who had resisted the practical elimination tests of school, offer small intelligence differences. If the same research is carried out on groups of individuals who did not receive high school education, the larger intelligence ranges which will be found will probably cause the correlation to be higher.

In none of the groups studied by me was any correlation found between height and intelligence, nor between lung capacity and intelligence, the latter being — .10 in 136 subjects.

Weight and volume of trunk yielded negative correlations with intelligence.

My correlation reviewed by a statistician (Wood) with different methods gave the following coefficients:

r 221 subjects,	Intelligence	—height		+.0415
" 221	"	"	—weight	−.183
" 221	"	"	—ratio H:W	+.230
" 75	"	"	—length of limbs	+.155
" 75	"	"	—volume of trunk	−.360
" 75	"	"	—morphologic index	+.356
" 136	"	"	—lung capacity	−.105

SUMMARY

1. Intelligence cannot be correlated with a simple physical trait such as height, weight, cephalic index, etc. A basis for correlation must be found in a compound physical trait which is made up of several anthropometric traits.
2. By morphologic type is meant the physical constitution of the individual when the development of the extremities and that of the trunk are reciprocally considered and compared. The "Morphologic index" of an individual is given by the ratio value of the extremities (length of one upper and one lower limb) to the volume of the trunk.
3. Among the three morphologic types differentiated by Viola, viz: the microsplanchnic, the normosplanchnic and the macrosplanchnic, the microsplanchnic gives the more intelligent units. This does not mean that the normosplanchnics and the macrosplanchnics are not intelligent. Normosplanchnics represent all degrees of intelligence: macrosplanchnics, while representing a less intelligent group, individually may be as intelligent as any other type.
4. The microsplanchnic type, being an intelligent type, is mentally hyperevolute. This fact constitutes another proof in support of Viola's thesis that the macrosplanchnic type and not the microsplanchnic is an infantile type and therefore retarded in his development.
5. The microsplanchnic type corresponds to the hyperthyroid type, namely to an individual who possesses a constitutionally hyperactive thyroid, an "Orthoplastic Hyperthyroidism" (Pende), or a "Hyperthyroid Temperament" (Levi and Rothschild). In this respect my research may constitute support to the common belief that the thyroid function bears relationship to intelligence
6. The morphologic type is the outcome of hereditary and accidental factors. Hormonic actions, which, as we know from the study of the endocrine glands, influence the physical and mental growth, the metabolism and the nervous reactions, are in part responsible for the individual differences in the same family and in the same ethnic group.

   During the prenatal life and during infancy and the adolescent

period, environment, diseases, traumata (physical and moral) improper diet, etc., may affect the function of one or more endocrine glands, and thus produce morphologic deviations which are usually accompanied by mental deficiencies or gains on account of interrelations existing between the morphoregulator and the neuroregulator hormones.

After the age of 25 external factors such as marriage, occupation, diet, climate, etc., may cause changes in the organism by which the morphologic type may be more or less modified from what it primarily was. Therefore morphologic index and ratio height to weight taken in adults should be referred to the age of 25.

7. Owing to racial tendency toward brachiskely and dolichoskely, one should avoid putting together in the same group, when taking the morphologic index, individuals of different races (namely the white, the yellow and the negroid).
8. My experimental study has shown that a positive correlation exists between intelligence and the ratio of height to weight. The average coefficient of correlation found in the group of 221 students was equal to + .228 with a P.E. equal to .044.
9. The ratio of height to weight gives an approximate indication of the morphologic type of the individual, since there is a correlation of about + .7 or more between that ratio and the morphologic index, as shown by the analysis of 125 students, whose anthropometric measurements were taken.

   Individuals showing a high morphologic index and a high ratio of height to weight correspond to the microsplanchnics; the lower indices and lower ratios correspond to the macrosplanchnics.
10. A higher positive correlation is found when instead of the ratio of height to weight, the morphologic index is taken as the expression of the type of the subjects. The morphologic index is found by dividing the length value of the two limbs by the volumetric value of the trunk.

    A group of 75 male students which gave a correlation of +.16 with the ratio of height to weight gave a coefficient of +.35 and a P.E. equal to .068, with the morphologic index.
11. No correlation was found in any one of the groups examined between height and intelligence. Similarly no correlation was found to exist between lung capacity and intelligence in 136 of the students making the group of 221, whose lung capacity had been measured. Weight gave a negative correlation.

In conclusion I am indebted to Prof. Woodworth for the helpful suggestions given me in the course of this work, to Prof. Thorndike and Mr. Wood for the use of the Entrance Examinations ratings, and to Dr. Meylan for the opportunity of the Gymnasium rooms and instruments he afforded me for my anthropometric research.

TABLE I

No.	Height	Weight	Ratio H:W	Int. Score
1	171	59.3	2.883	117
2	165	64.5	2.558	114
3	154	47.6	3.235	107
4	166	62.1	2.673	107
5	166	54	3.074	104
6	166	56.2	2.954	104
7	165	53.1	3.107	103
8	184.5	54.2	3.404	101
9	167.6	61.2	2.739	101
10	174.4	64.7	2.695	100
11	166.5	55	3.027	100
12	170	62.3	2.729	100
13	165.5	60.1	2.754	99
14	173	62.5	2.768	98
15	162.8	56.4	2.887	98
16	164	48.1	3.411	97
17	167	78.6	2.125	94
18	177.1	55.8	3.174	93
19	175	52.7	3.320	93
20	162.5	70.1	2.318	92
21	178	57.2	3.112	91
22	165.4	51.1	3.236	91
23	168.4	59.3	2.840	90
24	171	59.4	2.879	90
25	155.1	43.1	3.598	89
26	171	60.8	2.812	89
27	149.5	53.5	2.794	88
28	180.1	84.8	2.124	88
29	168	65	2.584	88
30	166.5	65	2.562	88
31	167	53	3.151	87
32	170.4	58.1	2.933	87
33	165	54.2	3.044	86
34	175.5	58.9	2.980	86
35	181.6	60.7	2.992	86
36	177	55.3	3.201	86
37	175	65	2.692	86
38	167	58	2.879	86
39	170.5	58	2.940	85
40	178	63.4	2.808	85
41	176.1	59.6	2.955	85
42	166	54.4	3.051	85
43	171.5	60.5	2.835	84
44	174	69	2.521	84
45	183	69.5	2.633	84
46	176	52.4	3.368	84
47	173.9	66.3	2.623	84
48	163.6	52.6	3.110	84
49	174.4	59.5	2.931	83
50	179	68.5	2.613	83

TABLE I—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
51	171	58.6	2.918	83
52	164.9	70	2.356	82
53	174.4	64.4	2.708	82
54	174	60.8	2.861	82
55	168.8	55.7	3.030	81
56	166.6	60	2.777	81
57	166	63	2.635	81
58	162	48	3.375	80
59	183.2	67	2.734	80
60	179.3	70.2	2.554	80
61	169.1	65	2.602	79
62	175.4	61	2.875	79
63	164	46	3.565	78
64	159.1	65	2.448	78
65	170	56.6	3.003	77
66	157.7	59	2.673	76
67	162.5	62	2.621	75
68	169.6	61.3	2.767	74
69	166.3	68.3	2.435	74
70	184	61.1	3.011	74
71	182	69.5	2.619	73
72	162	63	2.571	73
73	168.7	66.3	2.544	72
74	163.4	59.2	2.760	72
75	172	69.2	2.485	72
76	181.6	72	2.522	71
77	161	66.5	2.421	69
78	180.1	68.4	2.633	69
79	173	56	3.089	69
80	171.8	56	3.068	68
81	177.4	68.2	2.601	68
82	169	55.6	3.219	68
83	176	66.5	2.647	68
84	170.3	54.3	3.136	67
85	165	62.2	2.653	66
86	176.5	66	2.674	66
87	171.5	83	2.066	66
88	171.5	61	2.811	65
89	170	67	2.537	65
90	171.4	62	2.764	65
91	168.3	56	3.005	65
92	177	68	2.603	65
93	166.4	74.2	2.243	65
94	179.5	74.5	2.409	64
95	178.3	64.7	2.756	64
96	176	65.5	2.687	64
97	163.3	60	2.722	62
98	176.5	73.8	2.392	62
99	172.2	64.9	2.653	60
100	187	81.4	2.297	59

TABLE II

No.	Height	Weight	Ratio H:W	Int. Score
1	175.5	61.1	2.872	71
2	182.7	62	2.947	95
3	163.7	49.3	3.320	78
4	184	61.1	3.011	74
5	183.2	66.2	2.767	98
6	172.8	54.4	3.176	93
7	173.5	62.2	2.789	73
8	176.5	58	3.043	60
9	182.3	70	2.604	79
10	175.8	61.8	2.845	85
11	179.9	63.4	2.837	87
12	164.9	63.7	2.601	97
13	154	47.6	3.235	107
14	166	54	3.074	104
15	167.5	54.3	3.085	71
16	179.1	68	2.634	97
17	166	56.2	2.954	104
18	163.5	51.6	3.169	96
19	171.7	59	2.910	101
20	180.4	74.6	2.418	88
21	178.8	79	2.263	90
22	177	68	2.603	91
23	170.8	61.1	2.795	65
24	159.7	57.3	2.787	59
25	172.2	64.9	2.653	60
26	183.1	71	2.579	54
27	174.8	73.2	2.388	59
28	172	47.9	3.591	108
29	172	60.2	2.857	105
30	167.4	56	2.989	104
31	174.3	60.6	2.876	111
32	172.2	58	2.969	108
33	165.8	51	3.251	110
34	186.5	59	3.161	112
35	178	54	3.305	112
36	157.3	58	2.712	108
37	175.3	60	2.922	108
38	176	60.3	2.919	112
39	178.8	62	2.884	110
40	180.5	62.8	2.874	114
41	173	58	2.983	107
42	174.8	67.2	2.601	107
43	185.4	70.3	2.637	108
44	167.5	53	3.160	109
45	171.7	52.7	3.258	111
46	179.2	66.3	2.701	79
47	174.3	59.2	2.944	93
48	176.7	66.8	2.645	102
49	179.2	60.9	2.942	79

TABLE II—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
50	174	67	2.597	89
51	168.5	60	2.808	81
52	179	61.6	2.906	74
53	170.1	59.4	2.864	103
54	163.7	55.2	2.965	97
55	179.9	71	2.534	78
56	171.2	61.2	2.797	101
57	170	62.2	2.733	65
58	161.4	50	3.228	98
59	172	68.3	2.518	68
60	179.4	68.4	2.623	85
61	159.5	50.8	3.139	71
62	178.5	72.6	2.486	100
63	173.2	66	2.624	59
64	176.5	66.2	2.666	87
65	160.2	54.8	2.923	71
66	173.4	62.2	2.788	89
67	179	68.6	2.609	103
68	181.8	86	2.114	85
69	167.3	59.2	2.826	77
70	175	53.4	3.277	76
71	176.1	60.6	2.906	65
72	181	68.2	2.654	101
73	163.4	61.4	2.661	99
74	170.5	62	2.750	77
75	176	69	2.551	101
76	171	58.8	2.908	110
77	173.6	77.8	2.231	81
78	165.5	66.4	2.492	87
79	171.9	69	2.491	63
80	162.7	48	3.389	69

TABLE III

No.	Height	Weight	Ratio H:W	Int. Score
1	171	59.3	2.883	117
2	165	64.5	2.558	114
3	172	47.9	3.591	108
4	166	62.1	2.673	107
5	154	47.6	3.235	107
6	166	56.2	2.954	104
7	166	54	3.074	104
8	164.1	59	2.781	103
9	182	73	2.493	102
10	166.5	55	3.027	100
11	176	59.8	2.943	98
12	164	48.1	3.411	97
13	169	56.8	2.975	95
14	177.5	69	2.572	94
15	175	52.7	3.320	93
16	164.5	63	2.611	93
17	170	60.4	2.814	91
18	171	59.4	2.879	90
19	155.1	43.1	3.565	89
20	169	61	2.770	89
21	149.5	53.5	2.794	88
22	168	65	2.584	88
23	173	64.6	2.678	87
24	174	55	3.163	87
25	183	67.2	2.723	86
26	186.5	77.8	2.397	86
27	170.5	76.5	2.229	86
28	171	64.4	2.655	86
29	177	55.3	3.201	86
30	175	65	2.692	86
31	167	58	2.879	86
32	170.5	58	2.939	85
33	182	73	2.493	85
34	175	64	2.734	85
35	166	54.4	3.051	85
36	168	65	2.585	84
37	166	51	3.362	84
38	171.5	60.5	2.835	84
39	174	69	2.521	84
40	179	68.5	2.613	83
41	170	69.9	2.432	83
42	161.5	75	2.153	83
43	177	69.4	2.550	83
44	174	60.8	2.861	82
45	164	52.1	3.147	82
46	178	67.5	2.637	82
47	176	57.2	3.077	81
48	166	63	2.635	81
49	174	62.1	2.802	80

TABLE III—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
50	168	56.4	2.979	79
51	172	65.7	2.618	79
52	164	59	2.779	78
53	171	61.5	2.780	78
54	186	72	2.583	78
55	171	62.7	2.727	76
56	157	56.8	2.763	75
57	178.5	67.3	2.652	74
58	160	54.3	2.946	74
59	167	60.8	2.747	74
60	184	61.1	3.011	74
61	182	69.5	2.619	73
62	175	68.6	2.551	73
63	163	56	2.910	73
64	162	63	2.571	73
65	178	63	2.825	72
66	157	63	2.492	72
67	163.5	63	2.595	72
68	157.5	52	3.028	70
69	175	79	2.215	70
70	176	66.5	2.647	68
71	174	57.2	3.041	68
72	171	54	3.167	66
73	177	68	2.603	65
74	172	65	2.646	63
75	171	63.1	2.714	55

TABLE IV

No.	Int.	Length Limbs	Volume Trunk	Morphologic Index
1	79	136	26.73	5.08
2	63	140	26.88	5.21
3	80	142	24.70	5.75
4	117	144	21.57	6.67
5	88	122	20.99	5.81
6	103	133	20.18	6.59
7	70	144	29.56	4.87
8	87	142	21.45	6.62
9	86	149	26.71	5.58
10	68	145	28.72	5.04
11	93	145	21.24	6.82
12	86	139	31.92	4.35
13	95	139	21.02	6.61
14	86	153	32.08	4.77
15	83	141	30.05	4.69
16	89	126	20.26	6.22
17	79	137	21.79	6.28
18	84	140	26.15	5.35
19	76	136	28.85	4.71
20	66	129	22.49	5.73
21	72	132	26.53	4.97
22	83	138	28.96	4.76
23	83	130	29.86	4.35
24	84	138	19.12	7.21
25	86	142	22.77	6.23
26	88	142	24.68	5.75
27	84	142	25.56	5.55
28	86	145	26.73	5.42
29	65	142	25.23	5.62
30	73	133	24.20	5.49
31	86	139	24.18	5.74
32	85	149	31.31	4.75
33	74	140	29.62	4.72
34	108	136	19.91	6.83
35	107	133.5	24.59	5.43
36	104	135	24.34	5.54
37	74	126	26.41	4.77
38	78	157.5	25.38	6.20
39	91	140	22.33	6.27
40	114	137	22.39	6.11
41	73	144.5	25.09	5.75
42	85	139.5	20.65	6.75
43	84	144	26.08	5.52
44	74	146	27.81	5.25
45	55	138.5	25.73	5.38
46	97	133	21.75	6.11
47	75	129.5	23.04	5.62
48	104	138	22.80	6.05
49	100	142	22.32	6.36

