

UM Libraries Depository



103344702006



A STUDY OF THE EBBINGHAUS CONJECTURAL
METHOD.

by

Albert Paul Weiss, A. B.



Submitted in partial fulfillment of the
Requirements for the Degree of
Master of Arts
in the

Graduate School
of the
UNIVERSITY OF MISSOURI
1912

378.7M71
XW436

23979

CONTENTS

	Page
Introduction	2
Method	3
Specimen Quiz	4
Relation between rank and grade	7
Variations in the difficulty of the quizzes	10
Variations in the difficulty of the several blanks	12
Comparison of the conjectural with the question and answer method	14
Variability of the different degrees of ability	21
The form of the distribution of a single quiz	23
Variations between the different sections of the same class and between the sexes	26
Degree of Selection	28
Comparison of Actual and Normal distributions	32
Supplementary remarks	40
Summary	44

2012S.P.W

A STUDY OF THE EBBINGHAUS CONJECTURAL METHOD.

Introduction

While the need for persistent drill is clearly recognized in the elementary and secondary schools, when the student enters college it is supposed that he is prepared to at once assume responsibility for his studies, a responsibility which in his preparatory work was largely assumed by his parents and teachers. This sudden attenuation in discipline is usually accompanied by a novel and distracting change of environment. At best, it can be expected that the student will show personal initiative only in a few of the subjects which he will take during his first two years at college. The age at which the student enters the university makes a considerable degree of supervision desirable, if not absolutely necessary. Yet where the classes are large and the instructor cannot personally assure himself that every student is spending the proper proportion of his time in study, some method must be devised which reflects the habits of study, and if possible exerts control over these habits. One of the best means of doing this is the giving of frequent quizzes, but the time required for grading the papers may take up so much of the instructor's time, that little opportunity is left for self-improvement. Grading papers is of little benefit, either to the student or the instructor, and since the Ebbinghaus conjectural method reduced the time required for grading papers to a very great extent, it was thought worth while to attempt a quantitative study of the results secured by the method.

Method

The conjectural method was used to determine the relative scholarship of the students taking the course in the introduction to psychology. The method was used on two classes: One of which took the course during the Spring semester of 1911, and the other during the Fall semester of the same year. In each class there were about 150 students, mostly freshmen and sophomores. All lectures were given by Professor Max Meyer, who also devised the technique of grading.

The method was also used on the students taking the elementary course in inorganic chemistry during the Fall semester of 1911. This class was divided into two sections of about 80 students each. Professor Herman Schlundt has given access to the records of this class for purposes of comparison.

On the page which follows is an example of the type of quiz used in the department of Experimental Psychology.

The words written in red, were of course omitted and the student was expected to supply the appropriate words for the blanks thus left. Each blank correctly filled in was counted as one; one-half was allowed when the blanks were filled in with words which were incorrect, but which indicated that the student had more than the vaguest notion of the contents of the sentence or paragraph. When all the papers had been corrected, they were ranked in the order of their excellence. The best paper was marked 1, the second best 2, and so on until the poorest paper which was marked n, depending on the number of

Specimen Quiz

The first business of a careful psychological investigation into the phenomenon of attention is to distinguish between its manifestations and its conditions. Hence the question at once arises, whether we have in the process of attention something above and beyond the elementary conscious contents or whether the term simply indicates a special state in which all those contents may be presented under certain conditions. It seems that the second alternative is the correct one; for introspection discovers nothing really new in attention, nothing which is characteristic of the process as such. Those who maintain the contrary have nothing to allege but a sum of sensations, which occur just as well in other contexts. Strain sensations, for example, are ordinary concomitants of attention, and are referred to in the phrases 'strained' and 'relaxed' attention; but they are neither necessary constituents of attention nor attention itself. Some describe attention as an internal activity, concentrated in various degrees upon the particular ideas in consciousness. But this description applies to attention, as given to introspection, only on the assumption that we are able to perceive the internal activity as such. And this would mean that we are able to perceive an internal activity in addition to the contents of the perception upon which it is said to be concentrated; and above and beyond the concomitant strain sensations to which its functioning gives rise. We are unable to discover any such distinct act of consciousness in the state of attentive perception or imagination. Let no one mistake for such an internal act of consciousness the kinesthetic sensations which we have mentioned as concomitant. The psychological problem of attention may therefore be stated somewhat as follows: The psychologist must describe the way in which the simple and complex mental states occur and change with respect to their relative degrees of consciousness.

students in the class. For the purpose of determining the scholarship of the students, the actual number of blanks filled in was not recorded. At the end of the semester the ranks were added together and the student having the lowest rank number was considered the best student in the section; the student having the next lowest, the next best, etc.

For the purpose of the statistics used in this paper, two methods of grading were employed in addition to the one just described. Each quiz was graded on the basis of 100 where all the blanks had been acceptably filled in. If a quiz contained 20 blanks each blank was allowed the value $100/20$ or 5. If a student filled in only 12 blanks correctly, his grade would be 12×5 or 60.

In a third method of scoring, the average number of words filled in by the class was taken as basis and this average given the value of 50. More will be said of this method in a more appropriate place.

Briefly stated, the first method gives the rank or position of the student in his particular section without reference to the number of blanks actually filled in; the second method indicates the proportion of the blanks acceptably filled in, without reference to his position in the section; while the third method indicates the position of the student with reference to the average accomplishment of the section.

To facilitate comparisons, all standard deviations and probable errors are magnitudes of the same order. The formulae used are known as the Pearson formulae and are given on the following page.

Pearson Formulae

Standard Deviation $\sigma = \sqrt{\frac{S X_n^2}{n}}$

Probable Error $\frac{.6745 \sigma}{\sqrt{2n}}$

Correlation $r = \frac{S X_n \cdot y_n}{n \sigma_1 \sigma_2}$

Probable Error $\frac{.6745 (1 - r^2)}{\sqrt{n(1 + r^2)}}$

σ = Standard Deviation

x = deviation from the average in one quiz.

y = deviation from the average in another quiz.

n = number of students.

S = sum of n

r = degree of correlation

Formula for calculating grade when average accomplishment of the class is taken as basis

$$g = \frac{50b}{a}$$

g = Grade of student.

b = number of blanks filled in.

a = average (or median) number of blanks filled in by the class.

Relation between Rank and Grade

Table 1 on the following page, shows the distribution of the students in two ways: (1) by ranks; (2) by grades. Since the sections were not all of the same size, it was necessary to reduce the rank numbers of each section to terms of one of the other. The 2 oclock section being the largest, was used as basis.

The most apparent fact that the table reveals is that the distribution of the grades is not the same as that of the ranks. Of course, this is to be expected. When the method of ranking is used the limits of the numbers recording the ranks, depend only on the number of students present at the quiz. In the method of grading here used, the limits are the same for every quiz; namely 0 and 100. The ranks for two quizzes given to the same number of students would be the same, even though one of the quizzes were much more difficult than the other. This would not be the case with the grades. Graphically expressed, the curve showing the ranks for a single quiz is a straight line. However, as soon as a number of quizzes are added together, the straight line character is lost.

Curve No. 1 graphically shows the nature of the relation between grades and ranks. In order that the curve showing the ranks might fall in the same geometrical position as the curve showing the grades, their limits were made spatially the same.

