*Library copy
Lent to the University of Missouri*

A STUDY OF THE RELIABILITY OF THE

THORNDIKE HANDWRITING SCALE

BY

ARTHUR STRATTON HURT, B.S. in Ed.

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

IN THE

GRADUATE SCHOOL

OF THE

UNIVERSITY OF MISSOURI

1916.

378.7m71.
XH946

1

254.1916, E. & M.

TABLE OF CONTENTS

Introduction	Page
	3
Chapter I. Study A.	
Section I. Problems of Study A.	7
Section II. Material, personnel and method.	8
Section III. An intensive study of individual variation.	15
Section IV. An intensive study of group variation.	26
Chapter II. Study B.	
Section I. Problem and method of Study B.	35
Section II. Individual variation.	41
Section III. Group variation.	44
Section IV. The effect of two weeks of additional practice.	46
Chapter III. Study C.	
Section I. Problem of Study C.	50
Section II. Material, personnel and method.	50
Section III. Individual variation.	58
Section IV. Group variation.	61
Section V. Miscellaneous.	61
Chapter IV. Some problems.	
Problem I. Do inexperienced graders have a tendency to slight quality ten in grading ?	68
Problem II. In what part of the scale does the greatest variation in grading samples of handwriting occur ?	69
Problem III. How many individual averages or judgments must be combined to secure a trustworthy average ?	75
Problem IV. What is the minimum time of practice required in grading samples of handwriting before a variation low enough to be reliable can be secured ?	83

Problem V. What will be the effect upon variation of many regradings upon the same set of samples (1) separated from each other in point of time and (2) many regradings not separated from each other in point of time but graded on successive days ?	89
Bibliography.	98

Introduction.

The problem of these studies is to determine the usability of the Thorndike Handwriting Scale.

Some investigators, in attempting to make practical use of the scale, have become dissatisfied with it and have attempted to remedy matters by making scales of their own. The following are typical:

¹
Ayres felt that "general form" was too indefinite and broad to be estimated by a single judgment so he constructed a scale to measure legibility by the rate of reading writing. It was accurately constructed on the basis of legibility but in use it measures form as does the Thorndike Handwriting Scale.

²
Freeman also thinks that a sample of handwriting is too complicated for one judgment and that it can be evaluated more accurately by carefully analyzing it into several elements and making a judgment of each element separately instead of depending upon a judgment of the sample as a whole. To carry out his theory he invented five scales one for each of five elements as follows: (1) Uniformity of slant; (2) uniformity of

1. A Scale for Measuring the Quality of Handwriting of School Children, Russell Sage Foundation, New York City.
2. The Teaching of Handwriting. Houghton Mifflin Co. 1914.

alignment or the evenness of letters, top and bottom, in words and sentences; (3) quality of the line or stroke, whether the strokes are smooth, firm and even or tremulous and weak; (4) letter formation, whether open or closed loops, large or small letters, etc. and (5) spacing.

Other investigators have found the scale servicable in measuring handwriting. The following are typical:

¹
Starch has shown that, for grading samples of handwriting, an objective scale is more accurate than are subjective scales.

²
Kelly found that a judge will vary less with himself than with other competent judges.

In determining the practicability of the scale the maximum amount of variation in the estimate of the merit of a sample by the same grader at different gradings and by different graders must be determined. This maximum may be determined on the following basis: The variation in a class average must be less than one half the distance between the average merit of many samples from one grade and the average merit of many samples of the grade above or below since if it is greater than one half the merit the grades will become confused with each other.

The averages of the grades have been obtained by two investigators working independently.

1. The measurement of Handwriting, Journal of Educational Psychology, Volume VI, October 1913, page 455.
2. Teachers' Marks, Teachers College Record, page 107.

1

Charters graded many samples of handwriting of the grades two to eight inclusive from many sections of the state of Missouri and found that the average of the second grade lies in the neighborhood of 7.50 quality on the Thorndike Handwriting Scale while the eighth was approximately 11.00

2

Starch reached approximately the same norms by independent investigation in which he measured samples of handwriting from 4074 pupils in 19 schools in eight large cities located in the three states, Wisconsin, New York, and West Virginia. He found the average of the second grade to be 7.50 qualities and the eighth to be 10.90.

This means an average difference between each grade of approximately .50 of a quality measured by the Thorndike Handwriting Scale.

Variation. Any variation of over .20 of a quality is too large to make grading reliable. If it were known that a grader had a variation of .30 of a quality and he should grade the handwriting of a seventh grade class and find it averages quality 10.00 and grade an eighth grade class and find it to be 10.50 the true values might be 10.30 ($10.00 + .30$) and 10.20 ($10.50 - .30$) and thus the seventh grade which is given by his gradings at 10.00 would in

1. From an unpublished study of the Thorndike Handwriting Scale made during the spring of 1914.
2. The Measurement of Efficiency in Writing, Journal of Educational Psychology, February 1915, page 110.

reality be better than the eighth which he graded at 10.50 and the results would be confused. But since the grader was known to vary .30 of a quality no real difference in merit could be assumed. If however it were known that he varied .20 of a quality or less the grader's estimates would not be more than 10.20 ($10.00 + .20$) for the seventh grade and less than 10.30 ($10.50 - .20$) for the eighth.

It would seem reasonable that an average individual variation of .20 or less will be trustworthy in the practical interpretation of conditions in the typical school room.

Methods for reducing variation. Three methods were used in the ensuing studies by means of which it was sought to reduce variation. The methods are as follows: (1) Individual practice without instruction in the use of the scale; (2) practice with informal instruction in the use of the scale in which the average of a small group of graders was taken as the true measure of a sample; and, (3) practice and comparisons of the values assigned by the judges with the true values of the 50 samples which are given by Thorndike.¹ In addition to the investigation of these methods a number of related problems, growing out of the studies, were analyzed.

1. Teachers' Estimates of the Quality of Specimens of Handwriting, Teachers College Record, November 1914.

CHAPTER I.

Study A.

Section I.

Problems of study A. The object of this study is (a) to determine the variation, after considerable practice, in the judgments (1) of the same individual at different gradings and (2) of different individuals, and (b) to determine, if possible, some effective methods of reducing (1) individual and (2) group variation.

In school practice the problems are: (a), (1) To what extent may a teacher trust his own grading at different times and (2) to what extent may two teachers or superintendents trust each other's gradings when comparing the relative merits of the penmanship graded by each; and (b) to determine, if possible, some practical methods by means of which the variation which occurs may be reduced until the results may be accepted as trustworthy or reliable.

Unfortunately the Thorndike Handwriting Scale was put to a hard test. It was constructed from samples of handwriting selected from grades five to eight and should be used only to measure those grades but in practical school

1. Individual variation means in these studies the variation of any judge in his estimates of the value of a sample or the average of a set of samples.
2. Group variation in these studies means the variation between the judgments of different individuals when judging the same sample or sets of samples.

situations it has to be used for grading handwriting from grades two to twelve inclusive. This should be borne in mind in judging the intrinsic efficiency of the scale in the ensuing studies.

Section II.

For the sake of convenience study A will be evaluated under two heads as (a) individual variation and (b) group variation.

Material graded. In table I may be found a summary of the material graded in study A.

Table I should be read as follows: Set I consists of 73 samples which were graded five times on consecutive days beginning September 25, and the total number of gradings was 5 times 73 or 365. This set was graded again on two consecutive days beginning December 1. Set II consisted of 100 samples and was graded five times on consecutive days beginning November 30 etc.

The material graded, indicated in Table I, consisted of five sets of samples of handwriting. Sets I and II consisted of samples of "best handwriting in ink" while Sets III, IV and V were in "natural handwriting in ink." In each of the first three sets the subject matter consisted of the identical sentence used in the Thorndike scale: "Then the carelessly dressed gentleman stepped

Table I

A summary showing the number of sets, times graded, date of grading and the number of samples in each set.

Number of set	I	II	III	I	IV	V
No.samples in set	73	100	433	73	100	136
No. of gradings	5	5	2	2	5	5
Time of grading	9/25	10/30	11/6	12/1	12/3	1/5
	9/26	10/31	11/7	12/2	12/4	1/6
	9/28	11/2	11/9		12/5	1/7
	9/29	11/3	11/10		12/7	1/8
	9/30	11/4	11/11		12/8	1/9
			11/12			
			11/13			
			11/14			
Total gradings	365	500	866	146	500	680=3057

lightly into Warren's carriage and held out a small card. John vanished behind the bushes and the carriage moved along down the driveway." The subject matter of Sets IV and V was taken from matter familiar to the students but different from the preceding sets.

Set I was selected carefully to obtain as nearly as practicable an equal number of samples of each quality from five to fifteen inclusive, but with a smaller number of samples of both the upper and lower qualities. Sets II, IV and V were selected at random from the second to the twelfth grades inclusive. Set III consisted of samples, one taken from each member of the four classes of a large high school.

Personnel. The graders, five in number, were members of a class in Statistical Studies in the Theory of Teaching in the University of Missouri and were of normal intelligence and ability.

Four graders, numbers 1,2,4, and 5, were inexperienced in the use of the Thorndike Handwriting Scale. Four of the number, 2,3,4 and 5, were experienced teachers three of whom, numbers 2,3 and 4, were former superintendents while number five, a student in the university, was teaching part time in a small junior college. Graders 1,2,and 3 were college graduates while four and five were seniors in the School of Education. It might be mentioned that

Table II

A summary of the number of sets, date of grading and the daily arithmetical averages, of each grader, which are given in the chronological order in which they were graded.

Set	I					II			III				I		IV				V					
Date of grading	9/25	9/26	9/28	9/29	9/30	10/30	10/2	10/3	10/4	10/5	10/6-10	/11-14	12/1	12/2	12/3	12/4	12/5	12/7	12/8	1/5	1/6	1/7	1/8	1/9
Judge number 1	8.10	7.71	9.08	7.65	9.04	9.67	8.86	9.41	9.52	9.22	10.51	11.13	9.77	9.66	10.13	10.22	10.13	9.91	9.86	7.89	7.40	7.97	7.99	7.99
2	10.30	10.20	9.33	10.40	10.40	8.95	9.04	9.36	8.68	9.26	11.19	11.02	8.87	10.05	9.55	10.03	9.74	9.52	9.56	7.71	7.62	7.96	7.83	7.81
3	9.63	9.79	9.82	9.86	9.85	9.21	9.17	9.14	9.14	9.16	11.12	11.22	9.59	9.60	10.12	9.54	9.64	9.52	9.55	8.04	8.16	8.07	8.07	8.16
4	9.78	10.07	9.79	9.41	9.68	10.57	10.29	8.91	9.42	9.49	11.15	10.89	9.73	9.67	10.06	10.10	10.05	10.11	10.07	9.09	8.69	8.65	8.37	8.24
5	8.63	9.01	8.33	8.43	8.73	8.23	8.13	8.22	8.67	8.67	11.14	11.39	9.18	9.12	9.36	9.27	9.47	9.71	9.62	8.59	8.69	9.20	8.63	8.81
Group average	9.31	9.36	9.39	9.15	9.54	9.33	9.10	9.01	9.09	9.16	11.02	11.13	9.43	9.62	9.84	9.84	9.80	9.75	9.73	8.26	8.11	8.37	8.18	8.20
Group range	2.20	2.49	1.50	2.75	1.67	2.34	2.16	1.19	.85	.82	.63	.50	.90	.93	.77	.95	.66	.59	.52	1.38	1.29	1.25	.80	1.00
Average deviat.	.734	.796	.528	.838	.564	.636	.506	.354	.328	.196	.206	.140	.320	.200	.312	.350	.226	.204	.186	.460	.482	.446	.258	.253
Percent of variability	7.88	8.50	5.62	9.70	5.90	6.81	5.56	3.92	3.60	2.14	1.87	1.26	3.39	2.08	3.17	3.66	2.31	2.10	1.91	5.56	5.94	5.53	3.15	3.14

the students were not passive students of learning in an experiment to reduce variation but had a vital interest in the problem of standardization of grading.