TABLE IV—(*Continued*)

No.	Int.	Length Limbs	Volume Trunk	Morphologic Index
50	81	132	27.91	4.73
51	98	143.5	22.47	6.38
52	68	138	23.96	5.96
53	82	140	21.09	6.63
54	73	139	26.42	5.26
55	87	138	25.09	5.50
56	86	141	21.19	6.65
57	102	145	28.21	5.14
58	82	146	26.48	5.51
59	93	136.5	23.39	5.83
60	86	136	24.73	5.49
61	83	143	25.19	5.67
62	85	138	24.78	5.57
63	72	129	23.88	5.40
64	73	136	23.15	5.88
65	72	138	23.66	5.83
66	78	136	20.45	6.65
67	74	138	25.53	5.40
68	90	142	24.20	5.86
69	81	144	23.86	6.03
70	89	137	24.54	5.58
71	70	130	19.59	6.63
72	94	148.5	28.39	5.23
73	82	134	24.79	4.82
74	107	126.5	17.57	7.20
75	78	133	23.72	5.60

TABLE V

No.	Height	Weight	Ratio H:W	Int. Score
1	171	59.3	2.883	117
2	180.5	62.8	2.874	114
3	165	64.5	2.558	114
4	186.5	59	3.161	112
5	178.5	54	3.305	112
6	176	60.3	2.919	112
7	174.3	60.6	2.876	111
8	171.7	52.7	3.258	111
9	165.8	51	3.251	110
10	171	58.8	2.908	110
11	178.8	62	2.884	110
12	167.5	53	3.160	109
13	172	47.9	3.591	108
14	157.3	58	2.712	108
15	172.2	58	2.969	108
16	185.4	70.3	2.637	108
17	175.3	60	2.922	108
18	166	62.1	2.673	107
19	154	47.6	3.235	107
20	174.8	67.2	2.601	107
21	173	58	2.983	107
22	172	60.2	2.857	105
23	167.4	56	2.989	104
24	166	56.2	2.954	104
25	166	54	3.074	104
26	170.1	59.4	2.864	103
27	179	68.6	2.609	103
28	165	53.1	3.107	103
29	164.1	59	2.781	103
30	176.7	66.8	2.645	102
31	182	73	2.493	102
32	184.5	54.2	3.404	101
33	167.6	61.2	2.739	101
34	171.2	61.2	2.797	101
35	181	68.2	2.654	101
36	176	69	2.551	101
37	171.7	59	2.910	101
38	174.4	64.7	2.695	100
39	170	62.3	2.729	100
40	178.5	72.6	2.486	100
41	166.5	55	3.027	100
42	163.4	61.4	2.661	99
43	165.5	60.1	2.754	99
44	183.2	66.2	2.767	98
45	173	62.5	2.768	98
46	162.8	56.4	2.887	98
47	161.4	50	3.228	98
48	176	59.8	2.943	98
49	163.7	55.2	2.965	97

TABLE V—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
50	164	48.1	3.411	97
51	164.9	63.4	2.601	97
52	179.1	68	2.634	97
53	163.5	51.6	3.169	96
54	182.7	62	2.947	95
55	169	56.8	2.975	95
56	167	78.6	2.125	94
57	177.5	69	2.572	94
58	177.1	55.8	3.174	93
59	174.3	59.2	2.944	93
60	175	52.7	3.320	93
61	164.5	63	2.611	93
62	172.8	54.4	3.176	93
63	162.5	70.1	2.318	92
64	177	68	2.603	91
65	178	57.2	3.112	91
66	165.4	51.1	3.236	91
67	170	60.4	2.814	91
68	168.4	59.3	2.840	90
69	178.8	79	2.263	90
70	171	59.4	2.879	90
71	155.1	43.1	3.598	89
72	171	60.8	2.812	89
73	174	67	2.597	89
74	173.4	62.2	2.788	89
75	169	61	2.770	89
76	149.5	53.5	2.794	88
77	180.1	84.8	2.124	88
78	166.5	65	2.562	88
79	168	65	2.584	88
80	180.4	74.6	2.418	88
81	167	53	3.151	87
82	170.4	58.1	2.933	87
83	176.5	66.2	2.666	87
84	173	64.6	2.678	87
85	179.9	63.4	2.837	87
86	165.5	66.4	2.492	87
87	174	55	3.163	87
88	183	67.2	2.723	86
89	186.5	77.8	2.397	86
90	165	54.2	3.044	86
91	175.5	58.9	2.980	86
92	181.6	60.7	2.992	86
93	170.5	76.5	2.229	86
94	171	64.4	2.655	86
95	177	55.3	3.201	86
96	175	65	2.692	86
97	167	58	2.879	86
98	170.5	58	2.939	85

TABLE V—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
99	178	63.4	2.808	85
100	176.1	59.6	2.955	85
101	179.4	68.4	2.623	85
102	181.8	86	2.114	85
103	182	73	2.493	85
104	175	64	2.734	85
105	175.8	61.8	2.845	85
106	166	54.4	3.051	85
107	183	69.5	2.633	84
108	176.5	52.4	3.368	84
109	173.9	66.3	2.623	84
110	163.6	52.6	3.110	84
111	168	65	2.585	84
112	171.5	51	3.362	84
113	171.5	60.5	2.835	84
114	174	69	2.521	84
115	174.4	59.5	2.931	83
116	179	68.5	2.613	83
117	171	58.6	2.918	83
118	170	69.9	2.432	83
119	161.5	75	2.153	83
120	177	69.4	2.550	83
121	164.9	70	2.356	82
122	174.4	64.4	2.708	82
123	174	60.8	2.861	82
124	164	52.1	3.147	82
125	178	67.5	2.637	82
126	168.8	55.7	3.030	81
127	166.6	60	2.777	81
128	168.5	60	2.808	81
129	173.6	77.8	2.231	81
130	176	57.2	3.077	81
131	166	63	2.635	81
132	162	48	3.375	80
133	183.2	67	2.734	80
134	179.3	70.2	2.554	80
135	174	62.1	2.802	80
136	169.1	65	2.602	79
137	175.4	61	2.875	79
138	182.3	70	2.604	79
139	179.1	66.3	2.701	79
140	179.2	60.9	2.942	79
141	168	56.4	2.979	79
142	172	65.7	2.618	79
143	164	59	2.779	78
144	164	46	3.565	78
145	159.1	65	2.448	78
146	163.7	49.3	3.320	78
147	179.9	71	2.534	78

TABLE V—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
148	171	61.5	2.780	78
149	186	72	2.583	78
150	170	56.6	3.003	77
151	167.3	59.2	2.826	77
152	170.5	62	2.750	77
153	157.7	59	2.673	76
154	175	53.4	3.277	76
155	171	62.7	2.727	76
156	162.5	62	2.621	75
157	157	56.8	2.764	75
158	178.5	67.3	2.652	74
159	169.6	61.3	2.767	74
160	166.3	68.3	2.435	74
161	179	61.6	2.906	74
162	160	54.3	2.946	74
163	167	60.8	2.747	74
164	184	61.1	3.011	74
165	182	69.5	2.619	73
166	173.5	62.2	2.789	73
167	175	68.6	2.551	73
168	163	56	2.910	73
169	162	63	2.571	73
170	168.7	66.3	2.544	72
171	163.4	59.2	2.760	72
172	172	69.2	2.485	72
173	178	63	2.825	72
174	157	63	2.492	72
175	163.5	63	2.595	72
176	181.6	72	2.522	71
177	167.5	54.3	3.085	71
178	159.5	50.8	3.139	71
179	160.2	54.8	2.923	71
180	175.5	61.1	2.872	71
181	157.5	52	3.028	70
182	175	79	2.215	70
183	161.1	66.5	2.421	69
184	180	68.4	2.633	69
185	173	56	3.089	69
186	127.6	48	3.389	69
187	171.8	56	3.068	68
188	177.4	68.2	2.601	68
189	179	55.6	3.219	68
190	172	68.3	2.518	68
191	176	66.5	2.647	68
192	174	57.2	3.041	68
193	170.3	54.3	3.136	67
194	171	54	3.167	66
195	165	62.2	2.653	66
196	176.5	66	2.674	66

TABLE V—(*Continued*)

No.	Height	Weight	Ratio H:W	Int. Score
197	171.5	83	2.066	66
198	171.5	61	2.811	65
199	170	67	2.537	65
200	171.4	62	2.764	65
201	168.3	56	3.005	65
202	166.4	74.2	2.243	65
203	170	62.2	2.733	65
204	176.1	60.6	2.906	65
205	177	68	2.603	65
206	170.8	61.1	2.795	65
207	179.5	74.5	2.409	64
208	178.3	64.7	2.756	64
209	176	65.5	2.687	64
210	171.9	69	2.491	63
211	172	65	2.646	63
212	163.3	60	2.722	62
213	176.5	73.8	2.392	62
214	172.2	64.9	2.653	60
215	176.5	58	3.043	60
216	174.8	73.2	2.388	59
217	187	81.4	2.297	59
218	173. 2	66	2.624	59
219	159.7	57.3	2.787	59
220	171	63	2.714	55
221	183.1	71	2.579	54

TABLE VI

No.	Int. Score	Lung Capacity	No.	Int. Score	Lung Capacity
1	117	460	50	86	400
2	114	350	51	85	465
3	107	450	52	85	330
4	107	250	53	85	400
5	104	390	54	85	565
6	104	400	55	85	410
7	103	390	56	85	360
8	103	320	57	84	470
9	102	450	58	84	390
10	101	435	59	84	460
11	101	390	60	84	370
12	100	360	61	84	340
13	100	430	62	84	380
14	100	330	63	84	380
15	99	240	64	84	460
16	98	340	65	83	350
17	98	250	66	83	480
18	98	350	67	83	370
19	97	300	68	83	410
20	95	280	69	82	290
21	94	400	70	82	440
22	94	440	71	82	400
23	93	410	72	82	320
24	93	400	73	82	330
25	93	430	74	81	460
26	92	350	75	81	380
27	91	410	76	81	400
28	91	350	77	81	400
29	90	410	78	80	330
30	90	410	79	80	440
31	89	260	80	80	420
32	89	390	81	80	420
33	89	470	82	79	420
34	88	330	83	79	390
35	88	350	84	79	350
36	88	380	85	79	460
37	88	400	86	78	350
38	87	300	87	78	230
39	87	330	88	78	350
40	87	420	89	78	370
41	86	500	90	77	430
42	86	450	91	76	310
43	86	300	92	75	430
44	86	440	93	75	240
45	86	370	94	74	380
46	86	480	95	74	250
47	86	380	96	74	380
48	86	370	97	74	360
49	86	350	98	74	410

TABLE VI—(*Continued*)

No.	Int. Score	Lung Capacity	No.	Int. Score	Lung Capacity
99	73	360	118	68	380
100	73	340	119	67	330
101	73	350	120	66	410
102	73	410	121	66	470
103	72	360	122	66	360
104	72	380	123	65	330
105	72	360	124	65	400
106	72	330	125	65	500
107	72	340	126	65	350
108	71	420	127	65	370
109	70	320	128	65	340
110	70	410	129	64	450
111	69	420	130	64	390
112	69	470	131	64	500
113	69	410	132	63	330
114	68	460	133	62	300
115	68	460	134	62	410
116	68	350	135	59	520
117	68	450	136	55	440

TABLE VII

No.	Length Limbs	Volume Trunk	Morph. Index	Height in cm.	Weight in lbs.	Ratio H:W
1	126.5	18.84	671	157.5	113	139
2	123.9	20.28	610	152	98	155
3	139.4	24.19	576	173	120	144
4	140.2	25.70	545	172.5	135	127
5	129.5	24.60	526	161	121	133
6	136	25.97	523	166	115	144
7	138.1	26.50	521	171.5	135	127
8	141.6	27.40	516	169	124	136
9	146.7	28.50	514	179	135	132
10	157.7	30.69	513	187.5	176	106
11	150.3	29.40	511	178.5	128	139
12	145.5	28.67	507	172	132	130
13	152.4	30.31	502	181	133	136
14	156.5	31.32	499	191	145	131
15	134.7	27.40	491	169	128	132
16	132.3	26.97	490	166	121	137
17	147.3	29.70	495	181	146	123
18	144	30.18	477	172.2	145	119
19	143	30.00	476	174.5	160	109
20	144.4	30.44	474	179	151	118
21	145.6	31.30	465	179	139	128
22	146.5	31.80	460	178	149	119
23	142.7	31.00	460	183.5	146	125
24	136.3	30.10	452	170	126	134
25	140	31.20	448	181	155	116
26	145.6	32.70	445	181	160	113
27	133.8	30.41	440	172	131	131
28	144.1	32.53	442	179	145	123
29	137	31.43	436	167	145	115
30	138.1	31.73	435	171.5	135	127
31	138.4	32.00	432	170	146	116
32	130.8	30.70	426	162.5	135	120
33	141	33.04	426	171	145	117
34	140.3	33.51	418	172	144	119
35	144.9	34.77	416	180	158	114
36	155	37.51	413	189	175	108
37	132.6	32.46	408	170	145	117
38	144.9	35.44	408	177	166	106
39	137.6	34.61	397	164	147	111
40	134.4	33.83	397	173.5	142	122
41	142.4	35.90	396	175.5	165	106
42	127.7	32.02	398	167.5	130	128
43	138.8	35.83	387	172.5	150	115
44	142.2	37.05	383	177.5	154	114
45	138.8	36.34	381	172	150	115
46	145.5	38.34	379	180	165	109
47	138.2	36.68	376	170	161	105
48	130.6	35.00	373	175	152	115
49	148.5	40.68	365	180.5	175	103
50	132	36.57	360	170	160	106

SYNOPSIS OF THE TABLES

TABLES No. I, II and III give height (cm.), weight (kg.) ratio height to weight and intelligence ratings (Thorndike Intelligence Examination) of groups of 100, 80 and 75 college students.

TABLE No. V gives the same subjects of tables I, II and III in a single group of 221. Some subjects appeared in more than one table.

TABLE No. IV gives length of limbs, volume of trunk and morphologic index of the same individuals of table No. III.

TABLE No. VI gives lung capacity of 136 subjects among the group of table V.

Table VII—giving length of limbs, volume of trunk, morpholog'c index, height, weight and ratio of height to weight in a group of 50 University students.

BIBLIOGRAPHY

1. KLINE, L. W. Truancy as related to the migratory instinct. *Pedagogical Seminary*, 5, 1898.
2. SMEDLEY, F. W. Report of the Department of child-study and pedagogic investigation. Reprint from 46th Annual Report Board Education, Chicago, 1899–1900, 1900–1901.
3. SACK, N. Physical development of the children in the middle schools of Moscow, 1892. Quoted by Whipple in Manual of mental and physical tests. Baltimore, 1914.
4. GRATIANOFF. Quoted by Whipple, *ibidem.*
5. PORTER, W. T. The growth of St. Louis children. *Transactions Academy of Sciences*, St. Louis, **6,** 1894.
   Physical basis of precocity and dullness, *Ibid.*, **6,** 1893–4.
6. MACDONALD, A. Experimental study of school children, etc. Report United States Commissioner of Education, 1899.
7. DEBUSK, B. W. Height, weight, vital capacity and retardation. *Pedagogical Seminary*, **20,** 1913.
8. WEST, G. Observations on the relation of physical development to intellectual ability made on the school children of Toronto, Canada. *Science*, n. s., **4,** 1896.
9. GILBERT, J. A. Researches on the mental and physical development of school children. *Studies from the Yale Psychological Laboratory*, **2,** 1894.
   Researches upon school children and college students. *University of Iowa Studies in Psychology*, **11,** 1897.
10. DE GIOVANNI, A. Commentari di clinica medica. Padova, 1889–1893.
    Morfologia del corpo umano. Milano, 1891.
    Commentari, Parte generale. Milano, 1904.
    Studi di morfologia clinica. Vol. 2. Milano 1905.