It will be seen that the two curves resemble each other very closely. The curve showing rank has lost its straight line character

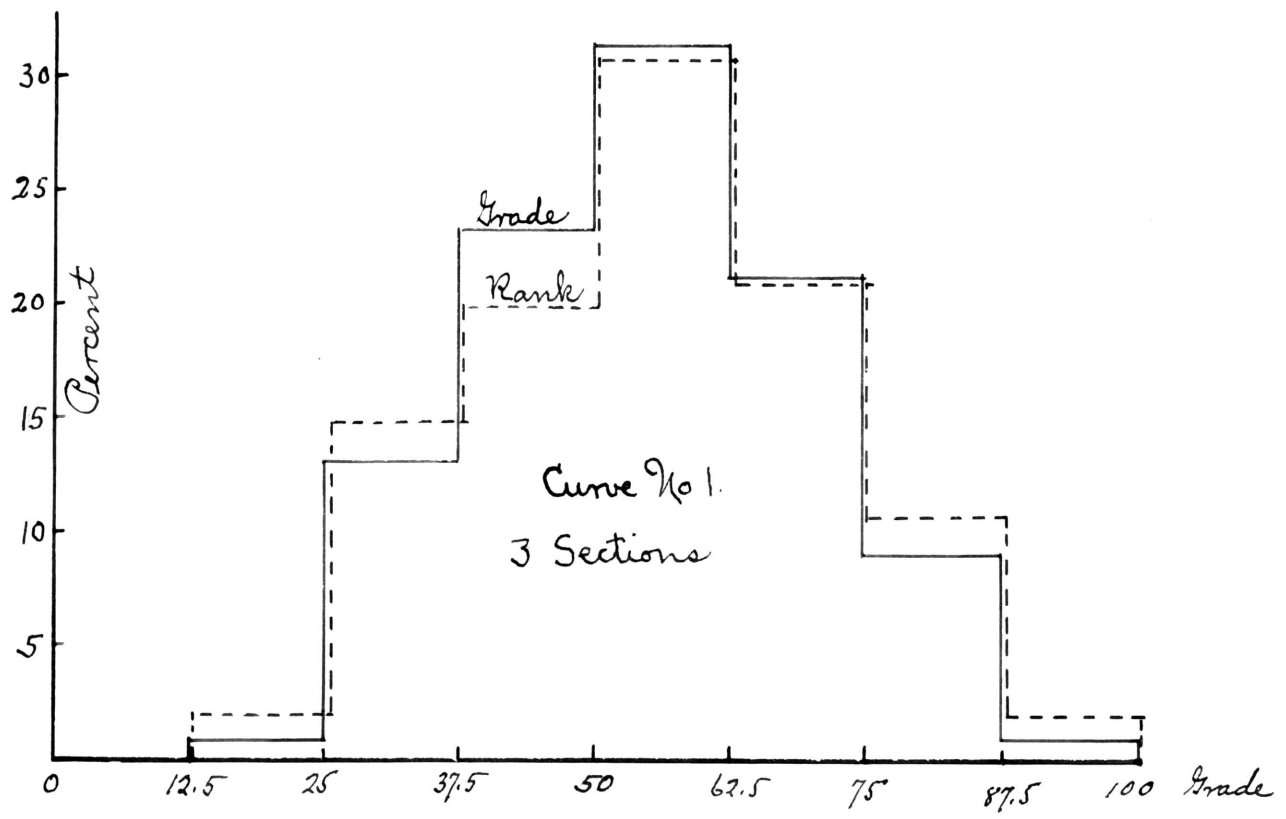
Table I

Distribution by Ranks

Average Rank Nine Quizzes	8 o'clock Section %	11 o'clock section %	2 o'clock section %	all three sections %
42	0	4	2	2
37	14	16	12	13
32	19.5	0	16	13
26	19.5	32	25	25
21	25	28	20	24
16	11	16	16	14
11	11	0	9	8
5	0	4	0	1
Number of Students	36	25	44	105

Distribution by Grades

Av. Grade	8 o'clock %	11 o'clock %	2 o'clock %	all three %
0 - 12.5	0	0	0	0
12.5 - 25.0	0	4	0	1
25.0 - 37.5	17	4	16	13
37.5 - 50.0	25	8	29.5	23
50.0 - 62.5	36	28	29.5	31
62.5 - 75.0	17	48	18	21
75.0 - 87.5	5	4	7	9
87.5 - 100	0	4	0	1
Number of Students	36	25	44	105



and tends to also measure actual accomplishment. The arbitrary units used in the method of ranking (in this case 1) have been eliminated and the steps between the ranks tend to become proportional to the steps between the grades. The advantage in the method of ranking lies in the fact that the marks secured by it are independent of the difficulty of the quizzes and the relative ability of the individual is thus more clearly shown.

Variations in the difficulty of the Quizzes.

Table 11 shows the average ^{class} grade made in 27 quizzes. The average grades range from 26.5 to 81.3. If the degree of difficulty is directly proportional to the average grade, this means that at least one quiz was three times as difficult as one of the others. While this seems to be a large variation, there is no reason for supposing it is peculiar to the conjectural method, or unparalleled by the ordinary question and answer method. An attempt to estimate the difficulty of a question with only personal opinion as a guide, is usually not accompanied by a feeling of conviction vivid enough to encourage the belief that such a subjective estimate is accurate. Where variations in the difficulty of ordinary quizzes become apparent, one is at a loss to determine how much of this is due to the personal equation. This makes it difficult to reduce the extent of variation into quantitative terms, and unless this is done we are not warranted in ascribing greater variability to the conjectural method.

Table II

Average Grade for 27 Quizzes

8 o'clock		11 o'clock		2 o'clock	
Quiz	Av. Grade	Quiz	Av. Grade	Quiz	Av. Grade
III	26.5	IX	41.3	IX	40.8
V	45.6	VIII	43.1	III	43.4
IV	47.6	III	57.9	V	51.0
II	51.6	VII	59.5	VI	51.8
VI	52.1	IV	63.7	I	55.5
IX	53.3	V	66.1	VII	56.7
VII	58.3	I	70.6	IV	56.8
I	65.4	II	71.7	II	57.0
VIII	66.8	VI	81.3	VIII	58.8
Number of Students	36		25		45

Series expressing the relative difficulty
of the blanks in two quizzes.

Quiz A. 1.0, 1.1, 1.2, 1.4, 1.6, 1.8, 2.4, 2.6, 2.8, 3.2, 3.6, 4.2, 4.8, 5.4

Quiz B. 1.0, 1.3, 1.6, 1.8, 2.0, 2.2, 2.5, 2.9, 3.5, 4.4, 4.9, 5.7, 6.8, 8.7, 10.3, 11.1

In this paper I have restricted my interpretation of the data to such considerations as could be objectively determined by an examination of the quiz papers. When a quiz is described as being difficult this means only that the average grade of the whole class is low. When a particular blank is called difficult, this means that relatively few students filled it in correctly.

Variations in the Difficulty of the Several Blanks.

In grading the quizzes, it was assumed that each blank was equal in value to every other blank. In order to see to what extent this was justified, the percentage of students who filled in a particular blank correctly, was taken ^{*inversely,*} as a measure of the relative difficulty of a blank. When this is done we again find a wide variation even in the difficulty of the several blanks. The two series below Table 11 express the relative difficulty of the blanks in two quizzes.

We find in quiz A a variation from 1.0 for the least difficult blank, to 5.4 the blank filled in by the least number of students. If we assume, for the moment, that the difficulty of the blanks is proportional to the figures we have selected, the 5.4 blank is more than five times as difficult as blank 1.0. In other words, a student who only filled in the two blanks 4.8 and 5.4 should receive the same grade as a student who filled in the seven blanks at the beginning of the series. If two blanks may be equal to seven blanks, it would seem as if the method of allowing each blank the same value as any other blank

is radically wrong. In order to ascertain the extent of error which results when all blanks are given the same value, the whole quiz was regraded. What was done was to change from a unit which was the same for every blank, to one which varied for the different blanks. By doing this, we would of course expect to get different grades, but if the position of the students in the one series is the same as their position in the other series, then the question of what method to adopt can be only one of convenience, not of accuracy.

Since the blanks in quiz B show even a greater variation than quiz A, the former was considered as best showing the effect of variations in the difficulty of the blanks, on the standing of the students. The extent of correlation between the two methods was found to be .92 with a probable error of .006. Since this leaves the position of the students practically unaltered, there is no necessity for using the more cumbersome method of graded or differentiated blanks. The poorer students who filled in relatively few blanks, did not fill in those blanks which were filled in by few students -- that is, the difficult blanks.

To one who has not corrected very many conjectural quizzes, this comes unexpected. An uncritical glance at the blanks is quite likely to lead to the conclusion that filling in any one of the blanks is largely a haphazard occurrence -- a matter of "luck" as it were. If this were the case we should expect the correlation to be zero, while the degree of correlation actually found was .92. To one who is more

familiar with the method it is surprising to find that time and again some particular blank will be "avoided" by nearly all the class except two or three students, who are not only considered good students in psychology but who have been found to also rank high in their other courses.

This is one of the strongest arguments to show that the conjectural method is not essentially different from any other method of testing which depends upon a comprehension, rather than a mere memory, of the subject matter.

Comparison of the Conjectural with the Question and Answer Method.