Number and frequency of grading. Each judge graded a set of samples a day with one exception. Set III, consisting of 433 samples, was divided into four groups for convenience in grading and one group was graded each day. By this method it took four days to grade the set once or eight days to grade the set twice.

Referring to Table I it will be seen that but thirty days, during the semester, were spent in grading samples of handwriting. Each judge made a total of 3057 gradings during the semester judging 915 different samples.

Table II gives a summary of the number of sets which are given in the chronological order in which they were graded, the daily arithmetical averages of each set by each grader and the date of each grading.

In the first horizontal column, at the top, are given the numbers of the sets. Just below are given the dates of the respective gradings. In the first vertical column, from the left, may be found a number for each grader. In the second vertical column from the left and at the top will be found the date September 25 and immediately below are the figures 8.10 which means that on September 25 judge number one graded Set I and found the average to be

the 8.10 quality on the Thorndike Handwriting Scale. Immediately below, in the same vertical column, will be found the number 10.30 which means that judge two graded Set I on the same date and found his average to be 10.30 qualities. In the same way the next three numbers of the vertical column may be interpreted. Just below the individual averages for September 25 will be found the group average for that day which was 9.31. Just below, in the ninth horizontal column, will be found the number 2.20 which is the group range for that day and means that on that particular day when five judges graded Set I judge one found his daily average to be 8.10 while judge number two found his average to be 10.30 or a difference of 2.20 of a quality. In the same vertical column, at the bottom, will be found the figures 7.88 which means that on September 25 the percent of variability was 7.88. The percent of variability is obtained by dividing the average deviation for that grading, in this case it is .734, by the group average for that day. In the same way the vertical column to the right may be interpreted.

Plan of selecting data from which conclusions will be drawn. The average was chosen as a basis of comparison because it is ordinarily used in school work when the merits of classes and schools are compared.

In individual variation conclusions will be drawn from the variation which occurs between the individual averages of the first and second gradings of each of the following sets: (a) Sets I and II which represent gradings made by inexperienced judges and (b) Sets IV and V which represent gradings made by judges after considerable practice in the use of the scale. The number of consistent judges on both the first two sets will be compared with the number of consistent judges on both of the last two sets.

In group variation conclusions will be drawn from the average group range of the daily individual averages of the first two gradings of each of the following sets: (a) Sets I and II which represent gradings made by inexperienced judges and (b) Sets IV and V which represent gradings made by judges after considerable practice in the use of the scale. A comparison of the average group range on the first two gradings each of the first two sets will be compared with the average group range on the first two gradings of each of the last two sets.

Section III.

An intensive study of individual variation.

Problems. (a) To determine the variation, after considerable practice, with instruction in the use of the scale, in the judgments of the same individual at different gradings and (b) to determine, if possible, some effective methods of reducing this variation.

The main object of this study was to determine if individual variation might be reduced within reliable limits by practice and instruction in the use of the scale.

Method. (a) General procedure used in the investigation of both individual and group variation.

In grading samples of handwriting only whole qualities were used in this study.

A regular schedule of grading was followed. For instance, at nine o'clock judge number one graded the samples and at ten judge number two graded the same set.

To prevent confusion, when grading, the samples were stacked, face up, and numbered on the front in a conspicuous place with Arabic numerals.

When regrading a set of samples no student was permitted to see any of the previous gradings that he had made on the set.

There was to be no exchange of opinion while the judges were actually grading a set of samples but the

judges were encouraged to have conferences at other times.

When regrading a set, if a grader remembered the value that he had given a sample on a previous day, he was to indicate the fact by drawing a ring around the value assigned and omit the grade from all his averages but this happened too infrequently to appreciably affect the results.

The individual daily practice was supplemented occasionally by a class conference in which problems, arising in the use of the scale, were discussed and plans devised for their possible solution. Some problems were solved by arbitrary definitions, other cases were settled by class agreement after a general discussion. An example of the former is the following: Some difficulty was encountered in estimating the value of a sample when a part, as a line, would be much better than the remainder. Since most of the samples graded consisted of "best handwriting in ink" it was decided to estimate the merit of the sample from the better part.

During the semester conferences were held on the following dates: Monday, November 2, 1914., Monday, November 9, Thursday, December 3, Monday, December 7, and Thursday, January 7, 1915.

It was found that inexperienced judges have much difficulty in using quality ten on the scale for it rarely happens, in actual practice, that one will find a sample similar to the one used as quality ten. To remedy this difficulty the ascending-descending procedure for finding quality ten was used as follows: If the value of any sample being graded is in the neighborhood of qualities 9, 10 or 11 of the scale the sample was treated in the following manner: Compare the sample with the scale beginning with a quality poorer than the sample and advance until the first quality is found that is undoubtedly better than the sample. Make a note of this value. Next begin with a quality that is better than the sample to be graded and pass down the scale until the first quality is found that is undoubtedly worse than the sample. Make a record of this value also. If the former is found to be quality 11 and the latter quality 9 the sample is to be evaluated as quality 10.

The individual graders kept a record of their grading and in it incorporated the following data:

- (1) The number of the set being graded;
- (2) the date of the grading;
- (3) the time actually spent in the grading;
- (4) the grader's name;
- (5) the number of each of the separate samples, and
- (6) the value given each on that particular day to be written opposite the number of

the sample. These records were preserved and compiled in Tables I and II.

(b) Specific procedure used in investigating individual variation and supplementing the general procedure mentioned in the preceding section.

After each member of the class had graded a set of samples several times the instructor would indicate a sample and request each member of the class to state orally his separate judgments of the sample. If it were found that any grader fluctuated much in his judgments the instructor would request him to compare the sample carefully with the scale and make a new estimate of the quality of the sample. After this the instructor would require him to give the reasons why he gave it the value which he indicated. If his judgment of the merit of the sample seemed poor or fluctuating or if several judges had a large individual variation on the sample the class would discuss the merits and demerits of the sample and each judge would grade the sample in turn. It was hoped by this method to find the true value.

This method was used at the following conferences:
Monday, November 2, 1914, Monday December 7, and Thursday,
January 7, 1915.

Qualifications and criticisms. One question which arose that was not settled by this study was: Would better results have been secured if the judges had graded samples once every day or once every other day?

Any distracting influences, as conversation, while a judge was grading a set of papers had a demoralizing effect upon the grading. A few cases of large variation were traced directly to this source.

Some of the graders experienced some difficulty in maintaining sufficient interest in the grading to do the work effectively at all times and would occasionally grade a set hurriedly and carelessly. The effects, in a few cases, were quite obvious when the grades were evaluated.

There is a question whether the method of single judgments on each sample is as effective in grading as the ascending-descending procedure. No attempt has been made in this study to determine the better method except in the case of quality ten.

With one exception, that of Set I, the sets of samples graded were not selected carefully so as to include specimens requiring the use of the higher qualities of the scale.

In grading very few samples were judged higher than quality twelve. If these were the true values of the samples they were not chosen wisely since, for best

practice, the samples should be chosen carefully to obtain about an equal number of samples of each quality from five to fifteen between which ranges practically all school room products of penmanship fall. A well chosen set would require the use of the entire compass of the scale.

Not all the time was devoted to grading samples of penmanship but periods of weeks elapsed in which no grading took place. During these periods unrelated subject matter as spelling and reading was studied. When practice was resumed it was noticed that the variation was usually greater than it was at the close of the last practice.

Enumeration of the characteristics observed in determining the merits of samples of handwriting.

To grade effectively a judge must estimate the value of the sample taken as a whole. Inexperienced graders rarely ever do this but try to find a specimen on the scale which most closely resembles the sample being graded and then compare certain elements as spacing, slant, etc. It was observed that a grader emphasized one element of the sample at a time. This would be in the focus of attention but, as grading continued, this element became more obscure and eventually

passed below the conscious level. This left the judge's mind free to attend to other elements. For this reason a judgment of a sample by an experienced grader is more habitual than one by an inexperienced grader. Experienced graders do not depend upon general impression alone, For instance, when a sample was found, the first impression of which was not satisfactory, thus leading to vacillation in judgment, the grader looked for some element, as alignment, to confirm his judgment.

A record was kept of the reasons and factors that were uppermost in each grader's mind when he made his judgments and at the close of the semester the data was organized and condensed in the following form:

Characteristics to be observed in determining the merit of samples of handwriting.

- (a) Alignment. The letters should be even, above and below, in words and sentences.
- (b) Uniformity. There should be uniformity in the form of letters as
 - (1) the slant of each of the letters should be the same and
 - (2) the stem of each of the letters should be of uniform shape.
- (c) Spacing or the distance between the letters and words should be
 - (1) uniform and of
 - (2) normal distance.

It was observed that correct spacing, alignment, regularity of slant, neatness and normal height were highly correlated with merit.

On the contrary it was found that judges paid attention to the factors enumerated above and it made it difficult to judge the sample as a whole or by general impression.

These factors found early in the semester agree rather closely with those later stated by Freeman as follows: (a) Uniformity of slant; (b) uniformity of alignment; (c) quality of the line; (d) letter formation and (e) spacing.

1. The Teaching of Handwriting, Houghton Mifflin Company, 1914.

Conclusions.

Individual variation. Table III was compiled from the first two gradings each of Sets I and II and represent gradings made by judges unpracticed in the use of the scale with the exception of judge number 3 who was experienced.

Table III

A summary of the individual averages of the first two gradings each of Sets I and II.

Number of grading	Set I			Set II		
	1st	2nd	Ind. var.	1st	2nd	Ind. var.
Judge #1	8.10	7.71	.39	9.67	8.86	(.19)
2	10.30	10.20	(.10)	8.95	9.04	(.09)
3	9.68	9.79	(.11)	9.21	9.17	(.04)
4	9.78	10.07	.29	10.57	10.29	.28
5	8.68	9.01	.33	8.23	8.13	(.10)

It will be observed that on the two gradings of Set I two of the judges, two and three, had an individual variation of less than .20 of a quality. On Set II the same judges were consistent and maintained the desired low variation. They were the only judges that maintained a trustworthy variation on both sets.

Table IV, compiled from Sets IV and V, seem to be typical of gradings made by experienced judges for each

member of the class had made 1877 separate gradings before evaluation Set IV and may be said to be practiced in the use of the scale.

Table IV.

A summary of the individual averages of the first two gradings each of Sets IV and V.

Number of grading	Set IV			Set V		
	1st	2nd	Ind. var.	1st	2nd	Ind. var.
Judge # 1	10.13	10.22	(.09)	7.89	7.40	.49
2	9.55	10.08	.53	7.71	7.62	(.09)
3	10.12	9.54	.58	8.04	8.16	(.12)
4	10.06	10.10	(.04)	9.09	8.69	.40
5	9.36	9.27	(.09)	8.59	8.69	(.10)

On Set IV three of the five judges, 1, 4, and 5, had an individual variation of less than .20 of a quality. On Set V three of the judges, 2, 3 and 5, came within the desired limits their variation being much less than .20 of a quality. But one judge, number 5, was consistent on both sets.