Commentari di clinica medica desunti dalla morfologia del corpo umano. Parte speciale. Vol I–II. Milano, 1907–1908.

Nuovi studi di morfologia clinica. Vol. IV. Milano, 1909.

11. Viola, G. La tecnica antropometrica a scopo clinico. Le dimensioni dell' uomo medio normale veneto. Padova, 1905.

L'indirizzo individualistico in medicina e il metodo morfologico del De Giovanni. Milano, 1905.

La tecnica antropometrica a scopo clinico: Criteri e metodo di comparazione antropometrica fra le individualità umane. Padova, 1905.

Le leggi di correlazione morfologica dei tipi individuali: La seriazione asimmetrica del "valore del tronco" e i suoi rapporti con la patologia. Padova, 1908.

Le leggi di correlazione morfologica dei tipi individuali: L'antropometria a scopo fisiognostico. Padova, 1908.

Le leggi di correlazione morfologica dei tipi individuali: L'abito tisico e l'abito apoplettico quali prodotti di una legge naturale di deformazione del tipo etnico. Padova, 1909.

Il metodo antropometrico "di deformazione" per la classificazione clinica dei tipi morfologici. Padova, 1909.

Le leggi di correlazione morfologica dei tipi individuali: L'abito tisico e l'abito apoplettico nei rapporti con l'infantilismo e la precocità. Roma, 1909.

12. Stratz, C. N. Naturgeschichte des Menschen. Stuttgart, 1904.

13. Ranke, J. Der Mensch. Vol II. Leipzig, 1894.

14. Schlesinger, F. *Zeitschrift für Kinderheilkunde*, 1920, **27**, 207.

15. Pende, N. Endocrinologia. 2*a* edizione. Milano, 1921.

16. Naççarati, S. The Oculocardiac Reflex (Dagnini-Aschner Phenomenon); its use in medicine and psychology. *Archives of Neurology and Psychiatry*, *5*, No. 1, 1921.

Psychological Examinations of College Students

BY

F. EDITH CAROTHERS, PH.D.

NEW YORK
1922

ACKNOWLEDGMENT

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

SECTION I

HISTORY OF FRESHMAN TESTS

The recommendation in 1882 by Galton [1] of the establishment of anthropometric and medico-metric laboratories for the examination of individuals represents the first definite recognition of the need of examining individuals in order to give them vocational guidance. Galton saw the importance both to science and to individuals of collecting complete life-histories of people which should include photographs, anthropometric measurements, and medical facts. To meet this need he established his now famous laboratory in the South Kensington Museum, London. There, by payment of a small fee, individuals could go and have certain physical measurements made and undergo tests for keenness of vision and hearing, dynamometer pressure, reaction time, etc.

Several years later, at the World's Columbia Exposition in 1893,[2] Professor Joseph Jastrow arranged a laboratory devoted to tests of a strictly psychological nature. Prior to Jastrow's work, however, Cattell proposed [3] and tried out a series of ten mental tests and measurements on students in the psychological laboratory of the University of Pennsylvania. In devising his series of tests Cattell followed Galton in combining physical measurements with psychophysical and strictly mental tests. He went a step farther, however, by emphasizing the necessity of standardizing methods of procedure in administering tests so that results secured by different experimenters might be comparable. In addition to the Pennsylvania students, tests were also given to the students of Cambridge University and Bryn Mawr College.

Galton's work stimulated other investigators to devise tests for measuring the capacities of individuals. Of particular interest is the list of ten fundamental traits or properties proposed by Kraepelin [4] as the basic factors to be considered in examining both normal individuals and the "mentally sick." These so-called fundamental dispositions include: the mental capacity to do work, the ability to

[1] Fortnightly Review, 1882, p. 332.

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

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

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

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

With the accumulation of data and the gradually increasing clearness of conception of the meaning of tests, methods of administering them were revised. In 1896 [5] appeared the first report of the results of mental and physical tests made on freshmen only. It concerned the work done by Professor Cattell and Dr. Farrand on one hundred Columbia University students in 1894–5 and 1895–6. At this time there was conceived the plan of testing Columbia students during their freshman and senior years. Their tests comprised ten records and twenty-six measurements. Such physical measurements were taken as the color of hair and eyes, height and weight, breathing capacity, sensation areas, and strength of right and left hands. Other measures were of a sensory character, while certain simple tests of a mental character were taken, such as the rate of perception and the perception of space and time. In addition, a personal record-blank was filled out by the student and a record of the impressions made upon him by the subject was filled in by the experimenter both before and after testing. The tests were given individually, the investigators and several assistants acting as experimenters, and required from forty minutes to one hour for their completion. The underlying purpose in giving these tests is clearly stated in this statement by Cattell and Farrand: [6]

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

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

[6] Above reference.

These Columbia freshman tests continued to be given each year under Professor Cattell's direction. In 1901 [7] an account and discussion of the results was published by Wissler. He discusses the changes and additions made in the tests and considers the records of 250 freshmen, a small number of seniors, and some Barnard girls. The tests employed were: length and breadth of head, strength of hands, fatigue, eyesight, color-vision, hearing, perception of pitch, perception of weight, sensation areas, sensitiveness to pain, perception of size, color preference, reaction time, rate of perception, naming colors, rate of movement, accuracy of movement, perception of time, association, imagery, memory, (auditory, visual, logical, and retrospective). Records of stature, weight, etc., together with data concerning parentage, personal habits, and health, the physical measurements taken in the gymnasium, and academic marks were also secured. From the similarity of the results of freshmen tested each year, Wissler concluded that freshmen entering Columbia from year to year are a homogeneous group and represent a type. His general conclusions are:

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

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

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

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

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

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

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

Various other proposals were made for the scientific study of the college student. In 1899 President Harper of Chicago recommended that special study be made of the college student's character, intellectual capacity, and tastes, by the questionnaire method. In 1906 Thorndike [8] called attention to the fact that the entrance examinations given by the College Entrance Board of the Association of Colleges and Preparatory Schools of the Middle States and Maryland did not measure at all accurately the candidate's capacity and emphasized the need of the scientific study of this matter. Williams [9] also stressed the importance of studying the college student. Like President Harper, he recommended the questionnaire method for ascertaining facts concerning the student's personality, and suggested the use of Whipple's information test for obtaining a knowledge of the student's range of information. He also pointed out the need of vocational advisors for freshmen.

Calfee [10] in 1913 reported the results of four general intelligence tests on 103 freshmen (51 boys and 52 girls) of the University of Texas. The tests used were card-dealing, card-sorting, alphabet-sorting, the mirror test, and the spirometer test for vital capacity. She finds inter-test correlations for the boys and girls combined ranging all the way from +.50 to .00. The correlations between the tests and college grades range from +.32 (card sorting and grades) to +.16 (mirror test and grades). The correlation between the lung test and grades is −.11. Considering the girls' records alone, the inter-test correlations range from +.45 to +.19, and the correlations with college grades from +.28 to +.13, and with the lung test the correlation is .00.

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

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

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

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

tion.[11] In 1912 he applied seven tests [12] individually to freshmen in Beloit College, and three years later, in 1915, he gave the same tests to thirty-nine of the same subjects. Waugh's inter-test correlations range from −.43 to +.54, and he finds some improvement in the tests from freshman to senior year.

During the year 1913–14 Bingham [13] gave nine tests to 200 Dartmouth freshmen, seven of them being given individually. As a number of psychology students, unpracticed experimenters, assisted Professor Bingham in his testing, the results of his investigation are somewhat inaccurate. He gives norms for the nine tests, (median, standard deviation and coefficient of variability) and the range from the poorest to the best. As no correlations are reported we have no information as to the relationships between the tests. Bingham's chief contribution consists in his use of the method of ogive percentile graphs. The data in seven of his tests are presented in this form, thus serving as a scale. Given the score made by any individual, the experimenter by reference to the chart can readily assign him a rank among his classmates. The speed with which a student may be thus assigned his relative position in any given trait makes this method a most convenient one for the instructor.[14]

At the University of Texas the same year Bell [15] gave nine tests [16] to about seven hundred and fifty freshmen. Bell definitely states that his aim was to devise a series of tests that would "be of assistance to college authorities in aiding freshmen to adjust themselves to their environment." The time required for testing was from forty to forty-five minutes. The tests were given not individually, but in groups averaging a little less than twenty each. The time-limit method was used. This, together with his arbitrary method of scoring the tests may account in some measure for the unsatisfactory nature of his results. He weighted each test so that a perfect

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

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

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

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

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

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

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

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

1. The correlations between freshmen university grades vary from + .34 (mathematics—history) to + .59 (English—history, science—history).

2. The highest correlation between class marks and test scores is + .31 (English—Completion).

3. Among the tests themselves the highest correlations are found between the Association and Recognition tests, and between the Directions, Alternatives and Completion tests.

4. There is a considerable difference in the results of the tests with the best and the poorest students, but the scores are so variable as to be of little value for individual diagnosis.

The investigations of Calfee, Waugh, Bingham, and Bell illustrate the striking change that has taken place in the character of mental tests since the early Columbia tests were first instituted. In place of sensory and motor tests we now employ tests which will measure diverse mental functions. Motivated by this same desire to secure a group of tests for college students indicative of mental ability, and correlative with college grades, Rowland and Lowden [17] began to try out groupings of psychological tests in 1912–13 and carried out their investigations over a period of three years. The tests were conducted individually on all the students in Reed College, twelve students of experimental psychology assisting in conducting the tests. The first grouping of tests was tried out on 54 students during 1912–13, after which the grouping was revised and given to 195 more subjects. No inter-test correlations are reported. The highest correlation between university grades and the groupings was between the grades and the letter-group g-r-s-t, cancellation, opposites, logical memory, judgment (syllogism), rote memory,

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

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

Psychological tests have also been conducted for several years at Vassar College. Results of tests made upon Vassar freshmen during the years 1914,[18] 1915, and 1916[19] show data collected from four sources, namely: 1. Answers to a questionnaire calling for information regarding the student's imagery, interests, language facility, and habits; 2. Results of the tests;[20] 3. Freshmen academic grades; 4. Reports of promising students by their instructors. To determine roughly the correlation between academic marks and test scores, the difference between the average class standing of students having test scores in the first or highest quarter and the average class standing of students with test scores in the last quarter was found. If there was a marked difference the experimenters concluded that a positive correlation existed. According to this rough method they found a positive correlation between academic marks and the tests except Hard Directions. On the whole, the results of the Vassar tests appeared to indicate that ability in the tests correlates well with ability in freshman studies, while inability to do well in the tests is correlated with a similar inability to do well in freshman studies. Moreover, students designated as "promising" by their instructors tend to manifest a high grade of performance in the tests. (14.5% of 317 freshmen tested in 1917 who passed all the tests in the Terman Superior Adult Tests were rated by their instructors as being of only average ability.) The experimenters also found that the relation between success in freshman tests and academic success in three years' work is less than that between success in freshman tests and academic success in the freshman year. Inasmuch as there were thirty different testers, each one being assigned a small group of freshmen, little confidence may be placed in the accuracy of the data. The tests as conducted at Vassar are of value more for the opportunity they afford students of psychology to acquire training in experimental methods of procedure than for any contribution they make to our knowledge of freshman standards of performance in various tests.

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

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

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

An interesting contribution in connection with the application of psychological tests to college freshmen is that of Kitson [21] at the University of Chicago. With the general purpose of devising a "system for measuring the mental capacity of college students in order to guide their college work," Kitson selected sixteen tests.[22] About half the tests were given by the group method. The time required for testing was two and one half hours. From forty complete records Kitson computed norms of performance in the various tests. In addition, a graphic chart was arranged for each student to show his standing in each test and to furnish a net score combining his standing in all the tests. In the particular tests used, Kitson found a significant positive correlation only between: 1. Memory for meaningful material seen and heard (+ .54); 2. Between the first and second reproductions of this material (+ .49); 3. Between the Opposites and Constant Increment tests (+ .40). When correlations were computed of standings in each test with standings in the net score, they were found to be somewhat higher. The correlation between college marks and psychological tests was found to be + .44 (P.E. .09) but from forty records secured from a second group of freshmen tested the correlation was found to be only + .20 (P.E. .11). Kitson explains this low correlation on the ground that many other factors besides intelligence enter in to determine standing in school studies, such as the personal factor of the instructor, the student's will power, social surroundings, economic conditions, and physical condition. The correlation between the psychological tests and intelligence as estimated by the dean was + .57 (P.E. .05). Twenty-one of the 1915 freshmen were retested in seven of the tests in their Sophomore year and improvement was shown in every test except one. (Numbers heard.) Comparison between the net score for freshman and sophomore year shows a correlation of + .88 (P.E. .03).

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

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

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

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

Other minor investigations have been made on freshmen with the same purpose. Sunne,[23] working at Newcomb College, found a low correlation between college grades and an information test tried on twenty-five freshmen, and with ninety-nine freshmen who were given a series of tests found correlations of tests with grades ranging from 0 to + .25. Haggerty [24] found a correlation of a quality of reading test and omnibus test with medical marks of + .62 and + .60, respectively, and of the two combined of + .65, in the case of sixty-nine candidates for medical school who had already completed two years of college.

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

A little later Irving King and James M'Crory [26] followed Kitson's method more definitely. In the fall of 1916 they tested 276 women and 268 men freshmen in seven different tests: the Courtis Standard

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

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

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

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

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

At Northwestern University Uhl[27] obtained inter-test correlations ranging from + .18 (Trabue Completion K and Information), to + .42 (Trabue Completion M and Information), for a group of one hundred freshmen tested in the fall of 1916. His series contained only four tests: Trabue Completion K and M, a hard opposites list of twenty words, and an information test which consisted of the seventy most familiar words in Whipple's list plus thirty new words. Test correlations with the first semester English and Mathematics grades were determined and found to range from + .48 (English and Mathematics), to + .16 (Completion K and Mathematics). When he had three mathematics instructors rate these one hundred students for ability, Uhl found a correlation of + .93 between their ratings and the Mathematics grades of the students. This high correlation was no doubt due to the tendency on the part of the teachers to make their judgments of the students practically equivalent to the students' course grades. The correlation between the instructor's judgments and the ranks of these same students in their last year of high school was + .59, and with all the tests combined was + .36. Uhl thinks his tests fail to measure accurately, the information test being the most unsatisfactory, and attributes his low correlations to the homogeneity of his group, the relative simplicity of the tests, and the unreliability of school marks.

Thurstone's[28] work represents a further development in the use of psychological tests. At the Carnegie Institute of Technology the attempt is made to use psychological tests as a criterion for admission. A series of six mental tests was given to 114 freshmen of the Margaret Morrison Carnegie School in October, 1917. The problem

[27] Uhl, W. L. Mentality Tests for College Freshmen, Journ. of Educ. Psychol., 1919, 10, 13–28

[28] Thurstone, L. L. Journ of Educ. Psychol., March, 1919.