In order to determine how closely the grades secured by the conjectural method corresponded to those secured by the conventional method of giving the class a number of questions and having them write out the answers, the correlation was calculated between one conjectural quiz and one question and answer quiz, and then as control, the correlation between the two conjectural quizzes immediately following. Both pairs of quizzes being, of course, given to the same students. In Table III the extent of the correlations are shown.

From the similarity of these correlations, it does not seem as if the conjectural method tests a form of ability essentially different in nature from that required in writing out the answers to a question. The rather low degree of correlation in both instances is due to the fact that only single quizzes were taken. Where the

Table III

Psychology

Method	Standard Deviation	Probable Error	Number of Judges	Number of Stud.	Correlation	Probable Error
Conjectural Question	24.2	1.62	1	50	.34	.08
	16.7	1.14	1	50		
Conjectural Question	16.4	1.16	1	50	.36	.08
	18.6	1.32	1	50		

Chemistry

Conjectural Question	12.5	.65	25	80	.69	.03
	20.6	1.10	1	80		
Conjectural Question	10.9	.57	25	80	.63	.04
	19.7	1.07	1	80		
Class Work Laboratory	14.2	.78		79	.59	.04
	8.1	.31		79		
Class Work Laboratory	13.7	.45		80	.71	.03
	6.9	.37		80		

average of a number of quizzes is taken the correlation is higher.

By taking the data from the Department of Chemistry in which larger classes are found, the degree of correlation reaches the magnitude usually found when mental traits are compared. In the second part of Table 111 the average grades of 25 conjectural quizzes were taken and correlated with the final question and answer examination. This final examination was of two hours duration and covered the work of the semester. It was planned more carefully than the shorter conjectural quizzes and the students also devoted more time to preliminary preparation. It therefore represents more nearly the actual ability of the various students in the subject of chemistry than does any one of the conjectural quizzes given during the term. For the purpose of comparison, the class work of these same students was correlated with their laboratory grades.

As will be seen from the table, there is nothing characteristic in the data which will indisputably reveal the nature of the examination used. The correlation between the conjectural quizzes and the question and answer examination, might very well represent the correlation between the class work and laboratory grades.

The value of the conjectural method as a means for determining scholarship, will depend largely upon the extent to which the grades secured by it, correspond to the grades given by the instructors in the students' other courses. No matter how satisfactory the method may be to the instructor, if it gives grades which are widely at variance

with the grades the student receives in his other courses, it will not be adopted. In order to throw some light on this question, the grades resulting from the conjectural method were compared with the grades reported for the same students by the instructors in their other courses. The grades handed in to the registrar by Professor Meyer were correlated with the average grades handed in by the other instructors for these same students and this correlation was found to be .63 with a probable error of .033. Out of 104 students there were only four reversals -- cases in which the grades awarded in psychology were on the other side of the mean, than the grades awarded in other courses. Where the products of the deviations of 100 students gives only four negative quantities, and these relatively small, there seems no justification for the assertion which is sometimes made, that the conjectural method tests only a special ability such as mere memory, -- unless we assume that this is all that the other methods of examining at the university are doing.

The same correlation was determined for the Department of Chemistry, and while here the grade of the student was not entirely determined by the conjectural method, at least one-half of the information which went to make up a student's grade was so secured. The grades in chemistry as compared with the grades of the same students in their other subjects, show a correlation of .60 for one section and .63 for the other. Again there were only 6 reversals. The extent of correlation is practically the same as that found in psychology, and what would be expected if the conjectural method determines scholarship

in a manner similar to that of the ordinary methods of examination.

There is another way in which the question of rote memory can be tested. One of the quizzes was taken from a psychological text book which, undoubtedly, none of the students had ever seen, and the wording of which was different from that in the text they were using. There can be no question of "mere memory" here because the students had never seen the book. The results are shown in Table IV.

For the purpose of comparison we will take two conjectural quizzes taken from the text book the students were using, and with which they were supposed to be familiar. If "memory work" is a stronger factor in the conjectural than in other methods, we should expect these correlations to show it. We find again that the correlations are practically identical. Between the quiz taken from an unknown source and the quiz taken from the text book the students were using we find a correlation of .55, while the two quizzes taken from the text give .56. Here word memory has been excluded and the students are entirely dependent upon their comprehension of the subject; no essential difference in the configuration of the quiz results is found.

Finally, the effect of memory was investigated in yet another way. A quiz was given in which a number of the blanks called for the same word. Now if ONE of these blanks was filled in correctly, we can no longer say that the other blanks calling for the same word were left blank because the student had forgotten the word. The correct word had already been used. In this case the whole quiz was given a

(1) This is the quiz given as a specimen on page 4. The word referred to is "attention."

Table IV

Source of Quiz	Section	Number Stud.	Std. Dev.	Prob. Error	Correlation	Prob. Error
Unknown Text Book	8	49	30	2.0	.60	.06
Regular Text Book	8	49	24	1.7		
Unknown Text	11	54	28	1.8	.53	.06
Regular Text	11	54	24	1.6		
Unknown Text	2	57	30	1.9	.51	.06
Regular Text	2	57	22	1.4		

Average .55

Regular Text Book	8	54	28	1.9	.54	.06
"	8	54	30	1.9		
"	11	57	28	1.8	.60	.06
"	11	57	28	1.8		
"	2	56	30	1.9	.53	.06
"	2	56	30	1.9		

Average .56

value of 100. The various blanks again received values depending on the number of students who filled them in, and these values were so selected that ^hwere all the blanks had been filled in the sum was 100. By this method the relative difficulty of those blanks requiring the same word can be expressed by the series 1.4, 1.5, 1.6, 1.8, 2.0, 2.1, 2.4, 4.8, 5.7, 10.3. The range is from 1.4 to 10.3, or we might say the last blank was found to be seven times as difficult as the first one, in spite of the fact that the word required was the same in both cases, and was the same word which a student had already written in some of the other blanks.

To illustrate how such a condition may arise, we can simplify the factors involved by taking a few examples from arithmetic. Suppose we take four equations which require an increasing degree of arithmetical ability for their solution. The solution to each of the four equations to be identical. This would be analogous to filling in blanks in sentences which ranged from contexts which readily suggested the correct word, to such which required special training for their comprehension.

$$\overset{I}{7+6} = 13$$

$$\overset{II}{\sqrt{13^2}} = 13$$

$$\overset{III}{\sqrt[3]{144}} + \sqrt[3]{13^3} = 13$$

$$\overset{IV}{\sqrt[4]{28561}} = 13$$

In these four problems the answer is in each case 13. That 13 is the correct answer to the first problem is evident to any 10 year old child. The second requires somewhat more training in arithmetic. The third still more, while perhaps no one could by mere inspection tell

that the fourth root of 28561 is also 13. We have here an instance in which the solutions to a number of problems are all the same, but to determine this fact various degrees of comprehension are necessary. We must understand the whole equation before we can feel certain that one of its members is correct. Memory of the number 13 does not enter at all. Of course, the filling in of a blank in a conjectural quiz may not be of such a simple nature, yet in some respects it seems to be similar. This is substantiated by indirect evidence. In grading conjectural quizzes one frequently finds that synonyms are used, words which are not in the text but which are correct or partly so. A correct synonym can be supplied only when the student fully comprehends the context.

Variability of the Different Degrees of Ability.

It is a well-known fact that a grade of 80 given by a teacher inclined to be lenient, may express a lower degree of ability than a grade of 60 given by an instructor more rigorous in his requirements. Again, a grade of 80 in an easy quiz does not represent the same degree of ability as 80 in a very difficult test. Such ambiguities cannot be avoided as long as the subjective estimate of the instructor is the predominant factor in estimating ability, or as long as all tests are assumed to be equal in difficulty and actual, rather than relative, accomplishment is taken as basis. If an easy quiz is assigned a value of 100, a difficult quiz should receive a greater value. However, for the instructor to attempt to evaluate the difficulty of the various

tests is impracticable. Some objective measure is more satisfactory. Such an objective measure is used when the maximum accomplishment of the class as a whole is given the value 100 and the students given grades depending upon the percentage of the whole which they accomplish. This practice eliminates the personal equation of the teacher and the error due to variations in the difficulty of the several tests. The method has the disadvantage that it obscures the true variability of the students.