It is a significant fact that in Table III judges 2 and 3 were consistent on both sets but that in Table IV neither had a trustworthy variation on both sets. These judges did not seem to have been benefitted by the

attempt to reduce individual variation by practice plus instruction in the use of the scale but to have become more inconsistent. The ratio of consistent judges in Table III to the number engaged in grading was two to five or .40. The ratio of consistent judges in Table IV to the number engaged in grading was one in five or a ratio of .20.

A comparison of the ratio of the consistent judges in Table III with the ratio of the consistent judges in Table IV, composed of gradings made after a month of practice in the use of the scale, reveals the fact that there is a decrease.

It seems reasonably safe to assert that practice plus instruction in the use of the scale does not reduce individual variation.

A comparison of the individual variation of the four sets of samples used in Tables III and IV reveals the fact that judge number 5 was inconsistent only on the first set graded or Set I. Judges 2, and 3 were inconsistent only on Set IV. The data shows that no judge was within the practical range of reliability on all the sets but was reliable on some sets and on some was not.

Section IV.

An intensive study of group variation.

Problems. (a) To determine the variation, after considerable practice with instruction in the use of the scale, in the judgments of different individuals in the use of the Thorndike Handwriting Scale and (b) to determine, if possible, some effective methods of reducing this variation.

Material. The material for this study is found in Table V given below on page 33 which like Tables IV and III is an abstraction from Table II above.

Explanation of group variation. The range of the individual average on the second grading of Set V, Study A (see Table II above) after five weeks of practice and 2500 gradings was 1.29 of a quality. Judge number 1 graded the series of samples and found his average to be 7.40 qualities while judge number 4 graded the same set under similar conditions on the same day and found his average to be 8.69 qualities, a difference of 1.29 qualities or over twice the difference between the averages of the merit of two consecutive school grades.

If these judges had been grading samples from the same grade in different schools and should find as much difference as a whole quality in their averages there would be no ground for assuming that any difference in merit existed or that the writing of any grade in one

school was better or worse than that of corresponding grade in another. Thus the group range is so great that any comparison of the averages of sets of samples of handwriting, taken from the same grade but different schools, would be unreliable and misleading. This is all the more remarkable when the fact that the graders concerned had already graded 2500 samples, is taken into account.

Method. (a) General procedure used in the investigation of both individual and group variation. See Section II pages 8-14 . (b) Specific procedure used in investigating group variation and supplementing the general procedure mentioned in the preceding section.

Briefly stated the method to reduce group variation was group conformity and will be treated somewhat in detail under four subheads as

- (1) Discussion of problems.
- (2) Comparison of gradings made before a class conference.
- (3) Class grading.
- (4) Arbitrary definitions and rulings.

During this study five conferences were held the dates of which are given on page 16 . No conference was held before the grading of a new set so in all cases, at conference, all members of the class had made one or more gradings on a set of samples which might be used for

comparison.

Comparison of gradings made before a class conference.

The instructor, during the conference, would first call for the values assigned by the different graders to certain samples which he indicated. The comparisons revealed the samples upon which the members of the class varied most widely.

Class Discussion. The graders varying most widely were asked why they assigned the values which they had stated. Each sample, upon which the members of the class varied much, was examined carefully and its merits and demerits were discussed with the hope of thereby finding the cause of the variation. It was frequently found that one or more elements of the sample such as height, spacing, etc., were abnormal, but this was not always the case. For instance, during the conference of November 2 judge number 4 whose grading was unusually high was asked why he assigned such high values to certain samples. He replied, "because they were written by the muscular system." In this case the grader's prejudice played an important but undesirable part since the previously established standard of merit in the grader's mind frustrated the attempts to reduce group variation. Such cases were frequently settled by arbitrary rulings. In this case a ruling was made, for future work, that the

system or method of writing was not to be considered in estimating the merit of a sample of penmanship.

Class grading. In case the class could not agree upon a value for any particular sample the instructor would request each member of the class to pass in turn to the scale with the sample and silently make an estimate of its merit. These values were afterwards collected by the instructor, compared and discussed. It was observed that there was less group variation in grading after a conference than before.

There is a form of variation, a difference in degree, which is hard to detect unless a grader's average is checked against the grand average of many graders on the same material. This grader may not vary much on any one sample but may grade uniformly high or low as the case may be. In this study this type was detected by comparing each grader's average with that of the average of the group on the same material. Applying the principle that the average judgment of many graders is of more weight than that of any one, those graders whose averages were uniformly low were encouraged to grade higher while those that graded higher than the average of the group were encouraged to lower their estimates. The following concrete cases are typical: During the conference

of November 2 it was observed that the daily average of judge number 4 was much higher than the remaining members of the group. He was asked to attempt to conform more closely to the group average. The gradings of Set II for November 3 and later show that the judge made a decided improvement in the effort to reduce the variation. This is shown very clearly in Graph I next page. The daily average of judge number 5 was much below the average of the group and he was told to grade higher. Graph I shows a gradual improvement in the later grading.

Qualifications and criticisms. The qualifications and objections stated in Section III apply to this study as well. The following additional objection applies to group variation alone.

The instruction, received by the judges at the conferences, had an immediate effect in reducing group variation but it was not very lasting. In every case one or more gradings had been made on a set before a class conference. In order to obtain the best results it seems as if the instruction should be given (a) frequently and (b) just before the first grading of a new set of samples.

Graph I.

Graph showing variability based upon the daily
arithmetical averages of Set II, Study A.

Conclusions. Sets I and II, Table V, give results that seem typical of unpracticed judges while the results of the judgments on Sets IV and V, the last gradings of the semester, seem typical of judges with considerable practice. The first two gradings each on Sets I and II, and Sets IV and V will be made the basis of our comparisons to determine the effect upon group variation of practice plus instruction in the use of the scale.

Group variation in these studies may be measured by (a) group range, (b) average deviation and (c) percent of variability. In this study group range will be used.

Table V should be read as follows: The group range on the first grading of Set I was 2.20. On the second it was 2.49 etc. The average of the group ranges on the first two gradings each of Sets I and II or the average of four days is 2.30 qualities. The corresponding average on the first two gradings each of Sets IV and V is 1.10.

It will thus be seen that group variation is reduced by practice plus instruction in the use of the scale.

While the group range was reduced by the method used in this study yet it was never so small that the gradings could be accepted as trustworthy and reliable.

It is an interesting fact that no average of a single grader was within .10 of a quality of the group average

Table V.

A summary of the individual arithmetical averages
of the first two gradings each of Sets I,II,IV and V.

No.of grading Judge no.	Set I		Set II		Set IV		Set V	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
1	8.10	7.71	9.67	8.86	10.13	10.22	7.84	7.40
2	10.30	10.20	8.95	<u>9.04</u>	9.55	10.08	7.71	7.62
3	9.68	9.79	<u>9.21</u>	<u>9.17</u>	10.12	9.54	8.04	<u>8.16</u>
4	9.78	10.07	10.57	10.29	10.06	10.10	9.09	8.69
5	8.68	9.01	8.23	8.13	9.36	9.27	9.59	8.69
Group average	9.31	9.36	9.33	9.10	9.84	9.84	8.26	8.11
Group range	2.20	2.49	2.34	2.14	.77	.95	1.38	1.29
A.D. % of	.73	.79	.64	.51	.31	.35	.46	.48
variability	7.88	8.50	6.81	5.56	3.17	3.66	5.56	5.94

on the first two gradings of Set I. On Set II grader 3 on the first grading and 2 and 3 on the second were within .10 of a quality of the group average on those days. On Set IV and the first grading of Set V not a single grader came within the desired limits. On the second grading of Set V grader 3 came within the limits of .10 of a quality.

This study reveals the fact that some judges may learn to use the scale much more proficiently than others. Judge number 2, Table V, was very erratic in his judgments. He graded about a quality higher than the group average on Set I, he graded lower than the group average on Set II, lower than the group average on the first grading of Set IV, Table V, and higher on the second. On Set V he graded lower than the group average.

In Table V and those following which make a study of group variation when a grader's average is within .10 of a quality of the group average for that day the fact will be indicated by a red line drawn beneath the grader's average.

CHAPTER II.

Study B.

Section I.

Study B was made in the Summer Session of 1914 before Study A was made. It was different from Study A in the respect that the method was practice without instruction. The material used included that used in Study A and some additional material found in Sets VI to VIII.

Problem of Study B. The object of the study is to determine the variation, after considerable practice, in the estimates (a) of different individuals in the use of the Thorndike Handwriting Scale and (b) of the same individual at different gradings, without instruction in the use of the scale.

Method. The method briefly stated was individual practice without instruction in the use of the scale. As in Study A only whole qualities were used but differing from the preceding study the judges were not permitted to discuss their gradings with each other with one exception. On July 1 and 2 judges 1,4,5,8,10 and 11, who were at the extreme ends of the curve in group variation, were given instruction in the use of the scale with the hope of thereby reducing the variation.

As in Study A a regular schedule of grading was followed and if a judge remembered one day the grade that he had given a sample on a preceding day he was to indicate the fact by drawing a ring around the grade and omit the value from all his averages.

Material graded. The material graded consisted of eight sets of samples of handwriting. A summary of the number of times and dates of gradings and the number of samples graded in each set is given in Table VI.

Table VI may be read as follows: Set I consisted of 73 samples and was graded five times on consecutive days beginning on June 16. This set was graded again on two consecutive days beginning on July 1, again on July 22 and 23, and on August 4, and 5. Set II consisted of 100 samples and was graded on four consecutive days beginning June 22. In the same way the remaining portion of the table may be interpreted.

Nature of the material graded. The first five sets graded in Study B were the same as the sets described in Study A, Section II. Sets VI and VII were composed of samples of "best handwriting in ink" while Set VIII was written in "best handwriting with leadpencil."

The subject matter, the material written on the body of the sample, of Set VIII was the identical sentence used on the Thorndike Handwriting Scale, "Then the carelessly dressed gentleman stepped lightly

Table VI.

A summary of the number of sets, times graded,
date of grading and the number of samples in each set.

Set	I	II	III	I	IV	V	VI	I	VII	VIII	I
No. samples in set	73	100	433	73	100	136	96	73	116	128	73
Number of gradings	5	5	1	2	5	5	5	2	4	4	2
Week beginning	6/16	6/23	6/30	7/2	7/6	7/14	7/20	7/22	7/28	8/1	8/4

into Warren's carriage and held out a small card. John vanished behind the bushes and the carriage moved along down the driveway." The subject matter of Sets VI and VII was different from Set VIII but was taken from material familiar to the students.

Table VII gives a summary of the number of sets of samples given in the chronological order in which they were graded, the daily arithmetical averages of each set by each grader and the dates of each grading.

The table may be read as follows: The first horizontal column at the top gives the number of the set. The next below gives the date of each grading on each set. The first vertical column to the left gives the numbers corresponding to the names of the students arranged alphabetically after the first three. The second vertical column from the left may be read thus: On June 16 student number one graded Set I and found his average of the 73 samples to be 9.51 qualities on the Thorndike Handwriting Scale. Just below the average of number one will be found 9.59 which means that on the same day student number 2 graded the same set and found his average to be 9.59 qualities. In the same way the individual averages below may be read. To the right of 9.51 or the average of student number one on June 16 will be found the figures 9.32 which mean that on June 17

Table VII.

A summary of the sets, dates of gradings and
individual averages of each of the eleven judges of

Study B.