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

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

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

The past three years have brought a further development in the use of psychological tests for measuring the intelligence of college freshmen. Since 1918 the Army Alpha test has been administered to freshmen in several colleges with varying degrees of success. Professor Stone [29] reports that its use at Dartmouth justifies the recent proposal to admit students scholastically in the upper quarter of their class in approved schools. Strictly speaking, the work at Dartmouth should not be included in this history, since it deals with the results obtained in testing all the college classes rather than freshmen only. We mention it here, however, because the college authorities are now devoting particular attention to admin-

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

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

Similar to this Dartmouth study is that of Walcott's [30] at Hamline University. Here, too, not freshmen alone, but all students were given the Alpha test in the fall of 1918. Walcott's results are based on data secured from 61 men and 145 women. As in the Dartmouth investigation, a far greater proportion of men and women students secure a score in Alpha in the high grade intelligence group than was found in any of the army camps. The median score is 129 for the Hamline men and 133 for the women, with the same sharp differentiation between the poor and the good groups as Stone found at Dartmouth. The correlation between the results of the Alpha test for the women and their first term academic grades was + .47, slightly higher than the Dartmouth result. Although Walcott does not consider the army test the best device for determining the fitness of students for college work, he sees in the significant difference in score between the upper and lower half of the students tested, the practical use to be made of this fact in the placing of students.

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

[30] Walcott, G. D. "Mental Testing at Hamline University." School and Society, 1919, 10, 57–60.

[31] Filler, M. G. A Psychological Test. School & Society, 1919, 10, 208–209.

[32] Hill, D. S. Results of Intelligence Tests at the University of Illinois; School & Society, 1919, 9, 542–545.

Southern Methodist University[33] the effort was made to secure select groups of students in order to compare their scores with the average score for the school. Each student was asked to name men and women students whom they thought would make high scores in the Alpha test. For 16 men and 8 women named by from five to forty students as being able to make the highest scores, the average score for the men was 154, and for the women 156, justifying the judgment of the students. With a similar group of students named by the faculty as being able to make the highest scores even better results were obtained, the average for the men being 161 and for the women 167. In selecting a group of men and women whom they judged would make low scores the faculty were equally successful. Both faculty and students thus showed themselves fairly good in their ability to select students on the basis of intelligence, though this method of selection is inferior to selection on the basis of actual scores. The correlation between the Alpha scores of the women students and their college grades for the fall term was + .52. No correlations are given in the Illinois and Dickinson reports, which are only preliminary.

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

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

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

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

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

[34] Colvin, S. S. Psychological Tests at Brown University; School & Society, 1919, 10, 27–30.

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

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

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

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

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

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

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

The Thorndike tests, previously mentioned, are rapidly becoming more widely employed for freshmen testing than the Army Alpha. Jones,[26] writing in the Educational Review, clearly describes the general nature of these tests. Although conceding their practical value, he urges that they should be employed "not to the exclusion of other measures for determining fitness, but along with them." Evidence of a student's fitness to undertake college work should, in Professor Jones' opinion, include his preparation for college work, his character and promise, his health, and his intelligence denoted by his score in the Thorndike test. In a brief report before the New York Branch of the American Psychological Association this year Mr. Wood stated that the purpose of the Thorndike tests was fourfold: 1. To select those fit for a college course; 2. To aid college committees; 3. To assist the progress of schools; 4. To

[26] Jones, A. L. Psychological Tests for College Admission; Educational Review, 1919, 58, 271–278.

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

SECTION II

STATEMENT OF THE PROBLEM WITH A LIST OF THE TESTS EMPLOYED

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

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

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

SECTION III

METHOD AND TECHNIQUE OF THE INVESTIGATION

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

The subjects, as indicated, were Barnard students in their freshman year. The fact that they had had no training in experimental psychology, and were unfamiliar with the tests employed, made them a suitable group for testing. Out of a class of about one hundred and forty freshmen during 1915–16, one hundred were tested. This constitutes our first group of subjects whom we will designate as Group I. During the year 1916–17 (class of 1920), eighty-five freshmen were tested, and in the fall of 1917 fifteen more (class of 1921) were given the tests. These last two groups together constitute our second group of one hundred freshmen whom we will designate Group II. In addition, in order to determine the reliability of the tests, the series was divided into two equivalent parts in a manner to be described later. In the spring of 1919, during the period extending from March 14 to May 15, forty-five freshmen from the class of 1922 were tested twice on the same day, each test requiring forty-five minutes of the student's time.

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

As previously indicated, attempt to secure uniformity in administering the tests was also made. Besides giving the tests individually, the order in which the tests are listed was followed. In a few cases circumstances rendered it necessary to deviate slightly from this order, but in general it was followed rigidly. The result of the preliminary trial had been to indicate the most satisfactory manner in which the tests should be administered. The aim was to make the directions as clear, simple, and direct as possible. As a detailed account of the instructions given for each test will be considered in the next section, it is only necessary to mention here that the method of procedure agreed upon was carefully followed with one or two exceptions where misinterpretation of the directions resulted in the experimenter's repeating the instructions in a slightly different form.

SECTION IV

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

Test No. 1. Coordination

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

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

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

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

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

TABLE I

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

Test No. 2. Tapping

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

Apparatus: Tapping board with metal plate and electric counter.

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

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

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

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

TABLE II

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

Test No. 3. Cancellation

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

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

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

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

Results: Table III indicates the performance in this test.

TABLE III

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

Test No. 4. Checking

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

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

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

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

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

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

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

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

TABLE IV

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

Test No. 5. Color Naming

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

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

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

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

Results: The results are shown in Table V.

TABLE V

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

[39] Op. cit.

Test No. 6. Directions

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

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

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

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

Results: Table VI indicates the performance in this test.

TABLE VI

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

Test No. 7. Opposites

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

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

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

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

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

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

TABLE VII

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

Test No. 8. Verb-object

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

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

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

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

Results: The results are indicated in Table VIII.

TABLE VIII

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

Test No. 9. Mixed Relations or Analogies

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

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

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

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

Results: The results are shown in Table IX.

TABLE IX

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

Test No. 10. Word Building

For a test that would indicate ingenuity and skill in the manipulation of letters and give a measure of the subject's command of vocabulary, the word building test was used. The number of words written in a given time depends in part on whether the subject proceeds with a definite plan, combining, for example, "a" with all the other letters, then "e" with all the other letters, etc., or goes about the task in a vague or random fashion.

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

Method of procedure: The procedure as given by Whipple [40] was followed with the exception that the time-limit was three minutes instead of five.

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

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

TABLE X

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

Test No. 11. Word Naming

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

Materials: Stop watch. Sheet of paper and pencil.

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

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

Results: Table XI shows the results for this test.

Test 12. Knox Cube

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

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

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

TABLE XI

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

Materials: Five one-inch cubes.

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

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

Results: Results are indicated in Table XII.

TABLE XII

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

Test No. 13. Digit Span

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

Materials: Digit Span blank. Stop watch.

Method of procedure: These instructions were given: "I am going to read some numbers to you and as soon as I have finished saying them, I want you to repeat them in exactly the same order." The smallest number of digits given was five. Three trials were given for each number. The attempt was made to repeat the numbers without rhythm.

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

Results: Table XIII indicates the results of this test.

TABLE XIII

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

Test No. 14. Word Memory

Test No. 15. Logical Memory

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

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

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

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

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

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

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

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

TABLE XIV

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

TABLE XV

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

Test No. 16. Substitution

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

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

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

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

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

TABLE XVI

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

Test No. 17. Completion

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

Materials: Trabue Language Scale A. Stop watch.

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

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

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

TABLE XVII

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

[42] Trabue. Completion-Test Language Scales.

Test No. 18. Information

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

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

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

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

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

TABLE XVIII

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

Test No. 19. Vocabulary

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

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

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

[43] Op. cit. Vol. 2, p. 310.

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

Results: Table XIX shows the results for this test.

TABLE XIX

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

SECTION V

NORMS OF PERFORMANCE AND THEIR PRACTICAL APPLICATION

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

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

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

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

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

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

TABLE XX

NORMS AND STANDARDS OF PERFORMANCE (200 Cases)

Name of Test	Average Group I	Average Group II	Average Groups I and II	Probable Error	Range: Poorest; (Average of poorest 5)	Range: Best (Average of best 5)
1. Coordination	82.7	84.1	83.42 strokes	7.5	60.00	114.20
2. Tapping	376.26	368.54	372.4 taps	27.6	251.40	501.20
3. Cancellation	76.51	76.77	76.64 sec.	10.35	128.28	48.56
4. Checking	102.93	105.98	104.46	13.10	169.16	72.08
5. Color Naming	56.01	58.55	57.28	5.28	84.72	38.76
6. Directions	126.15	119.76	122.95	24.95	312.00	59.16
7. Opposites	51.08	50.88	50.98	6.01	79.80	33.92
8. Verb-object	65.55	67.35	66.45	7.54	107.80	44.96
9. Mixed Relations	139.64	131.66	135.65	21.34	279.00	77.96
10. Word Building	16.33	16.23	16.28 words	2.71	4.80 words	27.60
11. Word Naming	67.14	67.87	67.50	7.50	39.40	97.20
12. Knox Cube	9.20	8.82	9.01	1.01	4.80	12.00
13. Digit Span	7.39	7.67	7.53	.53	5. digits	10.40
14. Word Memory—Recollection	11.59	10.91	11.25 words	1.74	5.8 "	18.60
15. Word Memory—Recognition	35.84	35.07	35.45	5.45	13.60	48.80
16. Logical Memory—Recollection	6.19	6.50	6.34	1.34	2.20	10.20
17. Logical Memory—Recognition	36.75	37.47	37.11	5.11	13.20	48.40
18. Substitution 1st Half	64.33	66.68	65.51	7.51	101.28 sec.	43.80
19. Substitution 2nd Half	59.10	61.51	60.32	8.12	98.40 sec.	35.20
20. Substitution Whole	123.09	128.19	125.64	14.40	196.60 sec.	83.20
21. Completion	36.08	35.78	35.93	2.93	23.80	46.00
22. Information	106.63	104.71	105.66	18.33	54.20	168.20
23. Vocabulary	74.81	73.90	74.35	5.35	57.40	89.00