Another method is to grade not on the basis of total accomplishment, but on the basis of AVERAGE accomplishment. The method has the advantage that it is not necessary to perform the tedious addition, since the median may be used in place of the average. If then the median is set equal to some constant value, (in this paper the value 50 has been used) the results from one quiz can be directly compared with the results from any other quiz. For purposes of statistics, all standard deviations and probable errors will then be magnitudes of the same order. The algebraic expression for this operation may be found on page 6. The assumption involved in this method is that the accomplishment of the average student is less variable than that of the superior or inferior students. This assumption is justified by the following data.

Two quizzes were given to 150 students and the grades were divided into three groups: The one-fourth best students; the one-fourth medium students; and the one-fourth poorest students. The standard deviation and probable error was calculated for each group.

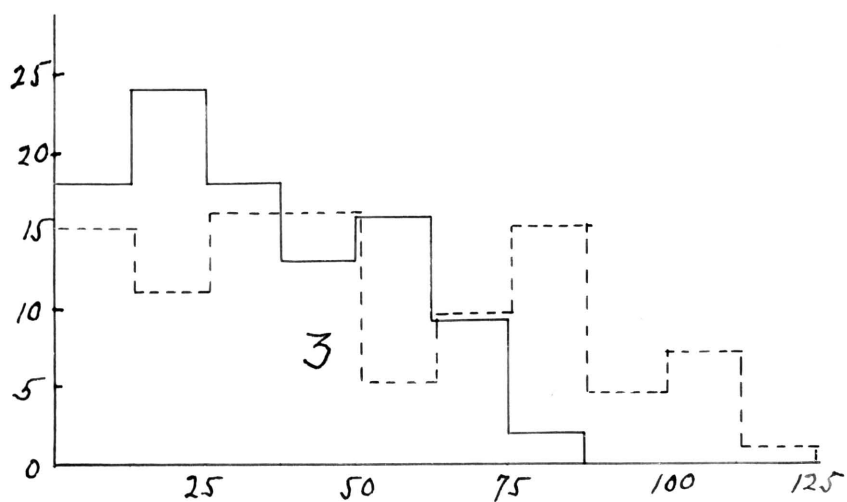
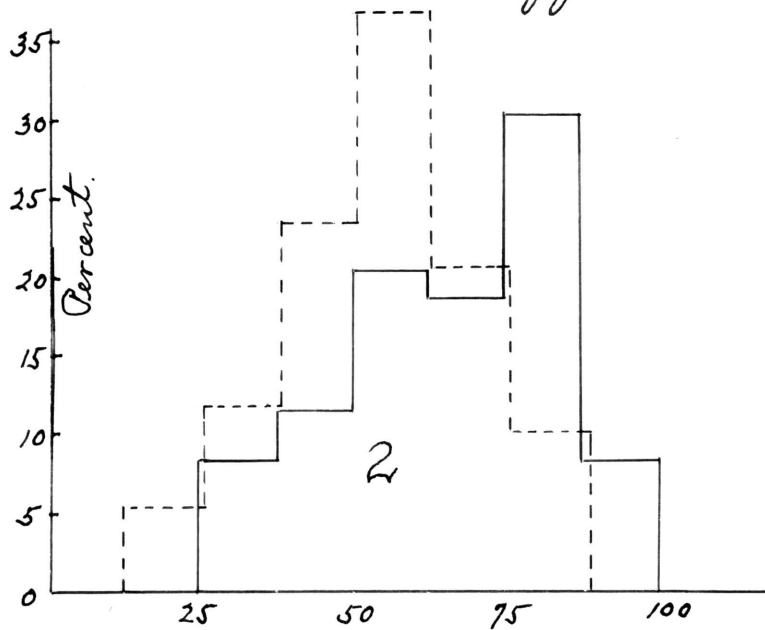
	Std. Dev.	Prob. Error
Best fourth	4.8	.33
Medium fourth	2.9	.23
Poorest fourth	8.5	.68

It will be noted that the least extent of variation is found in the medium or average group. Next in order is the best fourth, while the poorest fourth has a deviation of nearly three times that of the medium group. It would seem from this that the accomplishment of the average student is the least variable factor in this form of mental measurement and that the average accomplishment is the logical zero point rather than zero accomplishment. To avoid negative grades and bring the number series expressing the grades as much as possible within the conventional 1 to 100 series, the average accomplishment has been set equal to 50.

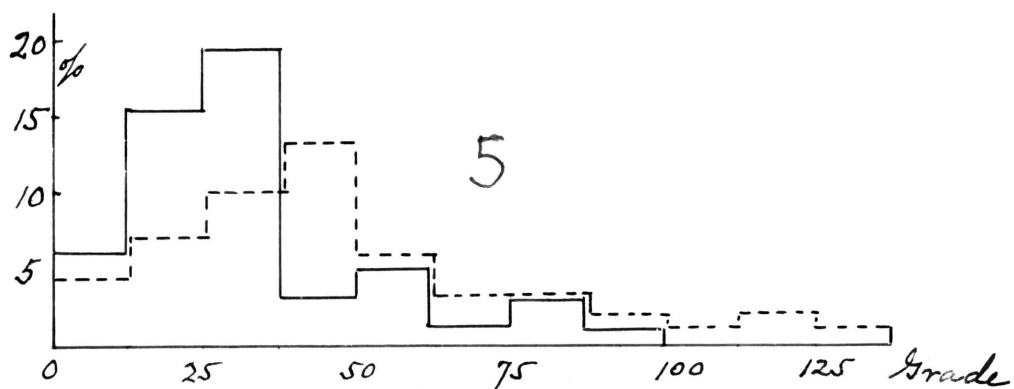
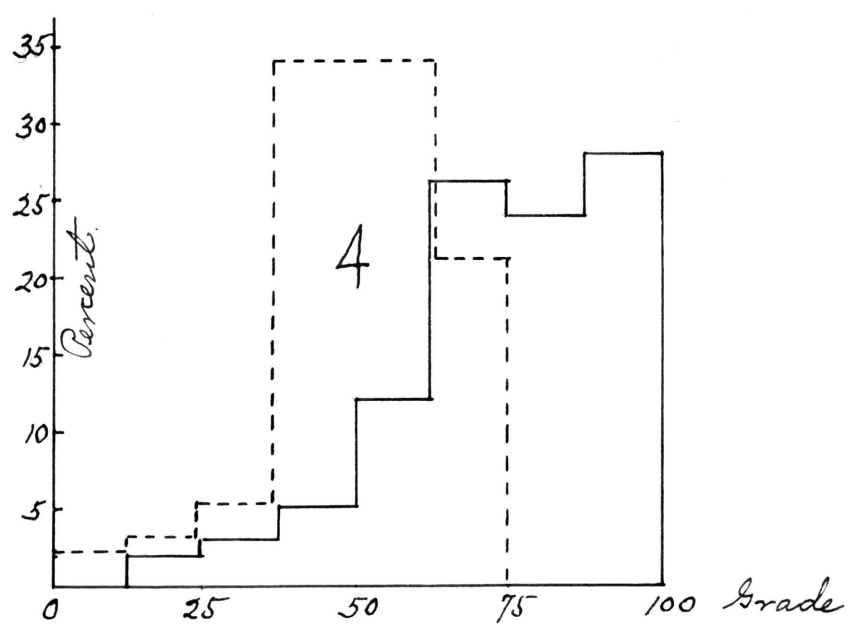
The Form of the Distribution of a Single Quiz.

A single test which is of short duration is not a good index of what the distribution will be when a larger number of quizzes have been given. A test may be so easy that the better students cannot express their maximum ability, or it may be so difficult that the poorer students cannot indicate the degree with which they comprehend the subject because the minimum required by the test is above the maximum which they are able to accomplish. In either case the relative scholarship of all classes of students is not tested. Curves 2, 3, 4, 5, show four quizzes which have abnormal distributions. The solid line shows the actual accomplishment, while the dotted line shows the accomplishment in terms of the class average. It is perhaps unnecessary to remark

Abnormal Quizzes



Abnormal Quizzes



that quizzes showing a decidedly skewed distribution are to be avoided. It is not likely that a class of any considerable size will show a markedly unsymmetrical distribution unless special selective factors have been operative. As a justification for this statement, the fact can be verified that when a large number of quizzes are given even to a relatively small class, the curve showing the distribution of the average grades, tends to approach the normal or binomial curve.

Variations between the Different Sections Of the Same Class and between the Sexes.