	Set I				Set II				Set III	Set I		Set IV				Set V				Set VI				Set I		Set VII				Set VIII				Set I						
Date	6/16	6/17	6/18	6/19	6/20	6/23	6/24	6/25	6/26	6/30	7/2	7/3	7/6	7/7	7/8	7/9	7/10	7/14	7/15	7/16	7/17	7/18	7/20	7/21	7/22	7/23	7/24	7/22	7/23	7/28	7/29	7/30	7/31	8/1	8/3	8/4	8/5	8/4	8/5	
Judge #	1	9.51	9.34	9.05	9.25	9.41	9.06	8.97	8.93	9.00	11.63	9.32	9.71	9.57	9.65	9.89	9.91	9.62	8.14	8.34	8.41	8.40	8.25	8.63	8.46	8.65	8.59	8.32	9.55	9.30	8.66	8.50	8.43	8.47	9.70	9.90	9.69	9.83	9.35	9.30
	2	9.59	9.36	9.22	9.42	9.38	9.50	9.17	9.37	9.31	11.35	9.48	9.35	9.46	9.62	9.53	9.54	9.47	7.82	7.68	7.64	7.68	7.65	7.90	7.82	7.73	7.76	7.80	9.35	9.38	9.47	9.32	9.47	9.46	8.86	8.80	8.83	8.85	9.21	9.30
	3	10.03	9.85	9.62	9.63	9.33	----				11.75	9.41	9.10	9.79	9.81	9.69	9.77	9.72	9.03	8.81	8.52	8.86	8.89	8.65	8.95	9.03	9.19	9.08	9.71	9.67	8.15	8.14	8.21	7.99	9.55	9.48	9.48	9.53	9.32	9.37
	4	10.34	9.62	9.96	9.44	9.65	8.26	8.24	8.21	8.21	11.71	10.17	9.81	10.07	9.91	10.11	10.10	10.10	7.98	8.04	8.13	7.84	7.78	8.92	8.19	8.24	8.16	8.09	8.88	9.19	9.31	9.18	9.22	9.00	9.19	9.08	9.04	9.14	9.34	9.03
	5	8.97	9.00	8.98	9.32	—	9.14	9.12	9.20	9.49	11.31	8.95	8.95	9.26	8.93	8.72	9.56	9.09	7.50	7.60	8.15	7.92		8.55	8.30	8.26	8.36	8.43	8.74	8.56	8.92	9.16	9.00	8.86	9.10	9.07	9.10	9.46	8.71	8.68
	6	9.65	9.74	9.70	9.63	9.21	9.31	9.27	9.27	9.30	11.88	9.61	9.32	9.69	9.45	9.04	9.19	9.22	8.23	8.14	7.96	7.83	7.62	8.60	8.13	7.87	8.04	8.00	9.40	9.37	8.95	8.90	9.25	9.20	9.75	9.48	9.78	9.79	9.51	9.50
	7	10.19	10.49	9.84	9.91	10.26	10.09	10.35	10.60	10.62	11.89	10.04	9.46	9.00	8.86	8.79	8.98	8.95	7.61	7.63	7.62	7.42	7.49	7.96	8.25	8.43	8.78	8.30	8.92	9.13	9.95	9.44	9.47	9.45	9.72	9.74	9.32	9.69	8.71	8.99
	8	9.85	9.48	9.75	9.21	9.38	9.10	9.93	9.06	9.00	10.89	8.69	9.38	9.36	9.70	9.74	9.58	----	8.16	7.77	7.87	7.71	7.49	8.47	8.37	8.16	7.89	7.96	9.37	9.05	9.50	9.17	9.36	9.13	9.95	9.44	9.42	9.65	8.71	8.59
	9	9.39	9.80	9.65	9.68	9.68	8.01	8.26	8.06	8.28	11.93	9.49	9.44	9.04	9.24	9.12	9.38	9.22	8.10	7.90	8.30	7.80	7.70	8.01	8.60	8.10	8.28	8.26	9.18	9.09	10.13	10.13	10.15	10.18	9.46	8.75	8.91	9.15	8.69	8.61
	10	9.02	8.72	8.65	8.79	8.75	8.38	8.04	8.76	8.82	11.32	9.60	9.52	9.95	9.92	9.98	9.77	9.88	9.04	9.00	8.95	9.05	8.99	9.17	9.31	9.18	9.10	9.12	9.71	9.78	8.54	8.70	8.56	8.68	10.08	10.07	10.08	10.08	9.68	9.68
	11	10.76	11.02	10.47	10.76	10.24	10.14	10.13	9.94	9.90	10.72	9.54	9.08	8.07	7.98	7.90	7.75	7.60	7.68	8.39	8.34	7.96	8.01	8.01	8.01	8.99	8.03	8.05	9.07	9.05	9.12	8.75	8.88	8.85	9.35	9.28	9.20	9.28	8.81	8.96
Group average		9.76	9.67	9.49	9.59	9.53	9.10	9.17	9.14	9.21	11.57	9.53	9.37	9.39	9.37	9.32	9.42	9.28	8.12	8.14	8.18	8.04	7.99	8.44	8.39	8.42	8.38	8.32	9.26	9.23	9.12	9.05	9.09	9.05	9.50	9.37	9.37	9.50	9.09	9.09
A.D. of variab.		.436	.292	.420	.342	.343	.537	.602	.490	.470	.305	.290	.190	.400	.449	.550	.525	.519	.366	.331	.303	.402	.439	.344	.373	.372	.341	.313	.276	.242	.410			.450	.301		.3021	.290	.335	.307
		4.46	3.02	4.42	3.50	3.60	5.90	3.56	5.40	5.10	2.60	3.10	2.00	4.20	4.79	5.91	5.67	5.59	4.50	4.68	3.70	5.00	5.49	1.00		4.43		3.70	2.98	2.62	4.90			4.97	3.15		3.22	3.00	3.68	3.37

student number 1 graded Set I as on the preceding day and found his average of judgements to be 9.32. In the same way the remaining averages may be read.

In the last horizontal column, at the bottom, will be found the percent of variability for each grading. The percent of variability is obtained by dividing the daily average deviation, given in the horizontal column just above, by the average of the group for that day. This average for the group is given in the third horizontal column numbering from the bottom.

Each judge made a total of 4393 gradings during the seven weeks of practice. Judges 1,2, and 3 had had considerable previous experience in the use of the Thorndike Handwriting Scale but the others had had none.

Personnel. The eleven judges,grading the eight sets of samples of handwriting, composed the members of a class in Statistical Studies in the Theory of Teaching during the summer session of 1914 in the University of Missouri. All the members of the class were experienced teachers six of whom were graduate students. There were four city superintendents, one high school principal, five high school teachers three of whom were teacher-training instructors and the eleventh member of the class was an eighth grade teacher.

Frequency of grading. Each member of the class graded a set of samples each school day for seven weeks.

There was one exception. Set III, consisting of 433 samples, was divided into four groups for convenience in grading and one group was graded each day. By this method it took four days to grade the set once. This grading was done on June 30, July 1, 2 and 3. On July 2 and 3 Set I was regraded.

Plan of selecting data from which conclusions will be drawn. In order to compare the value of the methods used in Study A and Study B the results obtained in grading Sets I, II, IV and V in Study B will be used. To get the effect of two weeks of additional practice the results obtained in grading Sets VII and VIII will be given.

Section II.

Individual variation. To determine the effect upon individual variation of practice without instruction in the use of the scale.

Evaluation of the first two gradings each of Sets I and II of Study B: Four judges 1, 3, 5, and 6, had an individual variation of less than .20 of a quality on Set I being respectively .19, .18, .03, and .09 of a quality. On Set II six judges, 1, 6, 4, 5, 10, ^{and 11} respectively .09, .02, .02, .04, .14 and .01 of a quality. But three judges, 1, 5 and 6 were consistent on both sets. The ratio of the consistent judges on both sets to the number grading was three to eleven or .27.

Table VIII.

A summary of the individual averages of the first
two gradings each of Sets I, II, IV and V.

Number of grading	Set I			Set II			Set IV			Set V		
	1st	2nd	Ind. var.	1st	2nd	Ind. var.	1st	2nd	Ind. var.	1st	2nd	Ind. var.
Judge # 1	9.51	9.32	(.19)	9.06	8.97	(.09)	9.57	9.65	(.08)	8.14	8.34	(.20)
2	9.59	9.36	.23	9.50	9.17	.23	9.46	9.62	(.16)	7.82	7.68	(.14)
3	10.03	9.85	(.18)	-----	-----	--	9.79	9.81	(.02)	9.03	8.81	.22
4	10.34	9.62	.72	8.26	8.24	(.02)	10.07	9.91	(.16)	7.98	8.04	(.06)
5	8.97	9.00	(.03)	9.14	9.12	(.02)	9.26	8.93	.33	7.50	7.69	(.19)
6	9.65	9.74	(.09)	9.31	9.27	(.04)	9.69	9.45	.24	8.23	8.14	(.09)
7	10.19	10.49	.30	10.09	10.35	.26	9.00	8.86	(.14)	7.61	7.63	(.02)
8	9.85	9.48	.37	9.10	9.93	.83	9.36	9.70	.34	8.16	7.77	.29
9	9.39	9.80	.41	8.01	8.26	.25	9.04	9.24	(.20)	8.10	7.90	(.20)
10	9.02	8.72	.30	8.38	8.24	(.14)	9.95	9.92	(.03)	9.04	9.09	(.05)
11	10.76	11.02	.26	10.14	10.13	(.01)	8.07	7.98	(.09)	7.68	8.39	.71

It will be noticed that on Set IV eight judges, 1,2,3,4,7,9,10 and 11, had an individual variation of less than .20 of a quality, being respectively .08,.16,.02,.16,.14,.20,.03 and .09 of a quality. On Set V eight judges 1,2,4,5,6,7,9 and 10 had a variation of less than .20 of a quality.

Six judges,1,2,4,7,9 and 10, were consistent on the first two gradings each of both sets. The ratio of the consistent judges to the number grading was six to eleven or .55.

There were three consistent judges on Sets I and II and six on Sets IV and V which were graded after a month of practice.

Since these results seem typical of the study,we may summarize by stating that practice, without instruction in the use of the scale, will reduce individual variation.

In Study A the number of consistent judges on the first two gradings each of Sets I and II was more than on the corresponding gradings of Sets IV and V. The method of Study A was not effective in reducing individual variation.

A comparison of the methods of the two studies reveals the fact that practice without instruction in the use of the scale, the method used in Study B, is more efficient in reducing individual variation than practice plus instruction in the use of the scale which

was the method used in Study A.

The probable reason for this difference is the fact that in Study A each grader's judgments were temporarily upset in an effort to reduce the group variation.

Section III.

Group variation. The problem in this section is to determine the effect upon group variation of practice without instruction in the use of the scale.

Sets I and II give results that may be termed typical of unpracticed graders while the results of the judgments on Sets IV and V seem typical of judges of considerable practice in the use of the scale.

Group variation in these studies may be measured by (a) group range, (b) average deviation and (c) percent of variability.

In this study the average group range of the first two gradings each of Sets I and II, or the average group range of four gradings, will be compared with the average group range of the first two gradings each of Sets IV and V.

The average, of the group ranges on the first two gradings each of Sets I and II, is 2.08 qualities. The corresponding average of the first two gradings each of Sets IV and V is 1.73. This reduction is too small to be significant.

Table IX.

A summary of the individual averages of the first two gradings each of Sets I,II,IV and V.