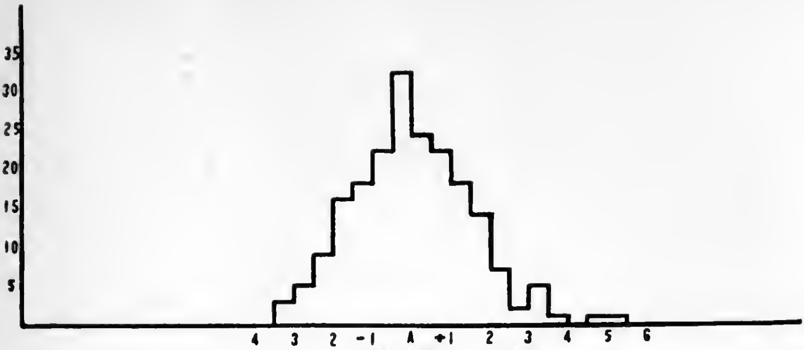


FIG. 1 COORDINATION

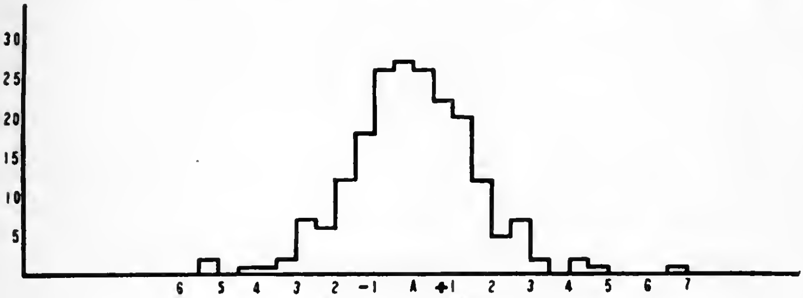


FIG. 2 TAPPING

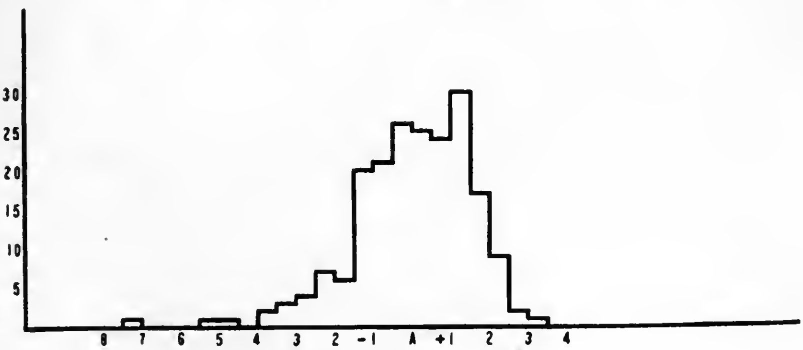


FIG. 3 CANCELLATION

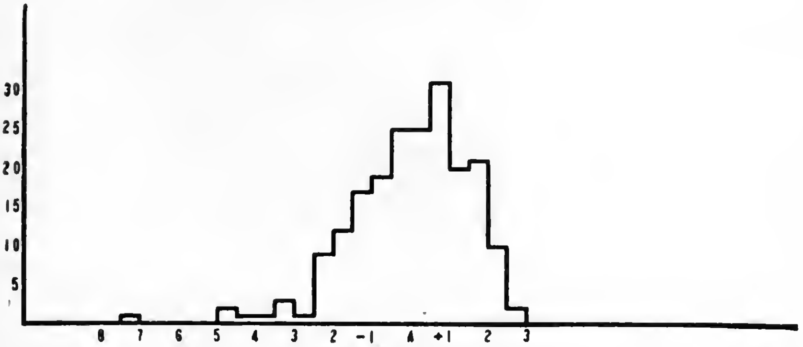


FIG. 4 NUMBER CHECKING

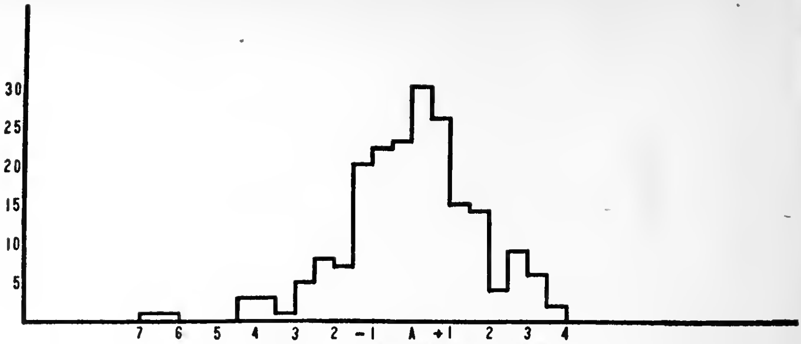


FIG. 5 COLOR NAMING

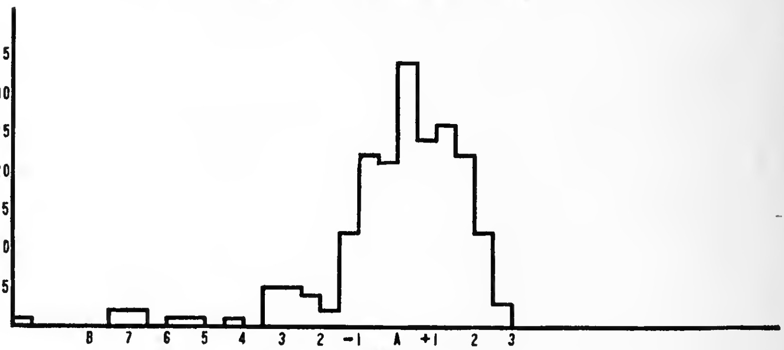


FIG. 6 DIRECTIONS

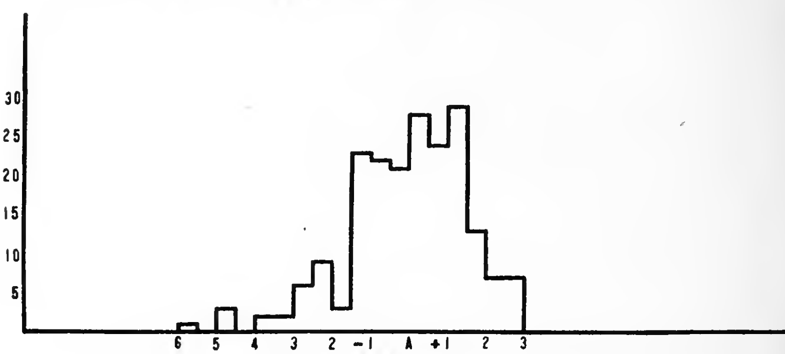


FIG. 7 OPPOSITES

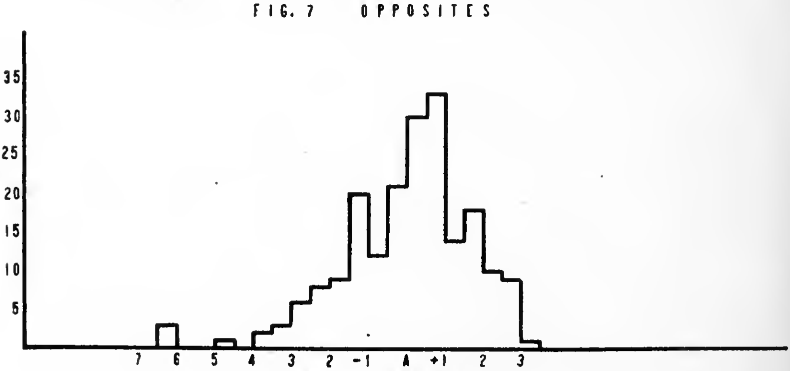


FIG. 8 VERB OBJECT

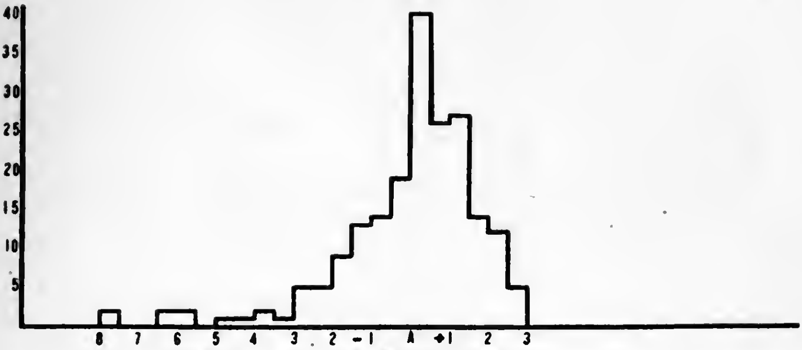


FIG. 9 MIXED RELATIONS

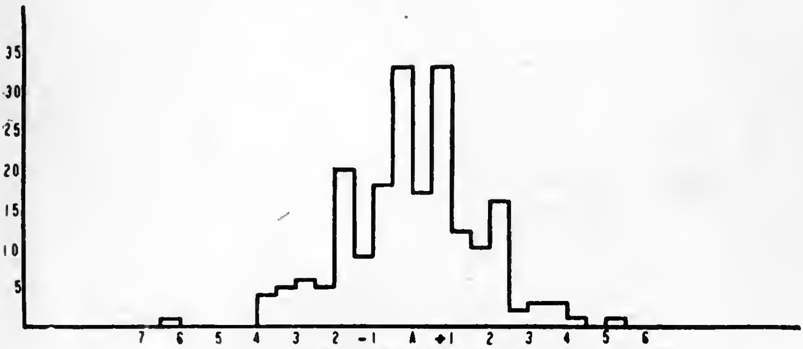


FIG. 10 WORD BUILDING

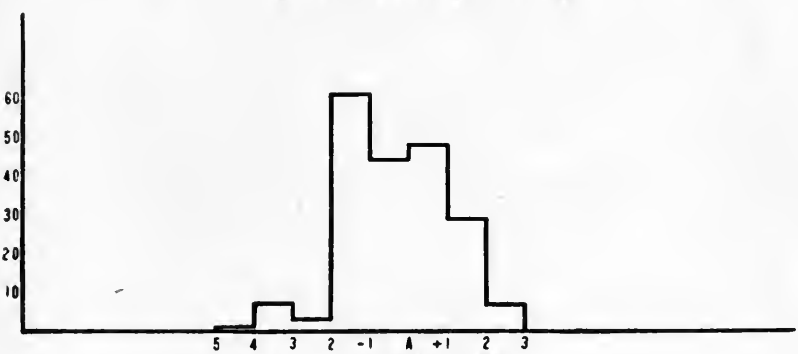


FIG. 12 KNOX CUBE

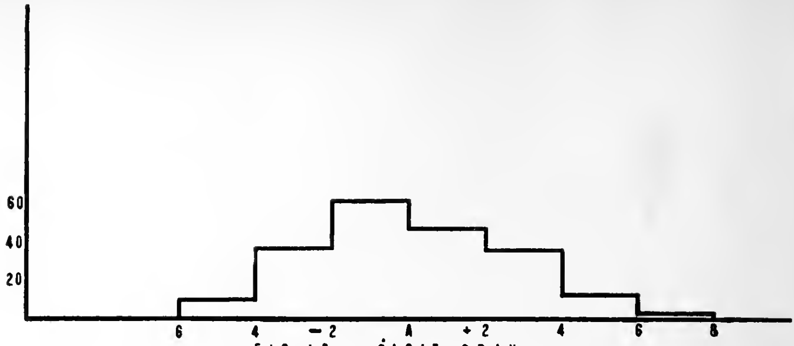


FIG. 13 DIGIT SPAN

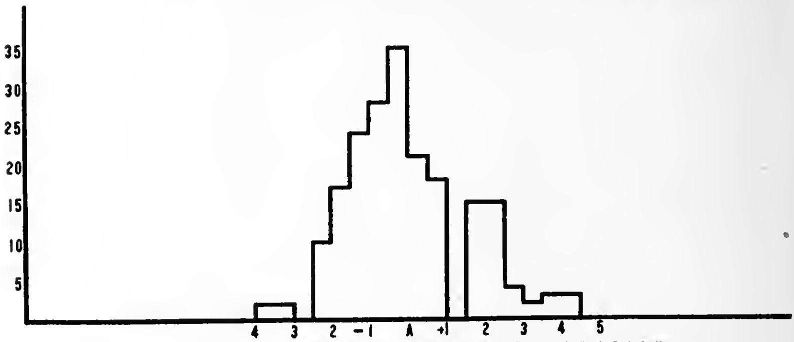


FIG. 14 WORD MEMORY RECOLLECTION

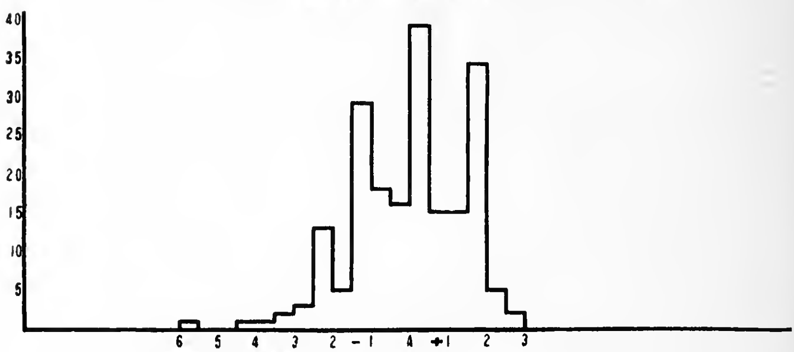


FIG. 15 WORD RECOGNITION

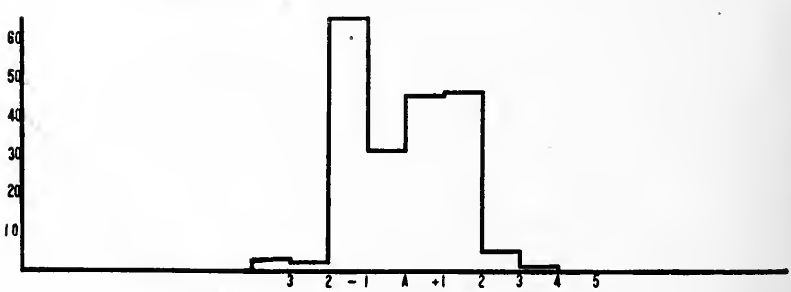


FIG 16 LOGICAL MEMORY–RECOLLECTION

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

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

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

The fourth group contains those tests which call into play powers

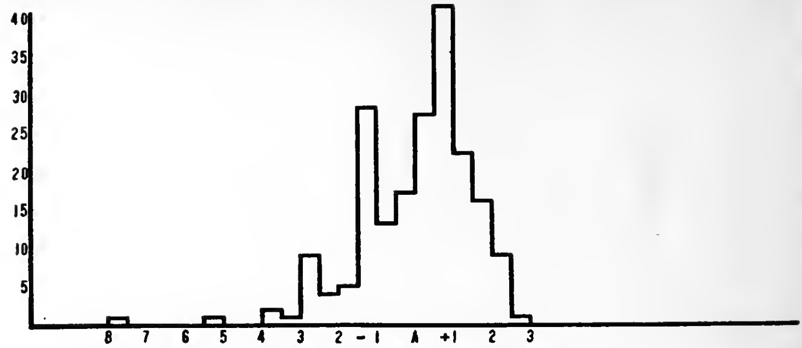


FIG. 17 LOGICAL RECOGNITION

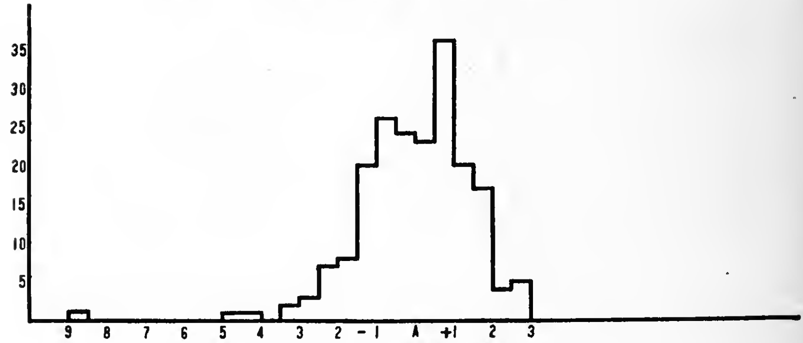


FIG. 18 SUBSTITUTION–FIRST HALF

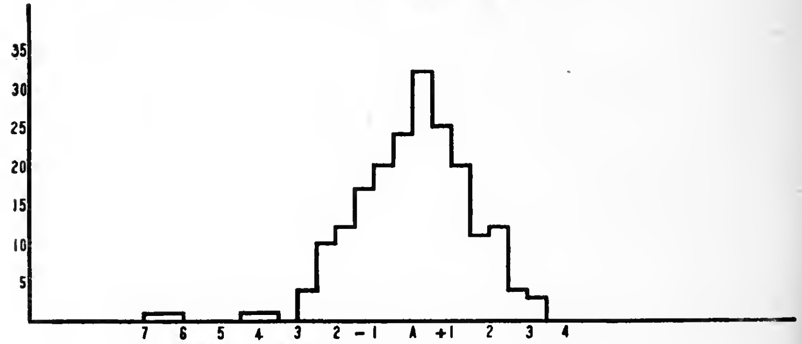


FIG. 19 SUBSTITUTION–SECOND HALF

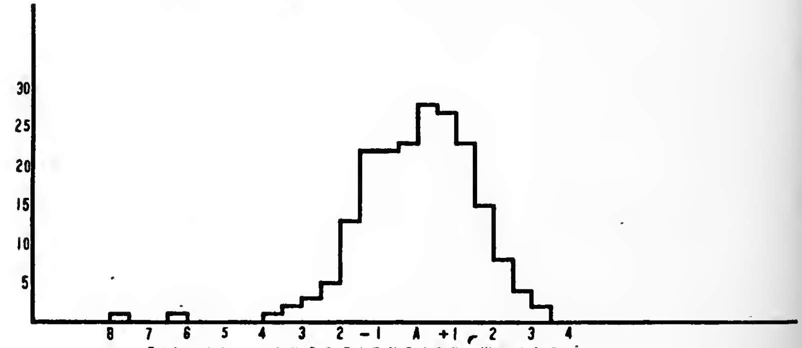


FIG. 20 SUBSTITUTION WHOLE

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

A word of explanation is needed here regarding the construction of the chart for Logical Memory (Recollection). The categories into which the scores fall are so few that the finest grouping possible is in 1 P.E. units instead of ½ P.E. units as in the other tests. As we said before, to secure uniformity we let the P.E. represent the same interval along the base line in all tests. Now, in order to keep the area of a given number of cases constant for all tests, it is necessary where we have scores in terms of 1 P.E. units to reduce the vertical scale proportionately. Therefore, we regard the measures as distributed evenly over the P.E. intervals and reduce the vertical scale one-half. In this test and in Word Recollection we find a greater number of cases below the average than above. The curve is skewed toward the poor end in Word Recollection, and toward the good end in Word Recognition and Logical Recognition.