Where a large class has been divided into a number of sections, it is often desirable to know the extent to which the sections are likely to vary from each other. To ascertain the extent of correlation between the three sections in the class in psychology, two quizzes were given which were identical for all three sections. The results were also considered with reference to sex. Table Va shows the correlation with the normal curve.

The correlation of the three sections taken together is .93. The probable error shows, however, that all these correlations should be considered as identical. This means that there is no significant variability between either the different sections or between the sexes. It may be objected that the normal curve is not a good means to use as a standard. To meet this objection the different groups were correlated among themselves, as shown in Table Vb.

Table IIa
Correlation with Normal

Section	Number of Quizzes	Number Students	Sex	Corre- lation	Prob. Error
8 o'clock	2	43	Men	.94	.02
11 o'clock	2	49	Women	.89	.04
2 o'clock	2	49	Mixed	.95	.02
Men	2	73	Men	.94	.02
Women	2	68	Women	.92	.03
All together	2	141	Mixed	.93	.02
Average				.92	

Table IIb
Correlations inter se

The two series which are correlated	Sex	Corre- lation	Prob. Error
8 with 11 o'clock	Men with women	.80	.07
8 with 2 o'clock	Men with mixed	.98	.01
11 with 2 o'clock	Women with mixed	.82	.07
Men with women	Men with women	.98	.01
8 with whole class	Men with mixed	.98	.01
11 with " "	Women " "	.93	.03
2 " " "	Mixed " "	.97	.01
Men " " "	Men " "	.98	.01
Women " " "	Women " "	.96	.02

Average of last 5 groups .96

The correlation of the 8-11, 8-2, and 11-2 sections in this case is .80, .98 and .82 respectively, or an average of .90. In Table Va the correlations of these three sections with the normal curve was found to be .93. This shows that correlating the different sections with the normal curve gives a closer degree of correlation than when the sections are correlated with each other.

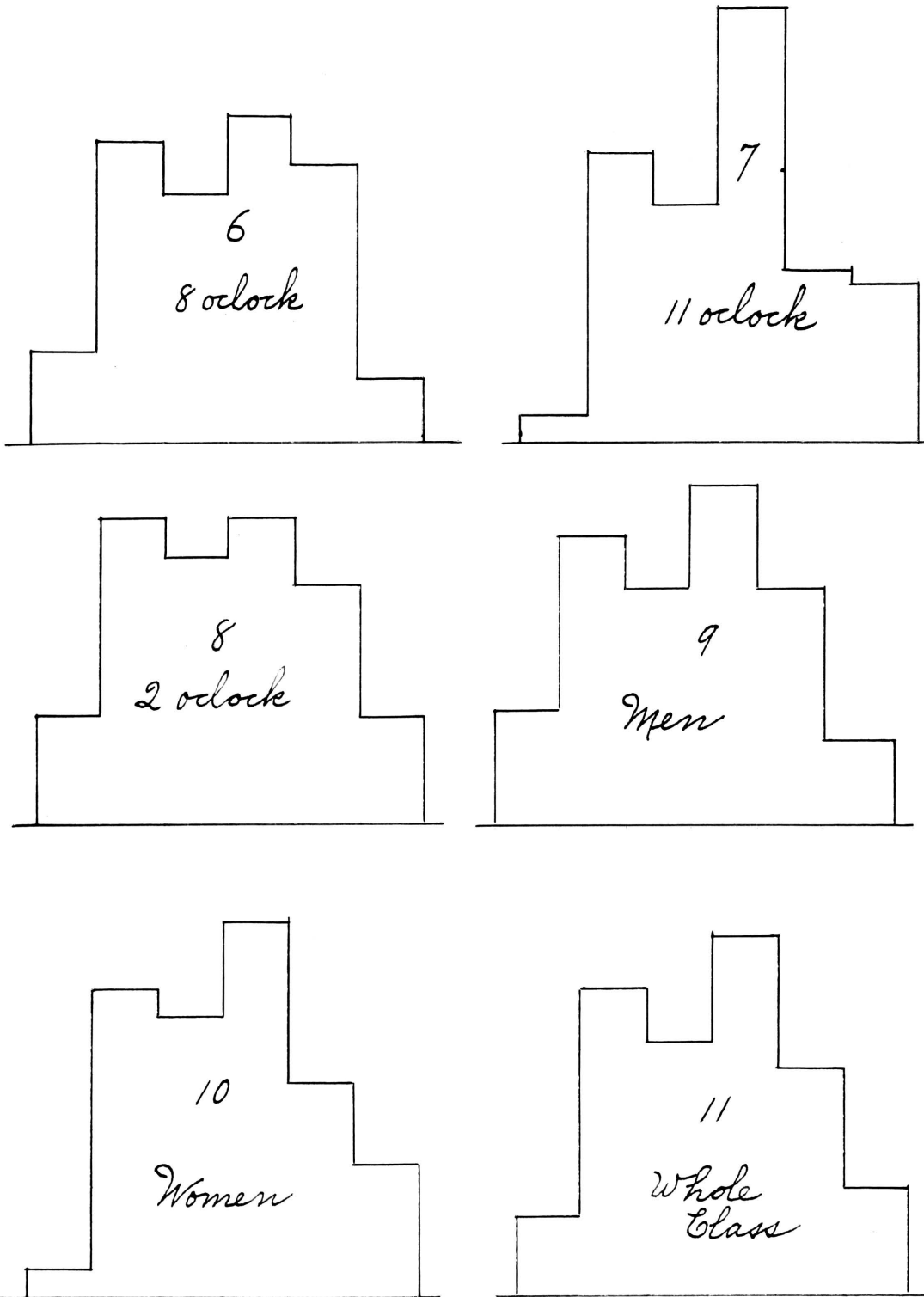
The last five lines of Table Vb give the correlations between the groups taken separately, and all the sections taken together. The result is the same as in the previous table, for when the probable errors are taken into consideration the correlations become practically identical. The average .96 is higher than in the preceding table but this would be expected since it is more likely that a combination of the three sections will possess the characteristics of any one of the groups to a greater extent than will the normal curve.

Curves 6, 7, 8, 9, 10, show the distributions graphically. The dimorphism which seems to characterize these curves is incidental. Curves 12, 13, 14, 15, 16, 17, show the distribution of the same students for all the quizzes given during the semester. Here the dimorphic character has disappeared.

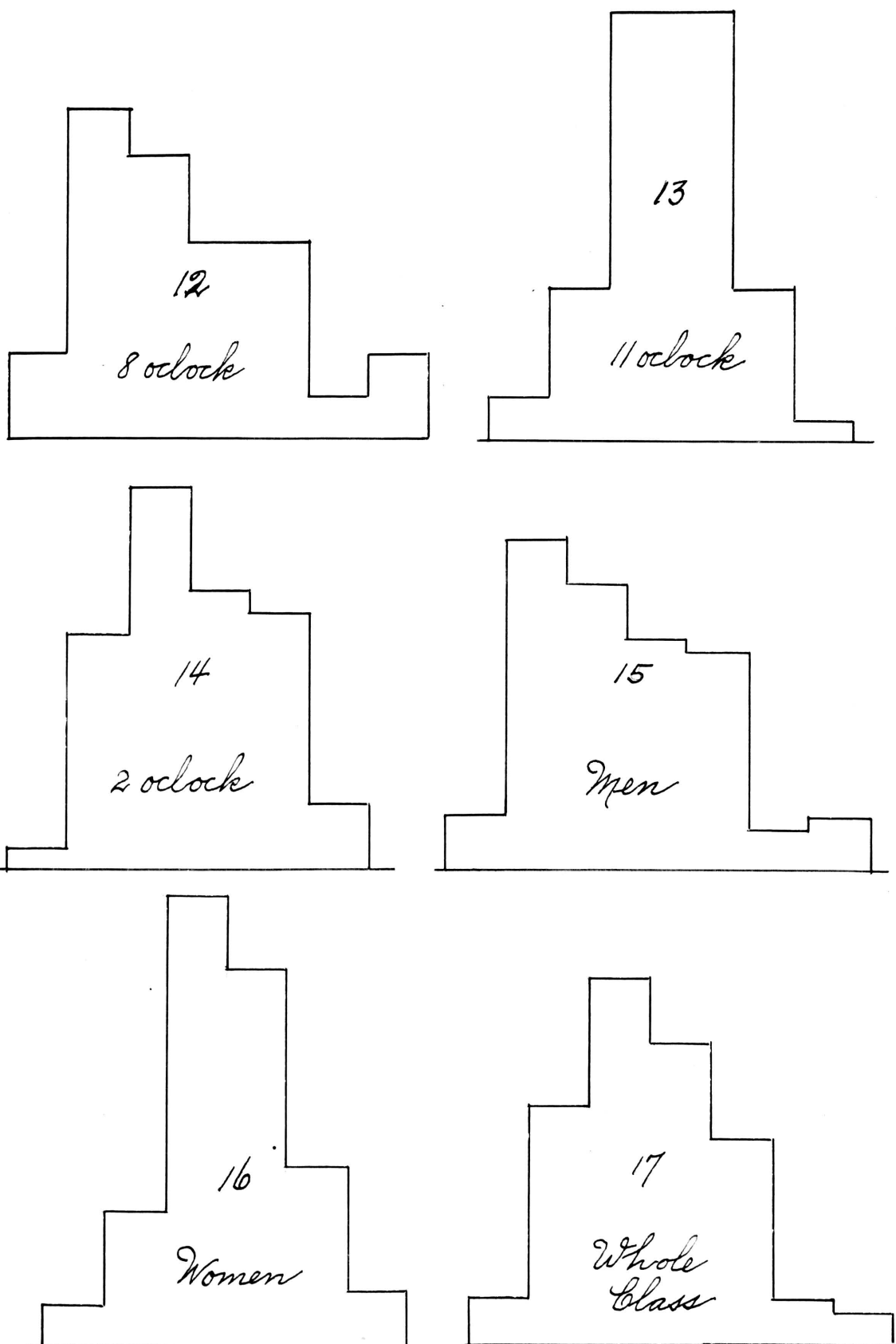
Degree of Selection.