Number of grading	Set I		Set II		Set IV		Set V	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Judge # 1	9.51	9.34	<u>9.06</u>	8.97	9.57	9.65	<u>8.14</u>	8.34
2	9.59	9.36	9.50	<u>9.17</u>	<u>9.45</u>	9.62	7.82	7.68
3	10.03	9.85	----	----	9.79	9.81	9.03	8.81
4	10.34	<u>9.62</u>	8.26	8.24	10.07	9.91	7.98	<u>8.04</u>
5	8.97	9.00	<u>9.14</u>	<u>9.17</u>	9.26	8.93	7.50	7.69
6	9.65	<u>9.74</u>	9.31	<u>9.27</u>	9.69	<u>9.45</u>	8.28	<u>8.14</u>
7	10.19	10.49	10.09	10.35	9.00	8.86	7.61	7.63
8	<u>9.85</u>	9.48	<u>9.10</u>	9.93	<u>9.36</u>	9.70	<u>8.16</u>	7.77
9	9.39	9.80	8.01	8.26	9.04	9.24	<u>8.10</u>	7.90
10	9.02	8.72	8.38	8.24	9.95	9.92	9.04	9.09
11	10.76	11.08	10.14	10.13	8.07	7.98	7.68	8.39
Group average	9.76	9.67	9.10	9.17	9.39	9.37	8.12	8.14
Group range	1.79	2.30	2.13	2.11	2.00	1.93	1.54	1.46
A.D.	.436	.292	.537	.602	.400	.449	.366	.381
Percent of var.	4.46	3.02	5.90	6.56	4.26	4.79	4.50	4.68

Conclusion. Individual practice, without instruction in the use of the scale, is not effective in reducing group variation.

One of the conclusions of Study A was that practice plus instruction in the use of the scale was somewhat effective in reducing group variation.

A comparison of the conclusions of the two studies on group variation reveals the fact that the method of Study A is more effective in reducing group variation.

Section IV.

The gradings on Sets VII and VIII, Study B, will be studied to get the effect of two weeks of addition practice. Any increase in the number of consistent judges on Sets VII and VIII, after the grading on Sets V and VI, may be attributed to additional practice and familiarity in the use of the scale.

Group variation, Table X, is read in the same manner as Table IX but cases of trustworthy individual variation are indicated by red rings around the variation. Student number 1 was consistent on the first two gradings each of Sets I and II. See Table VIII. It will also be observed that he was consistent on the first two gradings each of Sets VII and VIII. See Table X. In the same way the remaining data may be read.

Table X.

A summary of the individual averages of
the first two gradings each of Sets VII and VIII.

Number of grading	Set VII			Set VIII		
	1st	2nd	Ind. var.	1st	2nd	Ind. var.
Judge #1	8.66	8.50	(.16)	9.70	9.90	(.20)
2	8.75	8.88	(.13)	8.86	8.80	(.06)
3	9.47	9.32	(.15)	9.55	9.48	(.07)
4	8.15	8.14	(.01)	9.19	9.08	(.11)
5	9.31	9.18	(.13)	9.10	9.07	(.03)
6	8.92	9.16	.24	9.75	9.48	.27
7	8.95	8.90	(.05)	9.72	9.74	(.02)
8	9.95	9.44	.51	9.95	<u>9.44</u>	.51
9	9.50	9.17	.33	<u>9.46</u>	8.75	.71
10	10.13	10.13	(.00)	10.08	10.07	(.01)
11	8.54	8.70	(.16)	9.35	<u>9.28</u>	(.07)
Group average	9.12	9.05		9.50	9.37	
Group range	1.98	1.99		1.22	1.32	
A.D.	.410	.385		.300	.342	
Percent of var.	4.90	4.25		3.15	3.65	

Individual variation. On the first two gradings each of Sets I and II three judges 1,5 and 6 were consistent. See Table VIII. The ratio of the consistent judges to the number engaged was 3/11. On the corresponding gradings of Sets VII and VIII, Table X, after six weeks of practice in the use of the scale, eight judges, 1,2,3, 4,5,7,10 and 11, were consistent. The ratio of the consistent judges to the number engaged was 8/11. The ratio of the consistent judges on the first two gradings each of Sets I and II to the corresponding ratio on Sets VII and VIII was 3/8 or a decided improvement. Referring to Table VIII it will be seen that there were six consistent judges on the first two gradings of both Sets IV and V which represent gradings made by judges after four weeks of practice. The two additional weeks of practice resulted in an increase of two consistent judges.

Conclusion. The two weeks of additional practice was beneficial in reducing individual variation.

Just what the results might have been had the practice been continued cannot be told from this study but it seems reasonable that the individual variation would have been further reduced.

Group Variation. The average group range on the first two gradings each of Sets I and II was 2.08 qualities. See Table IX. The average group range on the corresponding gradings of Sets VII and VIII was 1.63, the former being the average group range made by the

inexperienced graders while the latter represents gradings made by experienced graders after six weeks of practice.

The average group range on the first two gradings each of Sets IV and V, after four weeks of practice, was 1.73 qualities while the corresponding average group range on Sets VII and VIII, after two weeks of additional weeks of practice, was 1.63 or a reduction of .10 of a quality. This reduction is too small to be significant.

Conclusion. The two weeks of additional practice produced no significant reduction in group variation.

Chapter III.

1

Study C

Section I.

The methods of Studies A and B did not reduce individual and group variation until the results of the judges after practice could be accepted as trustworthy so it became the object of this study to attempt to make the Thorndike Handwriting Scale more practical by using a method recently suggested by Thorndike.

The problem. To determine if the variation that occurs in estimating the quality of specimens of handwriting might be reduced within reliable limits by individual practice and comparisons.

Section II.

For the sake of convenience Study C was evaluated under two heads as (a) individual and (b) group variation.

Material graded. In Table XI may be found a summary of the material graded in Study C.

1. During the summer of 1915 twenty one students of the University of Missouri graded eight sets of samples of handwriting by the Thorndike Handwriting Scale.

2. Teachers' Estimates of the Quality of Specimens of Handwriting. Teachers College Record, November 1914.

3. The comparisons included a comparison of the value assigned to each sample of each grading on Supplement A with the true value of each. The true value of each sample had been determined previously by Thorndike.

Table XI.

A summary of the number of sets, number of samples in each, the number of gradings, time of each and the total number of gradings.

Number of the set	I	II	III	IV	V	VI	VII	Supplement A.
No. of samples in each	73	100	433	100	136	96	116	50
Number of gradings	3	2	1	2	2	2	2	8
Time of grading	6/21 6/22 7/9	7/2 7/3	7/5 7/6 7/7 7/8	7/13 7/14	7/15 7/16	7/19 7/20	7/21 7/22	6/17 6/18 6/23 6/24 6/28 6/29 6/30 7/12 7/19
Total gradings	219	200	433	200	272	192	232	400=2148.

Table XI may be read as follows: Set I consisted of 73 samples which were graded twice on consecutive days beginning June 21. This set was graded again on July 9. Set II consisted of 100 samples and was graded twice on consecutive days beginning July 2, etc.

The total number of gradings made during the study by each grader was 2148.

The material graded consisted of the first seven sets of samples used in Study B and in addition a special¹ set of 50 samples by Thorndike called Supplement A.

Supplement A consisted of 50 samples of handwriting of children in grades five to eight inclusive. The subject matter of each was the same and consisted of the sentence, "Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small card. John vanished behind the bushes and the carriage moved along down the driveway."

Personnel. The twenty one students concerned were members of a class in Problems of Instruction and were experienced teachers with the exception of student number 1. Seven were graduate students. There were six city superintendents, one county superintendent, one college professor, three principals, nine high school teachers, and one graduate student who had had no experience in teaching.

Nature and frequency of grading. The twenty one students were told to grade a set of samples each day with one exception by the ascending-descending procedure using the Thorndike Handwriting Scale. The graders were told that they might discuss the gradings on Supplement A but no others.

1. Teachers' Estimates of the Quality of Specimens of Handwriting. Teachers College Record, November, 1944.

The judges were told to work rapidly and not spend over 40 minutes on the first grading of a set, 30 minute periods being better.

If a grader remembered one day the value that he had assigned to a sample on a preceding day he was to indicate the fact by placing an asterisk to the right of the value assigned and omit the value from all his averages however, with the exception of Supplement A, this happened so infrequently that it did not appreciably affect the results.

The judges were told to use any numbers on the scale between four and eighteen except that they were to estimate the value to the nearest tenth of a quality. For instance, judge number 1 might estimate sample number five, the first sample of Supplement A, and give it any value 4.4.1, 4.2 etc up to eighteen which he thought it deserved.

However, at the close of the study it was found that judges 7, 13 and 15 estimated the merit of the samples only in whole qualities while judges 3, 9 - 12, 14, 18 and 19 used both methods. All the members of the group estimated Supplement A to the nearest tenth of a quality but from a cursory study of the results it seemed to make little difference which method was used.

A regular schedule of grading was followed and it was so arranged that four judges might grade at a time in the room.

Supplement A was used as a check upon the judges to bring them closer to the true values of the samples.

For instance, the true value of each of the samples of Supplement A had been determined by Thorndike and the true value was written on the back of the sample that the judge might refer to it when he had graded and recorded the judgment which he had given the sample. By this method each judge could determine immediately whether he had judged the sample too high or too low as the case might be. Thus the judge could examine the sample again, perhaps more critically than before, to find the reason of his variation from the true value. It was also hoped that the judgments of the class might be brought closer together by a comparison with a common true value.

Just how effective this method was in reducing variation will be seen in the conclusions at the close of the study.

Qualifications and criticisms. The judges were instructed to compare the values which they had assigned to the fifty samples with the true values which they did in determining the variations. Some judges disagreed with the estimates given by Thorndike as the true values. In some cases the true values seemed too high while in other cases they seemed rather low. A conference of the entire class, in which there would have been a discussion of the merits and demerits of certain samples, would

probably have assisted in settling some disputed matters.

It would seem that a careful analysis of the merits and demerits of each sample should accompany the list of fifty samples. It could be used as a reference and would assist in remedying the difficulty mentioned above.

This would be analogous to the work of Ballou in his "Scale for the Measurement of English Composition."

His analysis enumerates the merits and demerits of each sample. This analysis is obtained by taking the average judgment of competent judges and is the same foundation upon which Thorndike built his scale for judging handwriting.

Table XII gives a summary of the individual averages on the gradings of Sets I-VII in the chronological order in which they were graded. Table XII may be read as follows: In the first horizontal column at the top are given the number of each set and just below in the next horizontal column are given the dates of the respective gradings. In the first vertical column from the left may be found a number for each grader. In the second vertical column from the left and at the top may be found the date June 21 and immediately below are the figures 11.17 which means that on June 21 judge number 1 graded Set I and found his average to be 11.17 qualities. Immediately below in the same vertical may be found the number 9.26 which means that on the same day judge number 2 graded Set I

Table XII.

A summary of the number of sets, date of grading and the arithmetical averages, of each grader, given in the order in which they were graded. Supplement A was omitted because its averages were not significant.