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

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

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

Academic Grades

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

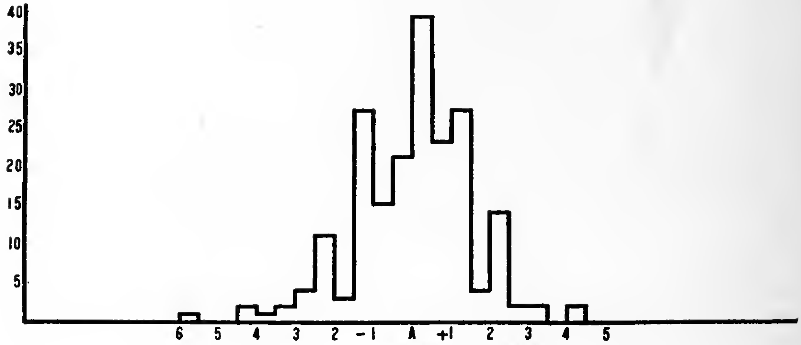


FIG. 21 COMPLETION

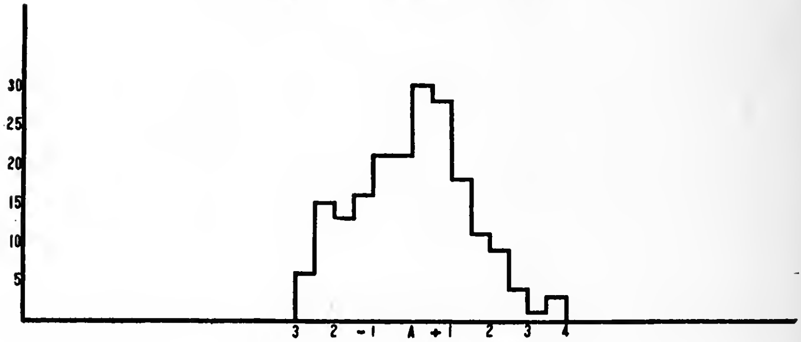


FIG. 22 INFORMATION

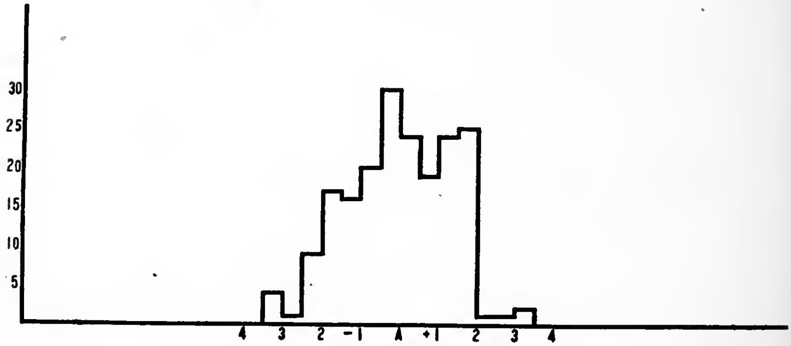


FIG. 23 VOCABULARY

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

TABLE XXI

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

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

Physical Measurements

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

TABLE XXII

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

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

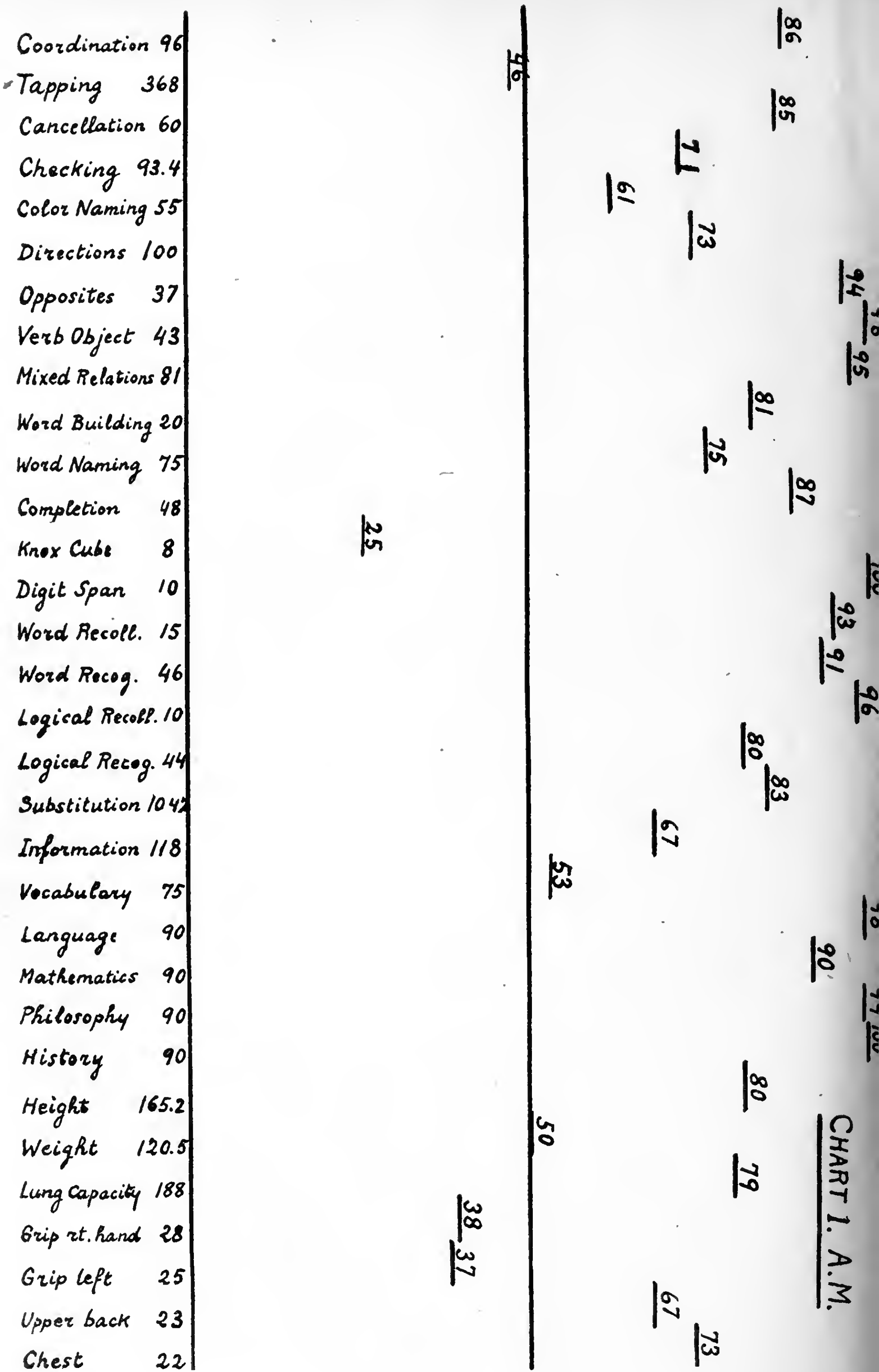

Coordination 96
Tapping 368
Cancellation 60
Checking 93.4
Color Naming 55
Directions 100
Opposites 37
Verb Object 43
Mixed Relations 81
Word Building 20
Word Naming 75
Completion 48
Knox Cube 8
Digit Span 10
Word Recoll. 15
Word Recog. 46
Logical Recoll. 10
Logical Recog. 44
Substitution 1042
Information 118
Vocabulary 75
Language 90
Mathematics 90
Philosophy 90
History 90
Height 165.2
Weight 120.5
Lung Capacity 188
Grip rt. hand 28
Grip left 25
Upper back 23
Chest 22

CHART 1. A.M.

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

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

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

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

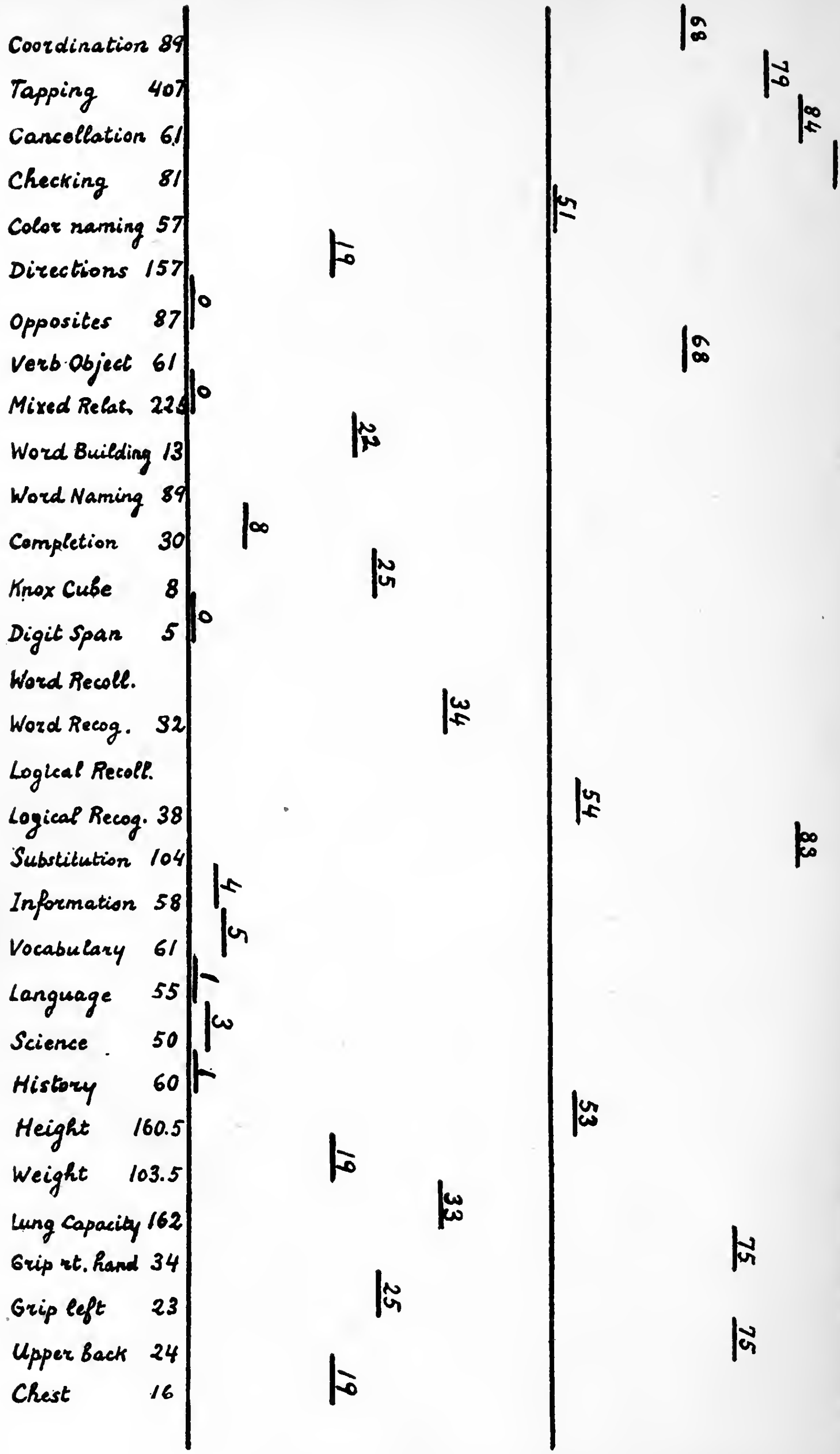

Coordination 89
Tapping 407
Cancellation 61
Checking 81
Color naming 57
Directions 157
Opposites 87
Verb-Object 61
Mixed Relat. 225
Word Building 13
Word Naming 89
Completion 30
Knox Cube 8
Digit Span 5
Word Recoll.
Word Recog. 32
Logical Recoll.
Logical Recog. 38
Substitution 104
Information 58
Vocabulary 61
Language 55
Science 50
History 60
Height 160.5
Weight 103.5
Lung Capacity 162
Grip rt. hand 34
Grip left 23
Upper Back 24
Chest 16
68
79
84
88
51
19
0
68
0
22
8
25
0
34
54
83
4
5
1
3
1
53
19
33
75
25
75
19

CHART 2. L.H.C.

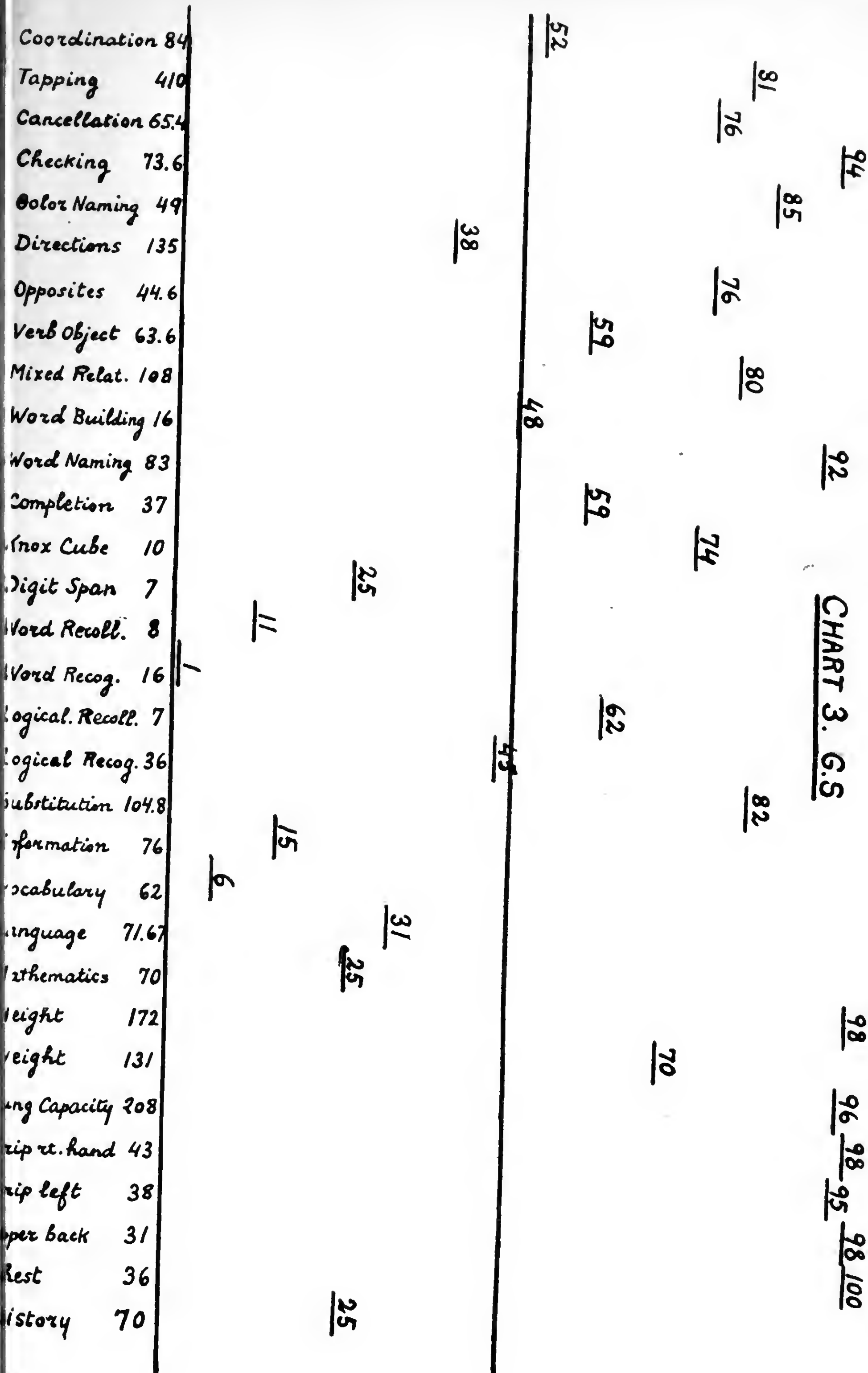


CHART 3. G.S
Coordination 84
Tapping 410
Cancellation 65.4
Checking 73.6
Color Naming 49
Directions 135
Opposites 44.6
Verb Object 63.6
Mixed Relat. 108
Word Building 16
Word Naming 83
Completion 37
Knox Cube 10
Digit Span 7
Word Recoll. 8
Word Recog. 16
Logical. Recoll. 7
Logical Recog. 36
Substitution 104.8
Information 76
Vocabulary 62
Language 71.67
Arithmetics 70
Height 172
Weight 131
Lung Capacity 208
Grip rt. hand 43
Grip left 38
Upper back 31
Chest 36
History 70
52
91
76
94
85
38
76
59
80
48
92
59
74
25
11
1
62
43
82
15
6
31
25
98
70
96
98
95
98
100
25