The results secured by the conjectural method can be used to get some idea as to the degree of selection which has been exercised on the university freshmen. In a trait which is distributed according to the normal curve, we should expect to find as many individuals below

Distribution of two quizzes V and IX combined



Distribution for all quizzes given during the semester



the average as there are above it. Interpreting the data secured by the conjectural method in this way, we should expect that the number of students below the average, divided by the number above would be equal to 1. Suppose this is found to be the case in the elementary classes. What conclusions can be drawn as to the distribution of this type of ability among college students as compared with the population at large? In the measurement of school children ~~where~~ we have as random a sampling of the population at large as it is possible to secure for mental measurements of a nature resembling the type of ability required to meet the conditions imposed by the tests given the class in psychology. Under these circumstances the normal type of distribution is the one expected. In the absence of actual data this would be the most probable assumption we could make with respect to the distribution of the trait in the population at large. If the ability of college students is above this, we should expect to find a greater number below the average of their class. If their ability is below, we should expect to find a greater number above the average.

At first glance this seems paradoxical, but we can easily assure ourselves that this is true. If the poorer students are eliminated the average of the class will be relatively high, due to the fact that the remainder of the class contains a relatively larger number of good students. We can easily verify this by combining the grades of all the students and then artificially creating special groups having any desired degree of selection. In Table Va it has been shown that the correlation between the actual grades and what would be expected from the

normal curve, is very high (approximately .90) Table VI shows such arbitrary groupings and also what we would expect if the class as a whole were of the same composition as the population at large -- remembering that for the population at large we have taken only what seemed the most probable form of distribution, that of the normal curve.

In view of the fragmentary nature of the data, it will not be profitable to extend the interpretation farther. However, the results here given support the view that the first year students in the university are not a specially selected group with respect to mental capacity of the form required to meet the conditions of the elementary course in psychology.

Comparison of Actual and Normal Distributions.

In mental measurements the forms of distributions actually found vary much from each other. In a great majority of cases no simple equation will, when plotted, reproduce the distribution actually found in the measurements. This fact has been urged as an objection against assuming that mental traits approach a form of distribution having a relatively simple equation. This objection can also be urged in physics. If the actual data secured from experimental work on falling bodies were converted into an equation, this equation would not be identical with the equation which is called the law of falling bodies. Not only would the empirical equation differ from the law, but it would also be more complex. As the experimental conditions are refined and perfected we find that the empirical equation approaches the theoret-

Table VI

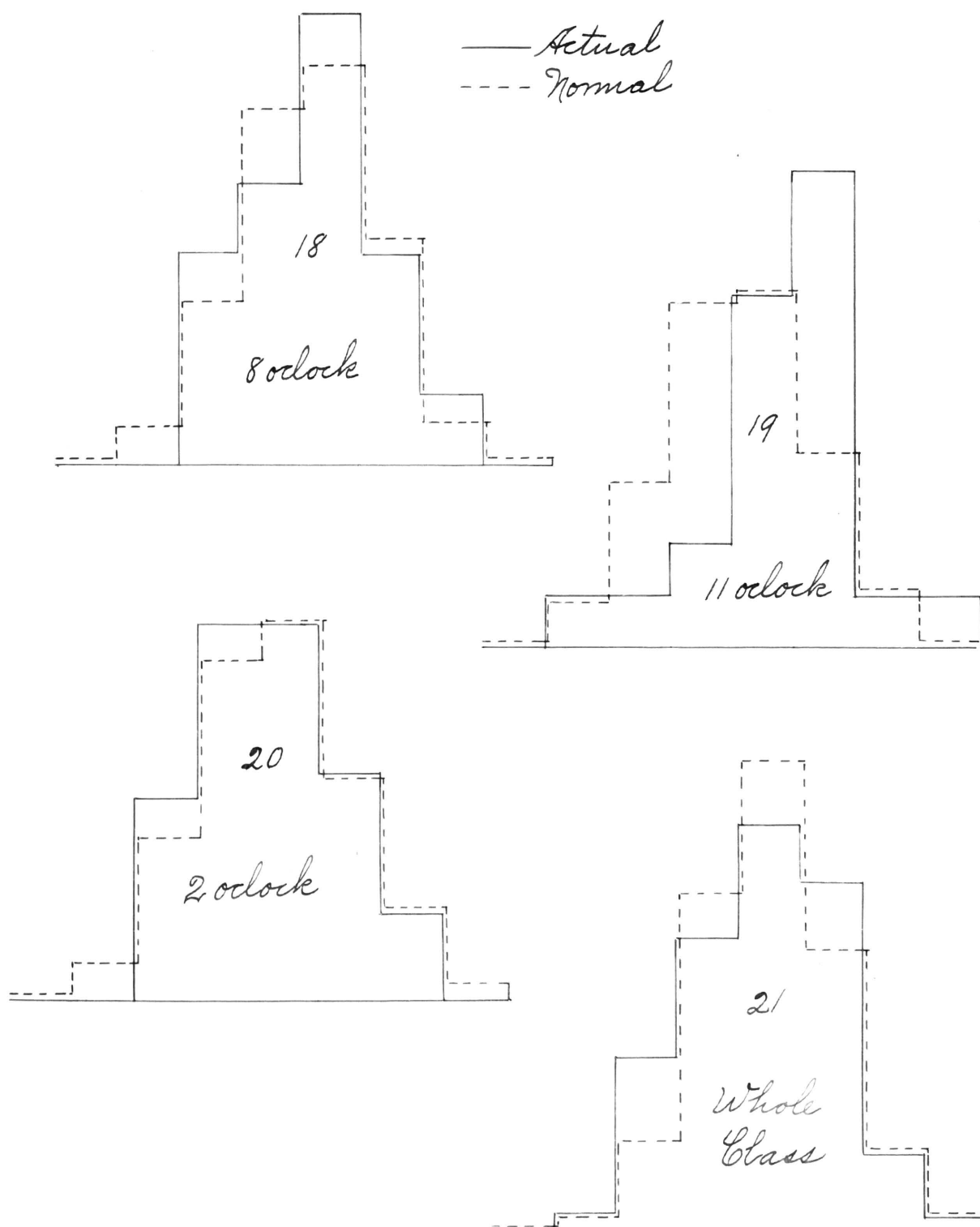
Group	below average	above average	actual ratio	expected ratio
Best fourth	58%	42%	1.38	more than 1.00
Poorest fourth	44½	55½	.80	less than 1.00
Middle fourth	49	51	.95	1.00
Whole Class	53½	46½	1.15	1.00

-ical equation more and more. Conversely, the fact that the experimental data in successive variations of the experiment approaches more closely to the values expected from the theory is held as evidence that these variations are real refinements and improvements. The only basis for this assumption is that physical laws are simple rather than complex. When we consider the progress which the science of physics has made since the discovery of such equations as Boyle's law relating to the elasticity of gases, of Ohm's law in the sphere of electricity, not to mention the contributions of Newton, Ampere and many others, we can understand why it is that students of mental measurements have tried to find a simple equation which expressed the quantitative relations which hold between the various forms of mental phenomena. As a guide, the theory of mental measurements recognizes that if a given type of distribution persists and becomes more pronounced as the groups become larger, that this is probably the type of distribution which would prevail in a group large enough to establish the actual type.