Set	Set I		Set II		Set III		Set IV		Set V		Set VI		Set VII	
Date of grading	6/21	6/22	7/2	7/3	7/5	7/9	7/18	7/14	7/15	7/16	7/19	7/20	7/21	7/22
Judge number 1	11.17	10.73	9.82	8.96	12.16	10.90	10.91	10.93	10.10	9.12	9.62	9.55	10.15	10.25
2	9.26	9.34	9.84	10.07	11.68	10.53	9.77	9.91	7.88	8.37	9.52	9.71	10.14	9.57
3	10.09	10.02	9.94	9.90	11.59	9.90	9.55	9.49	8.88	8.83	8.92	9.16	10.32	10.41
4	10.37	10.50	9.27	9.25	11.26	10.57	9.05	9.22	8.05	8.00	8.32	8.47	9.76	9.61
5	10.60	10.80	9.59	9.54	11.32	10.73	10.94	10.57	9.46	9.39	10.25	10.03	10.25	10.03
6	9.66	9.80	9.39	9.83	11.54	9.56	9.45	10.21	8.63	8.46	9.05	8.97	9.95	9.94
7	9.90	9.43	11.02	10.64	11.71	8.62	10.17	10.82	7.77	7.18	9.44	9.58	10.73	10.54
8	9.56	9.51	9.27	9.44	10.70	9.37	9.31	9.64	9.79	9.63	9.49	9.69	10.35	10.28
9	10.08	9.95	8.59	8.67	11.48	9.09	9.33	9.61	7.75	7.34	8.73	7.88	8.18	8.63
10	8.60	8.77	9.20	9.14	11.25	9.77	10.71	10.70	9.11	9.83	9.67	10.68	10.67	10.62
11	10.41	10.36	9.13	9.04	11.41	10.20	10.50	10.80	8.65	9.46	9.70	9.61	10.48	10.51
12	9.65	9.37	8.91	9.13	11.26	10.13	10.50	10.58	9.47	9.39	9.67	9.56	9.53	9.56
13	8.90	8.75	10.24	10.32	11.41	9.86	9.63	10.09	7.95	8.08	9.85	9.83	10.61	10.22
14	10.10	9.93	9.78	9.42	11.66	9.63	9.64	9.66	8.35	8.25	9.77	9.67	10.19	9.99
15	10.34	10.15	10.08	9.77	11.66	10.44	10.64	10.69	9.67	9.57	9.83	9.85	10.34	10.16
16	11.20	10.70	9.43	9.64	11.55	10.15	10.76	10.33	8.82	8.54	9.32	9.01	9.41	9.47
17	10.04	10.64	10.32	10.32	12.27	10.04	10.79	10.89	8.54	8.36	9.35	9.10	9.19	9.23
18	10.37	9.44	9.51	9.22	11.54	9.26	9.66	9.82	8.64	8.73	9.47	9.50	10.23	10.11
19	10.70	10.50	9.96	9.86	11.26	9.34	10.53	10.73	9.09	9.06	9.80	9.27	9.89	9.94
20	10.11	10.12	9.20	9.17	11.53	10.08	9.45	9.48	7.60	7.67	8.23	8.21	9.03	9.06
21	9.87	9.90	9.91	9.79	11.53	9.93	9.91	10.00	8.48	8.78	9.36	8.86	9.78	9.55
Group average	10.07	9.98	9.64	9.58	11.53	9.96	10.05	10.20	8.70	8.70	9.40	9.34	9.96	9.89

under similar conditions and found the average of his judgments to be 9.26 qualities. In the same way the next 19 numbers may be read. Just below the individual averages may be found the group average for that day which was 10.07 qualities.

Plan of selecting data from which conclusions will be drawn. In individual variation conclusions were drawn from the variation which occurs between the individual averages of the first and second gradings of the following sets of samples: (a) Sets I and II which represent gradings made judges of very little experience in the use of the Thorndike Handwriting Scale and (b) Sets VI and VII which represent gradings made by judges of considerable experience in the use of the scale.

The number of consistent judges on both the first two sets will be compared with the number of consistent judges on the last two sets.

In group variation conclusions were drawn from the average group range of the daily individual averages of the first two gradings of each of the following sets: (a) Sets I and II which represent gradings made by judges of very little experience in the use of the scale and (b) Sets VI and VII which represent gradings made by judges after considerable experience in the use of the scale. A comparison was of the average group range on

the first two gradings each of the first two sets with the average group range on the first two gradings on the last two sets.

Section III.

Individual Variation

The Problem. To determine if the individual variation that occurs in estimating the quality of specimens of handwriting might be reduced within trustworthy limits by individual practice and comparisons made occasionally with the true values of certain samples.

On the first two gradings of Set I fifteen judges, numbers 2 -6,8-10,13-15,19-21, had an individual variation of less than .20 of a quality. These cases of low and trustworthy individual variation were indicated by red rings drawn around the variation. On the two gradings of Set II twelve judges, 3-5,8-11,13,17,19-21, had a trustworthy individual variation. Eleven judges,3-5,8-11,13,19-21, were consistent on both sets. The ratio of the consistent judges on both sets to the number grading in the group was 11/21.

On the first two gradings of Set VI fourteen judges, numbers 1,2,4,6-9,11-15,18 and 20, had an individual variation of less than .20 of a quality. On the two gradings of Set VII seventeen judges, numbers 1,3,4,6-8,10-20, had an individual variation that was reliable.

Table XIII.

A summary of the individual averages on the first
two gradings of Sets I, II, VI and VII.

Number of grading	Set I			Set II			Set VI			Set VII		
	1st	2nd	Ind. var.	1st	2nd	Ind. var.	1st	2nd	Ind. var.	1st	2nd	Ind. var.
Judge #1	11.17	10.73	.44	9.82	8.96	.86	9.62	9.55	.07	10.15	10.25	.10
2	9.26	9.34	.08	9.84	10.07	.23	9.52	9.71	.19	10.14	9.57	.57
3	10.09	10.02	.07	9.94	9.90	.04	8.92	9.16	.24	10.32	10.41	.09
4	10.37	10.50	.13	9.27	9.25	.02	8.32	8.47	.15	9.76	9.61	.15
5	10.60	10.80	.20	9.59	9.54	.05	10.25	10.03	.22	10.25	10.03	.22
6	9.66	9.80	.14	9.39	9.83	.44	9.05	8.97	.08	9.95	9.94	.01
7	9.90	9.48	.42	11.02	10.64	.38	9.44	9.58	.14	10.73	10.54	.19
8	9.56	9.51	.05	9.27	9.44	.17	9.49	9.69	.20	10.35	10.28	.07
9	10.08	9.95	.13	8.59	8.67	.08	8.78	7.88	.10	8.18	8.63	.45
10	8.60	8.77	.17	9.20	9.14	.06	9.67	10.68	1.01	10.67	10.62	.05
11	10.41	10.36	.05	9.13	9.04	.09	9.70	9.61	.09	10.48	10.51	.03
12	9.65	9.37	.28	8.91	9.13	.22	9.67	9.56	.11	9.53	9.56	.03
13	8.90	8.75	.15	10.24	10.32	.08	9.85	9.83	.02	10.61	10.72	.11
14	10.10	9.93	.17	9.78	9.42	.36	9.77	9.67	.10	10.19	9.99	.20
15	10.34	10.15	.19	10.08	9.77	.31	9.83	9.85	.02	10.34	10.16	.18
16	11.20	10.70	.50	9.43	9.64	.21	9.32	9.01	.31	9.41	9.47	.06
17	10.04	10.64	.60	10.32	10.32	.00	9.35	9.10	.25	9.19	9.28	.09
18	10.37	9.44	.93	9.51	9.22	.29	9.47	9.50	.03	10.23	10.11	.12
19	10.70	10.50	.20	9.96	9.86	.10	9.80	9.27	.53	9.89	9.94	.05
20	10.11	10.12	.01	9.20	9.17	.03	8.23	8.21	.02	9.08	9.06	.02
21	9.87	9.90	.03	9.91	9.79	.12	9.36	8.86	.50	9.78	9.55	.23
Group average	10.07	9.98		9.90	9.93		9.40	9.34		9.96	9.89	
Group range	2.60	2.03		2.42	1.97		2.02	2.47		2.55	1.99	
A.D.	.487	.498		.431	.315		.369	.437		.467	.424	
Percent of var.	4.85	5.00		4.47	3.29		3.93	4.67		4.69	4.29	

Thus after three weeks of additional practice it was found that 17 out of 21 judges were able to reduce their individual variation within usable limits.

Conclusions. Individual practice and comparisons with the true value give a high probability of a single individual being able to reduce his individual variation within usable limits.

Five judges, numbers 4, 8, 11, 13 and 20, were consistent on the five sets. Judges 4 and 20 were consistent on all sets graded.

The study shows that certain graders may secure and maintain a trustworthy variation.

Six judges, numbers 2, 3, 9, 14, 20 and 21, had used the Thorndike Handwriting Scale previous to this study but the fact gave them no appreciable advantage in this study. Of the twelve consistent judges on the first two sets but three, 9, 20 and 21, were practiced before this study. Of the seventeen consistent judges on the last set but two, 14 and 20, had had previous practice on the scale.

A study was made of the individual variation between (a) two successive gradings on a set of samples close to each other in point of time and (b) two gradings made on the same set of samples but quite removed from each other in point of time during which time other grading of handwriting had been done.

The gradings of Set I in Table XIV seem typical. The first two gradings were made on two successive days while over two weeks elapsed between the second grading of Set I and the third grading on the same set.

There were fifteen consistent judges on the first two gradings but on the second and third there were but seven.

Conclusion. The variation between two gradings on the same set of samples made close to each other in point of time is less than the variation between two gradings on the same set much removed from each other in point of time.

Section IV.

Group variation. The average group range on the first two gradings each of Sets I and II, see Table XIII, was 2.26 qualities. The average group range on the first two gradings each of Sets VI and VII was exactly the same or 2.26 qualities. This is more than the difference between the average handwriting of four successive school grades and is too large to be reliable.

Section V.

Miscellaneous

A study was made of the individual averages that varied not more than .10 of a quality from the group average on that day. The averages of Table XIII seem

Table XIV.

A summary of the individual averages of each of
the 21 judges on the three gradings of Set I.

Number of
grading

Judge #	1	11.17	10.73	.44	10.73	10.90	.117
	2	9.26	9.34	(.08)	9.34	10.53	1.19
	3	<u>10.09</u>	<u>10.02</u>	(.07)	<u>10.02</u>	<u>9.90</u>	.88
	4	<u>10.37</u>	<u>10.50</u>	(.13)	<u>10.50</u>	<u>10.57</u>	(.07)
	5	<u>10.60</u>	<u>10.80</u>	(.20)	<u>10.80</u>	<u>10.73</u>	(.07)
	6	9.66	9.80	(.14)	9.80	9.56	.24
	7	9.90	9.48	.42	9.48	8.62	.86
	8	9.56	9.51	(.05)	9.51	9.37	(.14)
	9	<u>10.08</u>	<u>9.95</u>	(.13)	<u>9.95</u>	9.09	.86
	10	8.60	8.77	(.17)	8.77	9.77	1.00
	11	<u>10.04</u>	<u>10.36</u>	(.05)	<u>10.36</u>	<u>10.20</u>	(.16)
	12	9.65	9.37	.28	9.37	10.13	.76
	13	8.90	8.75	(.15)	8.75	<u>9.86</u>	1.11
	14	<u>10.10</u>	<u>9.93</u>	(.17)	<u>9.93</u>	<u>9.63</u>	.30
	15	<u>10.34</u>	<u>10.15</u>	(.19)	<u>10.15</u>	<u>10.44</u>	.29
	16	11.20	10.70	.50	10.70	<u>10.15</u>	.55
	17	<u>10.04</u>	<u>10.64</u>	.60	<u>10.64</u>	<u>10.04</u>	.60
	18	<u>10.37</u>	9.44	.93	9.44	9.26	(.18)
	19	<u>10.70</u>	<u>10.50</u>	(.20)	<u>10.50</u>	9.34	1.16
	20	<u>10.11</u>	<u>10.12</u>	(.01)	<u>10.12</u>	<u>10.08</u>	(.04)
	21	<u>9.87</u>	<u>9.90</u>	(.03)	<u>9.90</u>	<u>9.93</u>	(.03)
Group ave.		10.07	9.98		9.98	9.96	
Group range		2.60	2.03		2.03	2.28	
A.D.		.487	.498		.498	.449	
% of varia.		4.85	5.00		5.00	4.51	

typical. On the first two gradings of Set I judges 3, 9 and 14 were within .10 of a quality of the group average on both gradings. Judges 17 and 20 on the first grading and 21 on the second were also within .10 of a quality. On Set II judges 3 and 19 were within the limit on both gradings. In addition judges 1, 2 and 21 were within the limit of the first grading and number 6 on the second. On the second grading of Set VI but one judge, number 19, came within the limit. There were fewer judges within the limit on the last two gradings of the study, Set VII, than on the first two gradings of Set I.