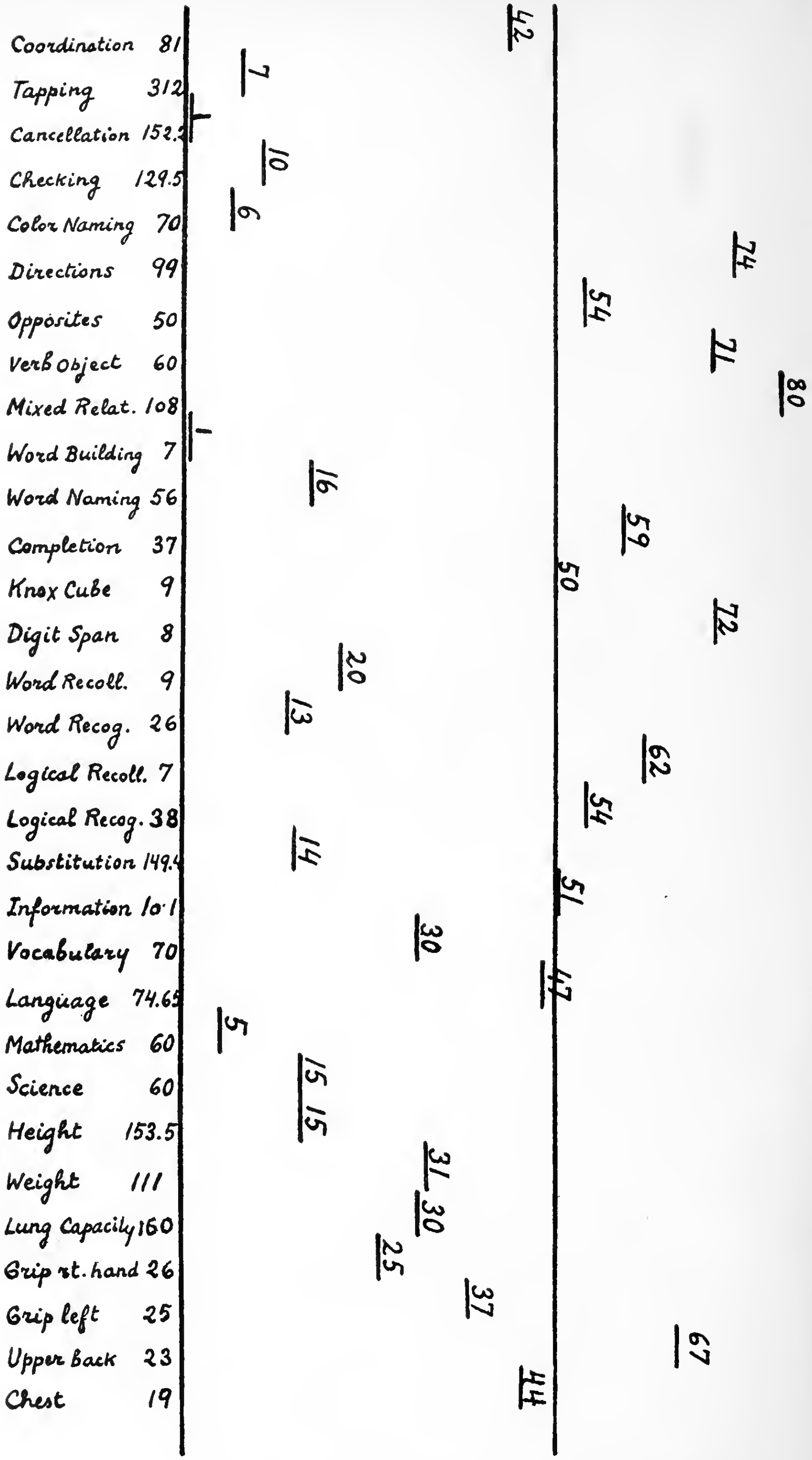

Coordination 81
Tapping 312
Cancellation 152.2
Checking 129.5
Color Naming 70
Directions 99
Opposites 50
Verb Object 60
Mixed Relat. 108
Word Building 7
Word Naming 56
Completion 37
Knox Cube 9
Digit Span 8
Word Recoll. 9
Word Recog. 26
Logical Recoll. 7
Logical Recog. 38
Substitution 149.4
Information 10·1
Vocabulary 70
Language 74.65
Mathematics 60
Science 60
Height 153.5
Weight 111
Lung Capacity 160
Grip rt. hand 26
Grip left 25
Upper Back 23
Chest 19

CHART 4. S.E.

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

[45] Thorndike, E. L. Mental and Social Measurements.

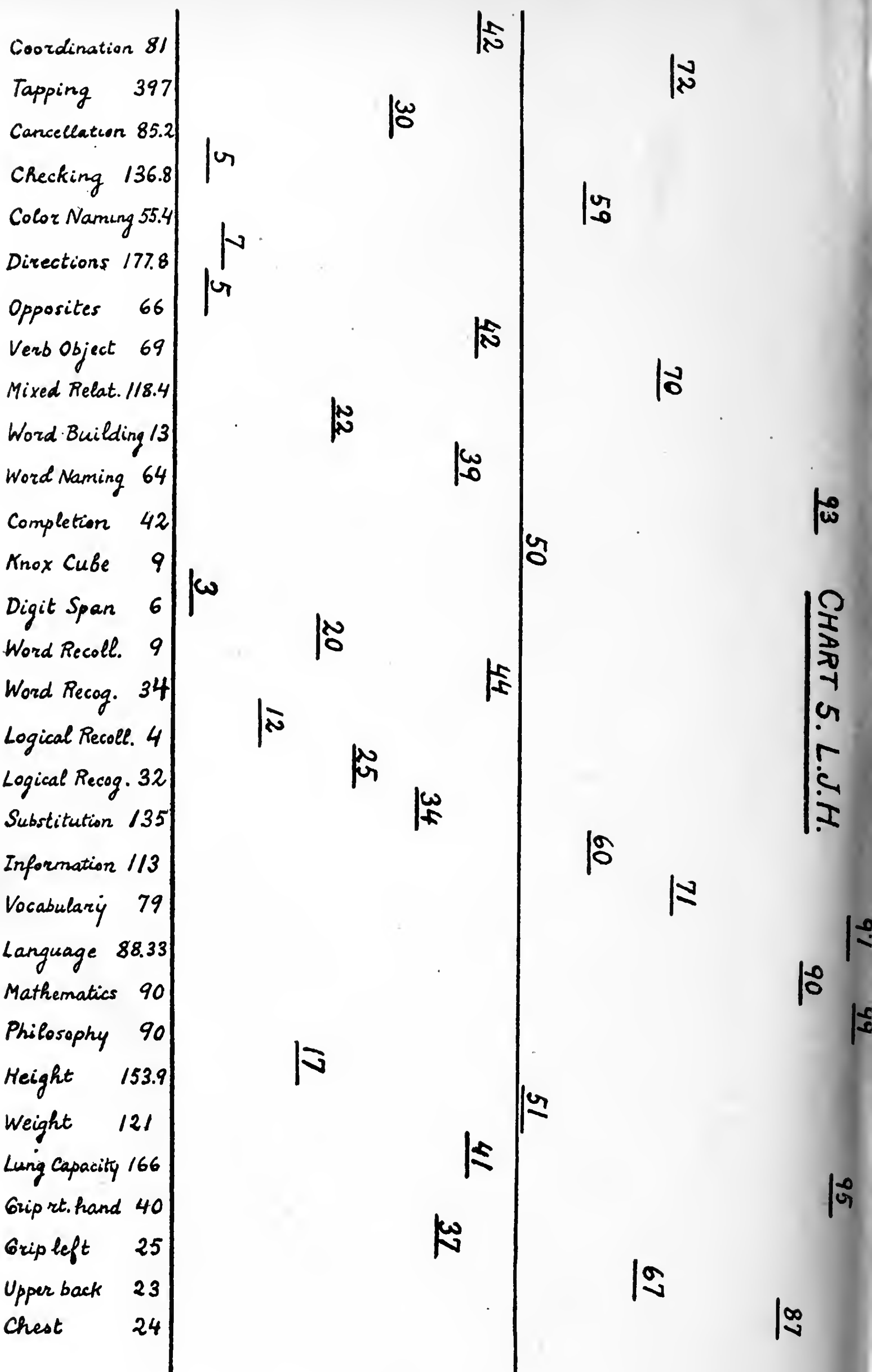


CHART 5. L.J.H.
Coordination 81
Tapping 397
Cancellation 85.2
Checking 136.8
Color Naming 55.4
Directions 177.8
Opposites 66
Verb Object 69
Mixed Relat. 118.4
Word Building 13
Word Naming 64
Completion 42
Knox Cube 9
Digit Span 6
Word Recoll. 9
Word Recog. 34
Logical Recoll. 4
Logical Recog. 32
Substitution 135
Information 113
Vocabulary 79
Language 88.33
Mathematics 90
Philosophy 90
Height 153.9
Weight 121
Lung Capacity 166
Grip rt. hand 40
Grip left 25
Upper back 23
Chest 24
42
72
30
5
59
7
5
42
70
22
39
93
50
3
20
44
12
25
34
60
71
90
17
51
41
95
37
67
87

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

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

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

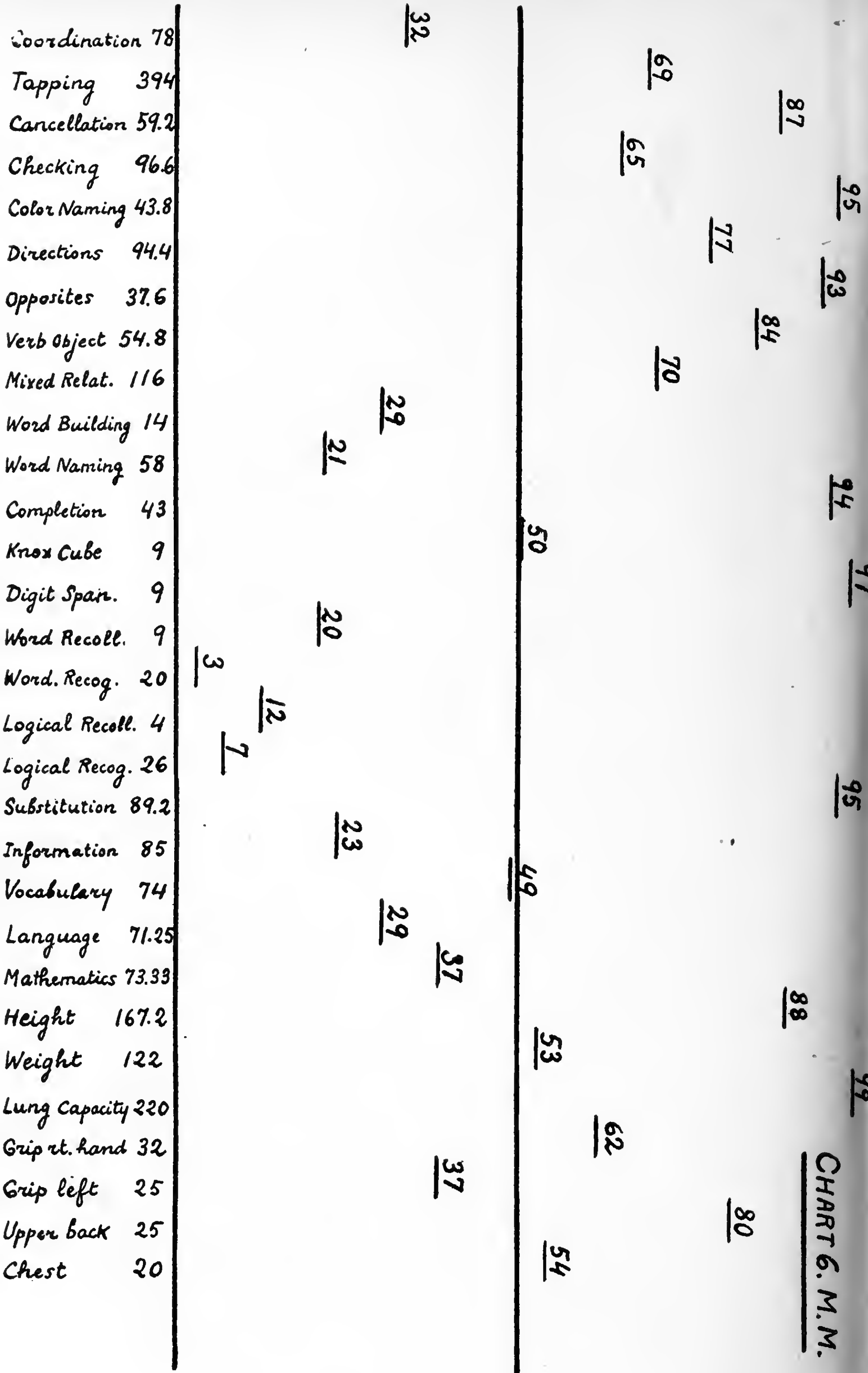


CHART 6. M.M.
Coordination 78
Tapping 394
Cancellation 59.2
Checking 96.6
Color Naming 43.8
Directions 94.4
Opposites 37.6
Verb Object 54.8
Mixed Relat. 116
Word Building 14
Word Naming 58
Completion 43
Knox Cube 9
Digit Span. 9
Word Recoll. 9
Word. Recog. 20
Logical Recoll. 4
Logical Recog. 26
Substitution 89.2
Information 85
Vocabulary 74
Language 71.25
Mathematics 73.33
Height 167.2
Weight 122
Lung Capacity 220
Grip rt. hand 32
Grip left 25
Upper back 25
Chest 20
32
69
87
65
95
77
93
84
70
29
21
94
50
97
20
3
12
7
95
23
49
29
37
88
53
99
62
37
80
54