Many investigators have found that the curve variously known as the Normal curve, binomial curve, the "bell-shaped" curve, normal probability surface, etc, seems to be the limit which numerous mental traits approach. The grades secured from the quizzes recorded in this paper seem to belong to this type of distribution.

An inspection of curves, 18, 19, 20, 21 will reveal the degree of correlation between the actual grades and what might be expected if they had been distributed in accordance with the normal curve. The

Comparison of the Actual with the Normal Distribution



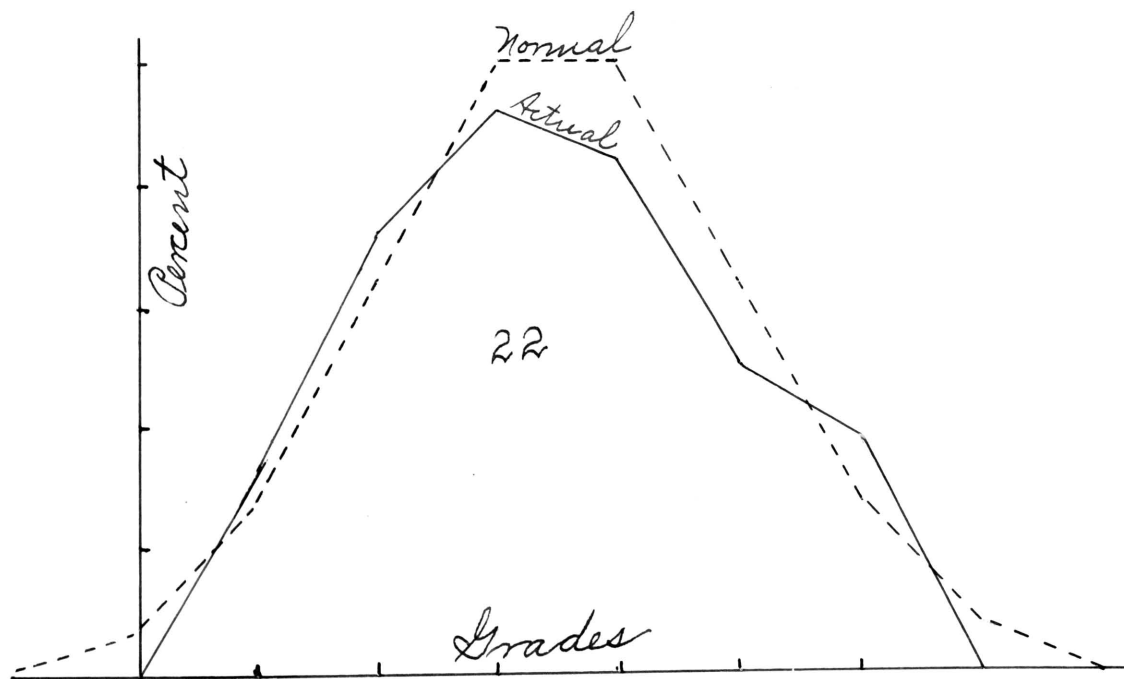
correlation is greatest where the number of students is greatest, but even in sections where the number of students is relatively small the tendency seems to be toward a normal type of distribution. At any rate, it does not seem as if any other curve having a simple equation resembles the actual distribution more closely than does the normal curve. Where the scholarship is of an unusual character as in curve 19 which shows a class of 25 students, all women, the correspondence is not so great; This is to be expected.

Where the extent of correlation has been calculated this was found to be very high, usually above .90 as will be seen by referring to Table Va. This fact is also considered as a justification for the method used in this table, i.e., of assuming that the normal curve represents a typical quiz more nearly than does the result obtained from any one single quiz. This could only be the case if the normal curve represents the limit toward which the form of distribution approaches, when obtained from material of the nature reported in this paper.

A single quiz may approach very closely to the normal curve if the number of blanks is fairly large and the variations in the difficulty of the blanks great enough to allow all members of the class to reach their maximum limit of accomplishment. Curve 22 graphically represents the distribution of such a quiz together with the expected normal curve. The correlation between the two curves is .97.

*Actual and Normal Distribution
of a single quiz*

Correlation .97



Curve 23 shows the actual distribution of all grades awarded during the semester, to about 175 students. Each of the sections of class received nine quizzes, and in all the curve shows 1536 grades. The curve which would be expected under a normal distribution is shown by the dotted line and the correlation between both curves is .996, as near to identity as it is possible to get in mental measurements, and a degree of correspondence which would be acceptable even in an experiment in physics. Of course, such close agreement between theory and practice can only be secured when the methods of grading have been refined to the utmost. It is doubtful whether any other method than the conjectural method will give the differentiation of the various degrees of scholarship with sufficient accuracy to secure such close correlation.

To state the results of this section more concisely: It was found that as the number of individual grades increases, the distribution approaches more and more closely the form of the normal curve. The following data is selected to show this gradual approach toward identity.

Number of Grades					Correlation with Normal			
44730
141970
1536996
Exact correspondence would be					.	.	.	1.000

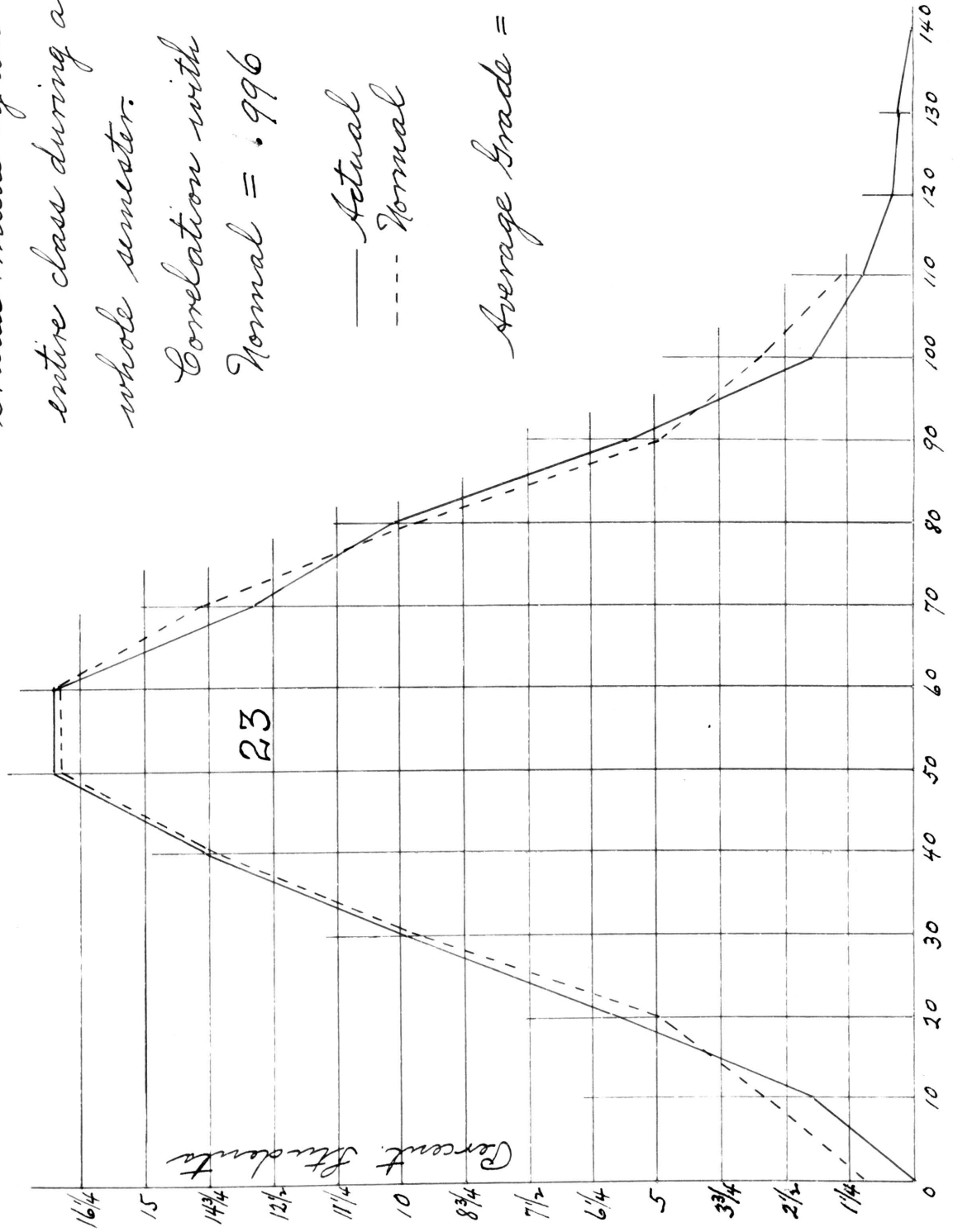
This series justifies at least the conclusion, that of all the curves having a relatively simple equation, the normal curve most

Distribution of the 1536
 Grades made by an
 entire class during a
 whole semester.