A study was made of the reliability of the averages of certain numbers of samples.

The data of Table XV was obtained from the first and second gradings of Set VII which consisted of 116 samples. First the individual averages of the two gradings of the full set were compared. Second the individual averages of the first forty samples were compared; the individual averages of the second forty samples were compared, and the individual averages on the last thirty six samples were compared.

Individual variation. There were sixteen consistent judges on the two gradings of the full set. There were fourteen consistent judges on the two gradings of the

Table XV.

A summary of the individual averages of the 21 judges of Study C (a) on the full set; (b) on the first 40 samples; (c) on the second 40 samples and on the last 36 samples. The data was taken from the first and second grading of Set VII.

No. grading	1st	2nd		1st	2nd		1st	2nd		1st	2nd	
	Full set			First 40			Second 40			Last 36		
Judge #1	10.15	10.25	(.10)	10.15	10.11	(.04)	9.75	9.84	(.09)	10.35	10.45	(.10)
2	10.14	9.57	.57	9.93	10.25	.32	9.85	10.19	.34	10.72	11.25	.53
3	10.32	10.41	(.09)	10.12	10.27	(.13)	10.05	10.25	(.20)	10.88	10.72	(.16)
4	9.76	9.61	(.15)	9.03	8.98	(.10)	9.72	9.58	(.15)	10.56	10.36	(.20)
5	10.25	10.03	.22	9.67	9.88	(.19)	10.22	9.44	.78	10.92	10.86	(.06)
6	9.95	9.94	(.01)	9.70	9.62	(.08)	9.73	9.70	(.03)	10.48	10.56	(.08)
7	10.73	10.54	(.19)	10.53	9.33	1.20	10.37	10.23	.86	11.36	11.14	.22
8	10.35	10.28	(.07)	10.06	9.95	(.11)	10.05	11.41	1.41	11.01	9.61	.40
9	8.18	8.63	.45	8.07	8.32	.25	8.05	8.48	.43	8.44	9.17	.73
10	10.67	10.62	(.05)	10.35	10.52	(.17)	10.40	10.27	(.13)	11.30	11.11	(.19)
11	10.48	10.51	(.03)	10.47	10.22	.25	9.80	10.60	.80	11.25	10.75	.50
12	9.53	9.56	(.03)	9.45	9.47	(.02)	9.26	9.32	(.06)	9.91	9.91	(.00)
13	10.61	10.22	.39	10.32	10.13	(.19)	10.35	10.05	.30	11.22	10.53	.31
14	10.19	9.99	(.20)	10.25	9.62	.63	9.95	9.72	.23	10.41	10.69	.28
15	10.34	10.16	(.18)	9.98	9.85	(.13)	10.07	9.95	(.12)	11.06	10.75	.69
16	9.41	9.47	(.06)	9.35	9.20	(.15)	9.21	9.35	(.14)	9.70	9.91	.21
17	9.19	9.28	(.09)	9.18	9.11	(.07)	9.21	8.98	.23	9.23	9.81	.58
18	10.23	10.11	(.12)	10.07	9.67	.40	10.02	10.02	(.00)	10.63	10.69	(.06)
19	9.89	9.94	(.05)	9.58	9.69	(.11)	9.75	9.95	(.20)	10.39	10.22	.17
20	9.08	9.06	(.02)	8.50	8.50	(.00)	8.58	8.40	(.18)	10.30	10.30	(.00)
21	9.78	9.55	.23	9.32	9.16	(.16)	9.50	9.35	(.15)	10.59	10.20	.39
Group average	9.96	9.89	16	9.71	9.80	14	9.71	9.74	11	10.51	10.42	10
Group range	2.55	1.99		2.46	2.20		2.35	3.06		2.92	2.08	

first forty samples; eleven on the second forty samples and ten consistent judges on the averages of the last thirty six samples.

Conclusion. If this study is typical the conclusion may be drawn that as the number of samples is increased the chances of large individual variation are decreased.

Group variation. There was nothing significant about the group range except that it was too large to be reliable.

Chapter IV.

Some problems.

Introduction. In solving the main problems of Studies A, B, and C several minor but related problems were found and solved. These problems were considered important enough to be preserved. They were collected and placed in this chapter for convenience but the order of the problems has no significance.

Problem I.

The problem. Do inexperienced graders have a tendency to slight quality number ten in grading?

Data. Set I was graded eleven times by the eleven judges of Study B. The occurrences of quality 10 in the gradings of these judges were assembled and furnish the data for Table XVI.

Method of selecting data from which conclusions will be drawn. The first two and the last two gradings of Set I were canvassed to find the number of occurrences of quality ten found by each grader. The total number found by the group on each grading was found and an average determined for the first two and the last two gradings each. A comparison of the averages is an answer to the problem since successive gradings on Set I should represent greater proficiency in the use of the scale.

The last two gradings of Set I, the tenth and eleventh, were made after seven weeks of practice in the use of the scale. Each member of the class had made 4226 judgments on samples of handwriting.

Table XVI may be read as follows: The first horizontal column, at the top, indicates the number of the grading. The first vertical column, numbering from the left, gives the number of each grader arranged in alphabetical order. The second vertical column from the left gives the number of tens found on the first grading of Set I. For instance judge number 1 graded the set the first time and found five values of quality ten. Judge 2 found three while judge 3 found but one value of quality ten on the first grading. The averages at the bottom indicate the average number of occurrences of quality ten found by each grader on each grading. For instance, on the first grading the average number occurrences of quality of ten was less than four, on the second grading it was just a little larger than five. On the tenth and eleventh gradings, after seven weeks of practice, the averages were more than six, a material increase if computed in terms of percent.

Conclusion. Inexperienced graders have a tendency to slight quality number ten and this tends to disappear, to a certain degree, as practice continues.

Table XVI.

A summary of the number of tens given by each
of the eleven graders on the first two and the

tenth and eleventh gradings of Set I.

Number of grading 1st 2nd 10th 11th

Judge number	1	5	8	5	5
	2	3.	4	3	4
	3	1	1	4	6
	4	6	4	4	6
	5	5	8	3	2
	6	0	2	7	8
	7	4	9	4	7
	8	7	9	9	5
	9	6	5	6	5
	10	0	2	11	11
	11	4	7	11	9
Total		41	57	67	68
Average		4-	5 $\frac{1}{2}$	6 $\frac{1}{2}$	6 $\frac{1}{2}$

Problem II.

The problem. In what part of the scale does the greatest variation in grading samples of handwriting occur?

Method of grading. The method of grading was individual practice, the samples being graded but once.

Method of finding the median. In this problem quality 10.50 will be considered the median of the scale. The median was determined as follows: Quality eighteen is a perfect copper plate and represents perfection a stage no penman can reach. For this reason quality eighteen will be eliminated when finding the median for the Thorndike Handwriting Scale. This leaves qualities four to seventeen inclusive. The median of these values is 10.50. Of the fourteen steps or qualities that remain seven are below and seven are above the value 10.50. Quality eleven belongs to the upper half and quality ten is included in the lower.

Material graded. The material graded consisted of a modified form of Set I. Nineteen samples were added to the original number seventy three of the set and these were graded once by ten judges.

Personnel. All the judges were experienced teachers with the exception of 1 and 7. All were students of the School of Education of the University of Missouri. Judges 3 and 4 were graduate students, the remaining judges were seniors except 5 and 9 who were juniors.

Qualifications and criticisms. All the judges did not represent equal degrees of proficiency in the use of the scale. For instance judges 3,4, and 5 were graders of much experience in the use of the scale while the remaining members of the group had had some experience except judges 1,7 and 8. The results would have been more trustworthy had all the judges been experienced.

Table XVII gives the individual judgments in detail and may be read as follows: The numbers in the horizontal column at the top indicate the number of the graders who are arranged in alphabetical order. The vertical column to the left gives the numbers corresponding to the samples and in each horizontal column except the first at the top will be found the ten judgments upon each sample. For instance sample number one was graded by judge number 1 who found the quality to be 13, in the same way judge number 2 found the value of the sample to be 13 etc.

To the right will be found a vertical column labeled "extreme range" which gives the extreme range of ten judgments on each sample. The next vertical column to the right gives the average of the ten judgments on each sample. The last vertical column to the right gives the average deviation on each of the 92 samples.

The median of the 92 average deviations was determined and it was found to be .90 of a quality.

Table XVII.

A summary of the judgments of a single grader by ten graders on 92 samples of handwriting.

Number of grader	1	2	3	4	5	6	7	8	9	10	Grader Average	Ave. Ten	A.D.
Number of sample													
1	13	13	11	13	11	12	14	14	10	14	4	12.5	1.20
2	16	16	14	15	12	15	16	16	14	16	4	15.0	1.10
3	6	4	6	8	6	7	6	6	4	5	4	5.8	.88
4	8	8	8	9	8	8	8	8	8	9	1	8.2	.32
5	10	9	9	10	9	8	9	9	9	10	2	9.2	.48
6	11	8	8	10	9	9	9	9	9	9	3	9.1	.48
7	9	8	7	9	8	8	8	9	8	8	2	8.2	.48
8	17	15	15	16	15	12	17	17	11	17	6	15.2	1.60
9	16	16	16	15	15	16	16	16	11	18	7	15.5	1.10
10	13	13	11	12	12	16	14	14	9	11	7	12.5	1.50
11	7	7	5	6	7	8	7	8	6	8	3	6.9	.74
12	9	11	9	10	10	10	9	14	9	11	5	10.2	1.08
13	12	12	9	10	11	11	12	9	9	11	3	10.6	1.14
14	5	5	8	8	9	7	6	7	5	6	4	6.6	1.20
15	14	11	12	13	12	12	16	14	11	14	5	12.9	1.30
16	14	13	12	13	11	10	17	16	11	13	7	13.0	1.60
17	6	5	5	5	6	6	5	5	4	5	2	5.2	.42
18	11	10	10	10	10	12	12	12	10	11	2	10.8	.80
19	9	9	9	9	9	9	9	9	9	10	1	9.1	.18
20	9	11	7	7	8	8	8	8	8	8	4	8.2	.72
21	10	10	11	10	11	10	10	13	10	12	3	10.7	.81
22	8	8	7	8	8	8	7	8	8	9	2	10.7	.84
23	9	9	8	9	9	8	9	9	9	9	1	8.8	.32
24	16	11	12	16	14	16	16	16	14	16	5	14.7	1.56
25	12	9	10	10	12	9	12	12	10	12	3	10.8	1.20
26	9	8	8	9	10	8	9	9	9	9	2	8.8	.48
27	13	13	12	14	13	12	16	16	13	15	4	13.7	1.24
28	11	9	11	11	11	9	9	13	10	12	4	10.6	1.08
29	16	16	11	16	12	15	15	16	16	15	5	14.8	1.32
30	8	5	9	9	9	7	9	8	8	9	4	8.1	.90
31	11	9	8	9	9	8	8	9	9	9	3	8.9	.44
32	12	12	15	12	12	15	15	13	11	12	4	12.9	1.28
33	12	12	13	14	12	10	15	15	11	15	5	13.0	1.50