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

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

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

SECTION VI

INTER-TEST CORRELATIONS AND THEIR SIGNIFICANCE

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

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

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

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

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

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

TABLE XXIII

	Coordination	Tapping	Cancellation	Checking	Color Naming	Directions	Opposites	Verb-object	Mixed Relations	Word Building	Word Naming	Knox Cube	Digit Span	Word Memory—Recollection	Word Memory—Recognition	Logical Memory—Recollection	Logical Memory—Recognition	Substitution Whole	Completion	Vocabulary
1. Coordination		+.48	+.30	+.30	+.10	+.01	+.08	+.13	−.11	+.11	+.27	+.12	+.10	+.12	+.20	+.27	+.17	+.25	−.01	−.19
2. Tapping	+.48		+.18	+.21	+.33	+.21	+.08	+.18	−.01	+.09	+.31	+.14	+.25	+.00	+.12	+.01	−.10	+.21	+.07	−.25
3. Cancellation	+.30	+.18		+.51	+.17	+.09	+.12	+.12	+.03	+.30	+.40	+.17	+.77	+.06	+.03	+.07	+.08	+.31	+.09	+.03
4. Checking	+.30	+.21	+.51		+.29	+.29	+.25	+.30	+.16	+.26	+.35	+.27	+.10	−.07	−.04	+.17	+.19	+.27	+.18	−.09
5. Color Naming	+.10	+.33	+.17	+.29		+.26	+.30	+.38	+.20	+.21	+.30	+.19	+.25	+.13	+.07	+.08	−.06	+.32	+.22	+.04
6. Directions	+.01	+.21	+.09	+.29	+.26		+.28	+.38	+.35	+.09	+.19	+.20	+.14	+.12	+.11	+.19	+.34	+.12	+.17	+.04
7. Opposites	+.08	+.08	+.12	+.25	+.30	+.28		+.44	+.57	+.34	+.17	+.17	+.18	+.06	+.02	+.19	+.15	+.21	+.34	+.07
8. Verb-object	+.13	+.18	+.12	+.30	+.38	+.38	+.44		+.38	+.23	+.32	+.13	+.07	+.05	+.08	+.11	+.09	+.27	+.13	+.07
9. Mixed Relations	−.11	−.01	+.03	+.16	+.20	+.35	+.57	+.38		+.23	+.22	+.25	+.18	+.00	−.04	+.25	+.17	+.02	+.48	+.08
10. Word Building	+.11	+.09	+.30	+.26	+.21	+.09	+.34	+.23	+.23		+.35	+.09	+.09	+.16	+.08	+.30	−.02	+.24	+.38	+.24
11. Word Naming	+.27	+.31	+.40	+.35	+.30	+.19	+.17	+.32	+.22	+.35		+.15	+.14	+.18	+.12	+.21	+.11	+.29	+.29	+.13
12. Knox Cube	+.12	+.14	+.17	+.27	+.19	+.20	+.17	+.13	+.25	+.09	+.15		+.15	−.13	+.07	+.24	+.26	+.18	+.24	+.06
13. Digit Span	+.10	+.25	+.77	+.10	+.25	+.14	+.18	+.07	+.18	+.09	+.14	+.15		+.09	+.13	+.07	−.03	+.08	+.22	−.06
14. Word Memory—Recollection	+.12	+.00	+.06	−.07	+.13	+.12	+.06	+.05	+.00	+.16	+.18	−.13	+.09		+.58	+.35	+.22	+.22	+.12	+.14
15. Word Memory—Recognition	+.20	+.12	+.03	−.04	+.07	+.11	+.02	+.08	−.04	+.08	+.12	+.07	+.13	+.58		+.28	+.26	+.29	−.02	+.04
16. Logical Memory—Recollection	+.27	+.01	+.07	+.17	+.08	+.19	+.19	+.11	+.25	+.30	+.21	+.24	+.07	+.35	+.28		+.56	+.19	+.21	−.21
17. Logical Memory—Recognition	+.17	−.10	+.08	+.19	−.06	+.34	+.15	+.09	+.17	−.02	+.11	+.26	−.03	+.22	+.26	+.56		+.14	+.16	−.15
18. Substitution Whole	+.25	+.21	+.31	+.27	+.32	+.12	+.21	+.27	+.02	+.24	+.29	+.18	+.08	+.22	+.29	+.19	+.14		+.15	−.01
19. Completion	−.01	+.07	+.09	+.18	+.22	+.17	+.34	+.13	+.48	+.38	+.29	+.24	+.22	+.12	−.02	+.21	+.16	+.15		+.26
20. Vocabulary	−.19	−.25	+.03	−.09	+.04	+.04	+.07	+.07	+.08	+.24	+.13	+.06	−.06	+.14	+.04	−.21	−.15	−.01	+.26	
Average correlation between each test and all others (omitting Vocabulary and Information)	.16	.15	.21	.22	.21	.20	.22	.21	.19	.20	.24	.16	.17	.13	.13	.21	.16	.21	.19	.01
Average correlation each test and others (omitting Vocabulary, Information and Word Memory Recollection and Recognition	.16	.17	.23	.25	.22	.21	.24	.23	.21	.21	.25	.18	.17			.20	.14	.20	.21	.00

NOTE: P. E. of a correlation of .20 for 100 cases is .0648. All coefficient of correlation that are .20 or over are reliable.

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

TABLE XXIV

GROUPING OF TESTS ON THE BASIS OF THEIR CORRELATION COEFFICIENTS

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

Thus Tapping and Coordination correlate + .48 with each other, but both tests show a much lower correlation with the other tests. (The correlations outside of the group range from + .33 to + .01). This agrees with Thorndike's theory that tests of the motor sensory level correlate rather closely with each other, but only loosely with tests of other levels. In Group II, Checking has an average correlation of + .36 with the others of the group, and also a much lower correlation with tests outside Group II, (ranging from + .30 to − .04). Similarly, in Group III, Opposites and Mixed Relations both have an average correlation of + .40 with the other tests in this group, but a lower correlation with any test outside the group, again conforming to Thorndike's contention that tests on the associative level correlate closely with each other, but rather loosely with tests on other levels. (The average correlation of Opposites with the tests outside Group III is + .15; the average correlation of Mixed Relations with tests outside Group III is + .10). In Group IV, also, Logical Recollection has an average correlation of + .40 with the other tests in the group, but a lower correlation with any test outside this group. (The correlations outside the group run from + .30 to + .01). Information and Vocabulary differ from the other tests of the series in that they are indicative of one's learning rather than one's innate ability. There is only a chance correlation between them and the other tests. A more detailed discussion of this relationship we will postpone till the following section. As for Knox Cube and Digit Span, perhaps the best plan is to consign them to the miscellaneous class. Knox Cube shows on the whole the closest correlations with the tests in Group II, but the average group correlation is not high enough to warrant us definitely placing it in this group rather than in Group IV. In like manner, aside from its surprisingly high correlation with Cancellation (+ .77), Digit Span shows no close relationship with any other test. If we omit these four tests, (namely, Information, Vocabulary, Knox Cube, and Digit Span), we do get very definite groupings of the other tests, as shown in Table XXII above, indicating that we are measuring different abilities. The rather high intercorrelations between the tests of each group, together with their low correlations with tests outside their own groups would support this view. There is no evidence from these results to support Spearman's theory that correlations are produced between all sorts of performance, the amount of the correlation being simply proportional to the extent that the performances concerned involve the use of a general common factor or "general ability." Our data give evidence neither of a common factor nor of a hierarchial arrangement of the correlations. Attempts to arrange the correlations to form a hierarchy met with even greater failure than Simpson reports.

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

TABLE XXV

Inter-test Correlations Corrected for Attenuation

	Coordination	Tapping	Cancellation	Checking	Color Naming	Directions	Opposites	Verb-object	Mixed Relations	Word Building	Word Naming	Knox Cube	Digit Span	Word Memory—Recollection	Word Memory—Recognition	Logical Memory—Recollection	Logical Memory—Recognition	Substitution	Completion
1. Coordination		+.67	+.48	+.39	+.13	+.01	+.11	+.19	−.17	+.16	+.39	+.18	+.14	+.35	+.43	+.48	+.24	+.37	−.01
2. Tapping	+.67		+.27	+.26	+.40	+.27	+.10	+.25	−.01	+.13	+.42	+.19	+.32	+.00	+.23	+.02	−.13	+.29	+.09
3. Cancellation	+.48	+.27		+.70	+.23	+.13	+.17	+.19	+.05	+.47	+.62	+.26	+1.00	+.18	+.07	+.13	+.12	+.48	+.13
4. Checking	+.39	+.26	+.70		+.33	+.35	+.30	+.38	+.22	+.33	+.44	+.35	+.12	−.18	−.07	+.26	+.24	+.34	+.22
5. Color Naming	+.13	+.40	+.23	+.33		+.32	+.36	+.48	+.28	+.27	+.38	+.24	+.29	+.33	+.13	+.12	−.07	+.41	+.27
6. Directions	+.01	+.27	+.13	+.35	+.32		+.36	+.52	+.52	+.12	+.26	+.28	+.18	+.33	+.22	+.32	+.46	+.16	+.22
7. Opposites	+.11	+.10	+.17	+.30	+.36	+.36		+.59	+.83	+.46	+.23	+.23	+.22	+.16	+.04	+.31	+.20	+.28	+.44
8. Verb-object	+.19	+.25	+.19	+.38	+.48	+.52	+.59		+.59	+.33	+.45	+.19	+.09	+.14	+.17	+.19	+.13	+.39	+.18
9. Mixed Relations	−.17	−.01	+.05	+.22	+.28	+.52	+.83	+.59		+.35	+.34	+.39	+.26	+.00	−.09	+.47	+.26	+.03	+.71
10. Word Building	+.16	+.13	+.47	+.33	+.27	+.12	+.46	+.33	+.35		+.50	+.13	+.12	+.45	+.17	+.51	−.03	+.34	+.52
11. Word Naming	+.39	+.42	+.62	+.44	+.38	+.26	+.23	+.45	+.34	+.50		+.21	+.18	+.50	+.25	+.36	+.15	+.41	+.39
12. Knox Cube	+.18	+.19	+.26	+.35	+.24	+.28	+.23	+.19	+.39	+.13	+.21		+.20	−.37	+.15	+.42	+.37	+.26	+.33
13. Digit Span	+.14	+.32	+1.00	+.12	+.29	+.18	+.22	+.09	+.26	+.12	+.18	+.20		+.23	+.25	+.11	−.04	+.11	+.28
14. Word Memory—Recollection	+.35	+.00	+.18	−.18	+.33	+.33	+.16	+.14	+.00	+.45	+.50	−.37	+.23		+1.00	+1.00	+.61	+.62	+.32
15. Word Memory—Recognition	+.43	+.23	+.07	−.07	+.13	+.22	+.04	+.17	−.09	+.17	+.25	+.15	+.25	+1.00		+.71	+.53	+.60	−.04
16. Logical Memory—Recollection	+.48	+.02	+.13	+.26	+.12	+.32	+.31	+.19	+.47	+.51	+.36	+.42	+.11	+1.00	+.71		+.95	+.33	+.35
17. Logical Memory—Recognition	+.24	−.13	+.12	+.24	−.07	+.46	+.20	+.13	+.26	−.03	+.15	+.37	−.04	+.61	+.53	+.95		+.20	+.21
18. Substitution Whole	+.37	+.29	+.48	+.34	+.41	+.16	+.28	+.39	+.03	+.34	+.41	+.26	+.11	+.62	+.60	+.33	+.20		+.20
19. Completion	−.01	+.09	+.13	+.22	+.27	+.22	+.44	+.18	+.71	+.52	+.39	+.33	+.28	+.32	−.04	+.35	+.21	+.20	
Average correlation between each test and all the others	+.25	+.21	+.32	+.28	+.27	+.28	+.30	+.30	+.28	+.30	+.36	+.22	+.23	+.32	+.26	+.39	+.26	+.32	+.27

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

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

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

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

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

TABLE XXVI

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

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

TABLE XXVII

TEST CORRELATIONS BETWEEN TRIAL 1 AND TRIAL 2—GROUP OF 45 FRESHMEN

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

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

SECTION VII

CORRELATIONS BETWEEN THE TESTS AND ACADEMIC MARKS

TESTS VERSUS MARKS AS MEASURES OF MENTAL ABILITY

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

	TABLE XXVIII LANGUAGE	TABLE XXIX MATHEMATICS	TABLE XXX SCIENCE
Grade	Frequency	Frequency	Frequency
F (50–60)	2	1	4
D (60–70)	14	14	6
C (70–80)	49	33	16
B (80–90)	30	30	12
A (90–100)	2	10	3

	TABLE XXXI—PHILOSOPHY	TABLE XXXII—HISTORY
Grade	Frequency	Frequency
F (50–60)	0	0
D (60–70)	1	4
C (70–80)	10	16
B (80–90)	12	4
A (90–100)	4	2

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

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

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

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

TABLE XXXIII

Correlations Between Tests and Academic Records

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

TABLE XXXIV

CORRELATION BETWEEN TESTS AND INTELLIGENCE QUOTIENT

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

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

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

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

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

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

Aikins [51] found a slight difference in the relative positions assigned to 17 students in a philosophy class by the students themselves on the basis of several ten-minute tests, and the positions he assigned them on the basis of four hour tests. Smith gives several plates, illustrating clearly the great discrepancies and marked lack of uniformity in marking systems at the University of Iowa.[52]

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

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

[48] Meyer, Max. The Grading of Students, Science, 28; 243–252.

[49] Dearborn, W. F. School and University Grades.

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

[51] Aikins, H. A. The Reliability of "Marks," Science, N. S., 1910, 32; 18–19.

[52] Smith, A. G. A Rational College Marking System, Journ. of Educ. Psychol., 1911, 2; 383–393.

subjects.[53] He also observed a marked lack of conformity to a standard in the case of grades given by individual instructors. When Jones [54] gave an opposites test and a memory test to each of two elementary psychology classes, taught by different instructors, he obtained these interesting results:

	Instructor "A" (28 students)	Instructor "B" (33 students)
Class standing and opposites	.09	.49
Class standing and memory	.44	.07

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

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

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

[53] Zerbe, J. L. Distribution of Grades. Journ. of Educ. Psychol., 1917, 9; 575–588.

[54] Jones, E. S. A Suggestion for Teacher Measurement. School and Society, 1917, 6; 321–322.

[55] Zerbe, J. L. Distribution of Grades.

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

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

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

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

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

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

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

[60] Starch and Elliott, School Review, 21, 254–259.

[61] Starch, D. The Reliability and Distribution of Grades, Science, 1913, 38; 630–636.

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

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

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

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

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

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

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

[63] Miner, J. B. The College Laggard. Journ. of Educ. Psychol., 1910, 1; 263–271.

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

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

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

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

SECTION VIII

CORRELATIONS BETWEEN PSYCHOLOGICAL TESTS AND PHYSICAL MEASUREMENTS. THEIR SIGNIFICANCE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The uniformity of the single correlations in exhibiting this tendency toward chance relationship is significant. In only eight cases out of the total number of 154 correlations, or, in fact, we might say in only six cases, since the correlations between Substitution First-half and lung capacity (+ .20) and Substitution Second-half and lung capacity, (+ .26) duplicate information yielded by the correlation between Substitution Whole and lung capacity (+ .24)—are there correlations of + .20 or over. The highest correlation is only + .26 (Substitution Second-half and lung capacity), which is too low to admit of diagnostic purposes. With these few exceptions, all the correlations between physical measurements and the tests—146 correlations in all—show approximately zero relationship. The large number of these correlations justifies us

TABLE XXXV

CORRELATIONS BETWEEN TESTS AND PHYSICAL MEASUREMENTS

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

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

It is interesting to know that the only other experimenter who has reported the results of a similar study with college freshmen supports this view. Although Wissler in his study of the results of the old Columbia freshman tests reports only two correlations between the physical tests and the psychological tests—namely, a correlation between length of head and logical memory of + .21, and between breadth of head and logical memory of − .05—the observation of the records of freshmen in other physical tests compared with their records in the psychological tests lead Wissler to conclude: "That the physical tests show a general tendency to correlate among themselves, but only to a very slight degree with the mental tests." [69]

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

[69] Wissler, Clark. Psychological Review Monograph Supplement, June, 1901.

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

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

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

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

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

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

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

On the basis of the relationship shown by the correlation coefficients we may divide the tests into three groups: (1) motor tests (Coordination and Tapping); (2) tests involving powers of perception and comprehension (Cancellation; Checking, Color Naming, Word Naming and Substitution); (3) tests involving associative relations (Directions, Opposites, Verb-object, Mixed Relations, Word Building and Completion); (4) tests which call into play powers of learning, viz., observation and retention—(Word Memory and Logical Memory); (5) tests depending on the subject's knowledge more than on her innate ability (Information and Vocabulary); (6) miscellaneous group (Digit Span and Knox Cube). There is only a chance correlation between Information and Vocabulary and the other tests. With the exception of this group and Digit Span and Knox Cube, the remaining groups of tests correlate closely among themselves but loosely with the other tests.

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

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

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

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

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

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

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

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

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

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

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

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

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

The advisor should be free to devote all his time to supervising the academic career of the students and to rendering needed advice. Obviously such a man should be a psychologist with both ability to interpret the various measures secured of each student's ability and tact in persuading students to follow his suggestions. From the attempts that have thus far been made in certain institutions to guide students' academic careers, it seems probable that with an able vocational advisor aided by a competent assistant such a system would be a distinct help in stimulating students to exert maximum effort in doing their college work.