Correlation with
 Normal = .996

— Actual
 --- Normal

Average Grade = 50



closely approaches the form of distribution secured from such tests as those reported in this paper and using the methods of grading which have been indicated.

A fact which is sometimes overlooked when trying to estimate the extent to which a given series of grades approaches the normal distribution, is that there is an indeterminate number of normal curves; also, a normal curve is only symmetrical when the mean happens to fall at the beginning of a class division. This is not the case in most examinations. To merely arrange the material into convenient classes and then conclude there is no normal distribution just because the curve secured does not resemble the conventional bell-shaped curve is unsafe. Attention must be given to the variability of the group. The extent to which a series of grades approaches a normal type can only be satisfactorily determined when they have been compared with the corresponding values of the normal probability integral.

Supplementary Remarks

As to the range of the subject matter which can be tested by the Ebbinghaus conjectural method, this depends largely upon the ingenuity of the instructor. At the University of Missouri it is now being used either exclusively or in part in the departments of Experimental and Educational Psychology, Chemistry, and Zoology, and Professor Karapetof of the School of Engineering at Cornell University has expressed himself as well pleased with the method as used in his classes

in electrical engineering. The flexibility of the method could scarcely be better expressed. The method seems best adapted to the large classes in the introductory^{courses} of the natural and social sciences where emphasis is placed on a general rather than on specific knowledge. In a general course in a science such as physics, having a vast body of material with which it is expected to make the student familiar, there is not time to do justice to the quantitative aspect. The instruction is therefore qualitative rather than quantitative. If the tests given are mainly algebraic problems, this qualitative aspect is neglected. A serious study of any science must, of course, be centered around its quantitative phases, yet that class of knowledge which we describe as general is principally qualitative. Its aim is to show the great variety of natural phenomena and at the same time indicate the unity which underlies it.

One criticism that has been urged against the conjectural method is, that it eliminates one of the sources of training in English composition. The rhetorical features of coherence, emphasis and unity in sentences are not expected to reach their maximum effectiveness in an examination. Effective expression in the sense in which we should expect it to manifest itself in an examination, is largely a matter of selecting the proper words. This element is not eliminated by the conjectural method. One can expect that idioms peculiar to the subject matter of the examination be used, and where these are given one has reasonable assurance that the student has prepared his lesson. The conjectural method offers opportunity for a form of training in language which is practically neglected by the ordinary method of examina-

-tion. This is a training in the READING of English. To intelligently fill in a blank a student must know exactly what the sentence means. It is a fact of common experience that unless an examination question is stated in the very simplest language, many of the class will misinterpret it. This tendency is exhibited very clearly in the quiz papers of the conjectural method. It frequently happens that a student will flatly contradict a statement made in one sentence, by the sentence following, thus indicating that he had little idea of what he was reading. It is a matter of surprise to one correcting conjectural quizzes how prevalent is the inability to understand what is read. The training which this method gives in discriminative reading, is one of its greatest advantages.

The nature of the conjectural quiz is such that it requires a definite answer to a definite question. It can be so constructed that unless a student has prepared for this particular phase of the subject, he cannot fill in the blanks correctly no matter how wide his range of general information may be. Nor can a student utilize those questions which he cannot answer as stepping stones to lead on to subject matter with which he is familiar but which has not been called for in the examination.

The conjectural method makes it possible to give large classes frequent quizzes, and the grading of the papers can be left to assistants with a reasonable assurance that no great amount of injustice will be done. To correct 100 papers of fifteen blanks should not take more than two and one-half hours and after some practice, this time is

reduced. A fair test is one in which every student scores, but which is yet difficult enough to prevent the best student from filling all the blanks. This will give the desired range to the grades. For short tests 15 to 20 blanks are sufficient and a quiz of this length may be given during the last fifteen minutes of the hour and the students dismissed as they finish.

The preparation of the context for conjectural quizzes is a simple process. A method which has proved successful in chemistry is to give as part of the quiz the names of the chemical substances to be studied and let the student supply the symbols or formulae. A description of the preparation of some chemical substance in which the significant words have been omitted has also been used. A form of conjectural quiz which lends itself well to the biological sciences is to present a drawing to the student and have him orient and label the various organs and details. Extracts may be taken from the literature at large, or where a text book is used one or more paragraphs may be selected from it. The source or nature of the material does not influence the relative standing of the students to a larger degree than would result from the inevitable accidental variations. To decide which words are to be omitted is more difficult. In general it may be said that any words which the instructor is likely to select as significant will serve, though at first there is a tendency to make the quizzes too difficult. After a few tests have been prepared the instructor learns how to make the quiz as specific or as general as he chooses. It is this element of control which makes the conjectural method so effect-

-ive in general courses, where the tendency to ramble into irrelevancies, vitiates the purpose for which examinations are given.

Summary

Recording the ranks of the student in class is simpler than recording the actual grades made, and shows the relative scholarship more clearly.

Conjectural quizzes vary in difficulty to a considerable extent.

The different blanks also vary much in difficulty, but the relative standing of the students will not be materially changed by all blanks to have the same value in grading.

The fact that the conjectural method does not test a special ability such as rote memory, is shown by:

(1) The grades handed in by the departments of experimental psychology and chemistry where the method is used, compare as well as may be expected, with the grades handed in by the other teachers of the same students where different methods of examination are used.

(2) The correlation between conjectural quizzes and question and answer quizzes, is the same as between conjectural quizzes only, or between class work and laboratory work.

(3) Where the subject matter is taken from a source unknown to the students, their relative standing is unchanged.

(4) Some blanks are not filled in, even though the correct word has already been used.

The grades of the students reduced to terms of the average accomplishment of the class, eliminate the personal equation of the instructor and the ambiguity arising from examinations of varying degrees of difficulty.

A single quiz is not a reliable measure of what the grade of a student will be if the average of a greater number of quizzes is taken.

No significant differences were found between the scholarship of different sections of the same class, or between the sexes.

Little if any support was found for the assumption that the high school acts as a selective agency with respect to scholarship.

The correlations between the grades secured by the conjectural method and the grades expected from the normal curve have a high positive value.

The conjectural method does not limit the training in effectiveness in expression, and it also trains in the comprehension of what is written.

Consistent work is rewarded and since quizzes may be frequently given, sporadic habits of study are penalized.

Time is saved both for the student and the instructor and the grades do not reflect so strongly the effect of accidental factors.

References

Ebbinghaus, H. Ueber eine neue Methode zur Pruefung geistiger Faehigkeiten und ihre Anwendung bei Schulkindern.
Zeitschr. f. Psych. U. Phys. der Sinnesorgane
Vol Xiii p. 401

Meyer, Max. Experiences with the grading system at the University of Missouri. Science Vol. XXXl p. 661, Apr. 28, 1911.

Wiersma, E. Die Ebbinghaus'sche Combinationsmethode
Zeitschr. f. Psych. u. Phys. der Sinnesorgane Vol. XXX p. 196

C. Spearman, The proof and measurement of association between two things. Am. Jour. Psych. Jan. 1904 Vol. XV p. 72

Statistical Methods -- C. B. Davenport

Thorndike, E. L. Mental and Social Measurements.



010-100965582

378.7M71
XW436