Number of grader	1	2	3	4	5	6	7	8	9	10		
Number of sample												
34	12	10	13	12	13	13	13	13	10	12	5	12.6 .88
35	11	8	9	9	9	8	11	8	8	9	3	9.0 .80
36	9	9	10	10	10	9	11	11	8	9	3	9.6 .80
37	10	10	10	10	10	9	10	12	8	11	4	10.0 .60
38	8	6	7	8	7	6	7	5	6	8	3	6.8 .84
39	16	11	12	12	12	13	12	13	11	12	5	12.4 .96
40	13	13	13	15	11	11	18	16	14	15	7	13.9 1.70
41	9	7	8	8	8	7	8	9	8	9	2	8.1 .54
42	14	11	12	13	11	10	14	13	10	13	4	12.1 1.30
43	8	6	5	6	6	6	7	5	6	6	3	6.1 .56
44	11	9	11	11	11	12	13	9	10	13	4	11.0 1.00
45	15	11	13	14	12	10	17	15	10	14	7	13.1 1.90
46	14	13	13	12	12	14	16	14	11	13	3	12.0 1.40
47	14	10	12	12	10	11	15	13	10	13	5	12.0 1.40
48	8	8	8	9	9	8	9	9	8	10	2	8.6 .60
49	8	8	6	6	7	7	7	5	6	7	3	6.7 .76
50	13	11	12	12	9	12	13	11	11	12	4	11.6 .72
51	13	13	14	15	13	15	16	16	12	13	4	14.0 1.20
52	14	12	12	12	12	13	14	13	12	13	2	12.7 .70
53	14	13	13	14	15	16	16	15	13	14	3	14.3 .96
54	11	9	10	10	9	10	11	9	9	11	2	9.9 .72
55	6	4	5	5	6	6	5	6	4	5	2	5.2 .64
56	15	16	13	11	12	13	13	13	12	13	5	13.1 .96
57	15	12	13	13	14	12	13	13	10	12	5	12.7 .90
58	8	8	7	8	8	6	8	9	8	8	2	7.8 1.16
59	6	8	9	11	11	9	8	7	7	9	5	8.5 1.30
60	9	9	12	10	10	9	12	9	10	12	3	10.2 1.08
61	7	5	6	8	7	6	7	6	6	7	2	6.5 .70
62	16	14	13	13	12	14	14	13	11	13	5	13.3 .96
63	8	8	10	10	10	7	13	9	8	11	6	9.4 1.40
64	11	11	12	11	11	11	14	16	11	13	5	12.1 1.34
65	9	7	8	8	8	8	8	8	7	8	2	7.9 .36

Number of grader	1	2	3	4	5	6	7	8	9	10			
Number of sample													
66	9	10	10	9	10	10	9	8	9	9	2	9.3	.56
67	14	16	14	16	13	16	16	14	12	16	4	14.7	1.30
68	16	11	15	15	12	15	15	16	12	15	5	14.2	1.52
69	13	13	13	14	13	11	15	13	12	15	4	13.2	.88
70	6	5	5	5	6	5	5	7	6	5	2	5.5	.60
71	11	11	10	9	9	8	9	9	9	9	3	9.4	.76
72	9	8	9	10	11	9	13	13	10	11	5	10.3	1.46
73	11	11	11	11	10	11	13	11	11	11	3	11.1	.38
74	14	9	12	15	11	10	15	14	13	14	6	12.7	1.66
75	8	5	5	6	7	6	8	5	6	6	3	6.2	.88
76	16	16	14	15	13	13	16	16	16	16	3	15.1	1.08
77	9	8	10	10	10	10	10	9	9	12	4	9.7	.91
78	9	8	9	8	8	8	9	8	8	9	1	8.4	.48
79	5	4	4	4	5	4	4	5	4	4	1	4.3	.42
80	13	13	12	16	14	13	17	13	13	13	5	13.7	1.18
81	13	11	13	14	12	12	14	16	11	14	5	13.0	1.20
82	8	7	6	10	8	7	9	11	8	9	5	8.3	1.16
83	6	5	6	9	7	8	7	8	7	7	3	7.0	.80
84	12	9	13	12	12	12	15	9	12	14	6	12.0	1.20
85	4	4	5	5	7	5	5	5	5	5	3	5.0	.40
86	9	9	11	10	9	8	8	8	9	10	3	9.1	.74
87	11	9	9	11	9	9	9	9	8	9	3	9.3	.68
88	7	8	7	7	7	6	7	6	7	8	2	7.0	.40
89	13	9	12	12	10	11	14	11	11	13	5	11.6	1.20
90	8	8	5	6	7	5	7	7	7	7	3	6.2	.90
91	16	13	13	14	13	13	15	14	16	16	3	14.3	1.16
92	14	12	12	13	12	11	13	13	11	13	3	12.4	.80

Of the 46 samples having an average deviation of .90 or a quality or more 38 of these had an average of 10.50 or more qualities.

Conclusion. The greatest variation in grading samples of handwriting occurs in the use of the upper qualities of the scale.

Substantially the same conclusions were reached by C.T.Gray, The training of Judgments in the use of the Ayres Scale for Handwriting, Journal of Educational Psychology, February 1915, page 88. Gray however failed to furnish any data as proof of his statement.

Problem III.

Section I.

The problem. How many individual averages or judgments must be combined to secure a trustworthy average?

For the sake of convenience this subject was handled under the three heads, Sections I, II, and III.

Section I is an introduction. Section II is a study of averages in which the material was grouped according to a method or selection. Section III is a study of averages based upon random selection.

Section II.

The object of this study is to examine the averages of an increasing number of graders and attempt to find the minimum number whose averages may be trustworthy.

The practical school man, knowing the unreliability of the ordinary grader of penmanship, will inquire "What is the minimum number of graders to use in order to secure a reasonable degree of accuracy in the averages on a sample or a series of samples of handwriting?"

This can be best answered by using some actual results, results secured under conditions possible for ordinary school practice if a moderate degree of care is exercised in preparation and technique.

Method. The plan was to secure the averages of an increasing number of graders by permutations and combinations. For instance, the averages of four

graders were taken at a time from the data secured from Study A. Comparisons were made by comparing the group range on each set in either case. Much of the detailed computations was omitted except the averages of four by permutations from the data of Study A.

Source of data. The data for the averages of two and three judges, 2, 3, and 11 was taken from the gradings of Study B. The data for the averages of four and five was taken from the gradings of Study A. The data for the average of sixteen was obtained by combining the averages of the five graders of Study A with the corresponding averages of the eleven graders of Study B.

Table XVIII and explanations. Table XVIII was compiled from the gradings of Studies A and B. The data from the first three sets was not used in either case for two reasons: (a) The graders of Study B graded Set III but once while the graders of Study A graded the set twice and (b) this omission allowed the judges in either case practice on approximately 1300 samples before their gradings were used.

Table XVIII may be read as follows: The horizontal column at the top indicates the number of the set of samples. The first vertical column to the left indicates the number of the graders being combined and their exact numbers as they were ranked in the original studies.

Average of one grader. When the data of Studies A and B was evaluated it was found that no judge had a consistent and dependable average for occasionally he would vary more than .20 of a quality on a set of samples. Since the average of no single judge was found to be trustworthy the averages of an increasing number were combined by permutations and combinations hoping to find a dependable average.

Averages of two. The daily averages of judges 2 and 10, Study B, were combined and found to be consistent. Those were the only two judges in both studies whose averages were dependable when combined. Their averages are shown in the second horizontal column from the top.

Averages of three. The daily averages of judges 2,3 and 10, Study B, were combined and found to be consistent. These were the only three judges in both studies whose averages were trustworthy when combined. Their combined averages are shown in the third horizontal column from the top of Table XVIII.

Averages of four. In the next five horizontal columns from the top will be found a study by permutations of the

Table XVIII.

A summary of some averages on Sets I, IV and V.

	Set I				Set IV				range				Set V				range				Summary
Judges	2&10	9.50	9.37	.13	9.70	9.77	9.75	9.66	9.67	.11	8.43	8.39	8.30	8.36	8.32	.13	yes				
	2,3&10	9.54	9.44	.10	9.82	9.82	9.73	9.69	9.69	.13	8.63	8.53	8.44	8.50	8.51	.19	yes				
	1 to 4	9.49	9.74	.25	9.97	9.98	9.89	9.77	9.76	.22	8.18	7.97	8.16	8.07	8.05	.31	no				
	2 to 5	9.42	9.61	.19	9.77	9.75	9.73	9.72	9.70	.07	8.36	8.29	8.47	8.23	8.26	.24	no				
	3-5&1	9.54	9.51	.03	9.91	9.78	9.82	9.81	9.78	.13	8.40	8.24	8.22	8.27	8.30	.18	yes				
	4,5,1&2	9.39	9.63	.24	9.78	9.91	9.85	9.81	9.78	.13	8.32	8.10	8.10	8.21	8.21	.22	no				
	5,1-3	9.35	9.61	.26	9.79	9.78	9.75	9.67	9.65	.14	8.06	7.97	8.30	8.13	8.19	.33	no				
The five		9.43	9.62	.19	9.84	9.84	9.80	9.75	9.73	.11	8.26	8.11	8.37	8.18	8.20	.26	no				
Eleven		9.53	9.37	.16	9.39	9.37	9.32	9.42	9.28	.14	8.12	8.14	8.18	8.04	8.99	.19	yes				
Sixteen		9.50	9.45	.05	9.53	9.51	9.47	9.52	9.43	.10	8.16	8.17	8.24	8.08	8.06	.18	yes				

reliability of the averages of four graders. The data was secured from the gradings of Study A. The group consisting of judges 3,4,5 and 1 were more consistent than any of the remaining groups of four and even more consistent than the larger group of five. It will be observed that the group range on Set V for the five was .26 of a quality while the corresponding range for the group of judges,3,4,5,and 1 was .18 which is trustworthy so it will be observed in this study that the averages of four judges were more dependable than the averages of five. Therefore an increasing number of judges does not necessarily mean a more reliable average, however, in the main, the greater the number of judges the more trustworthy the averages. This will be seen by inspecting the averages of the groups of eleven and sixteen judges in the table.

If the averages in this study may be taken as typical it would seem to indicate that the averages of two judges in some cases may be dependable even if no single judge has a trustworthy average.

Conclusion. In the main the greater the number of judges the more trustworthy the averages. But it is impossible from this study to determine the minimum number of judges whose individual averages might be combined to produce a trustworthy average.

A brief study was made of the variation of the group average of judges 2 and 10, and 2,3 and 10 from the group average of the eleven judges of Study B. All cases, in which the group average of judges 2 and 10, and 2,3, and 10 came within .10 of a quality of the group average of the eleven, was indicated by drawing a red line under the group average for that grading. It is interesting to note that their averages came within .10 of a quality of the group average of the eleven but on the gradings of Set I. It might also be mentioned that these were not the first gradings on Set I but the sixth and seventh.

Table XIX.

A summary of the individual judgments of 32 graders on Set I.

Sample number	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	26	27								
Grader's number																										
1	8	14	15	9	13	11	4	16	5	6	16	4	11	10	9	17	8	12	8	13	17	9	10	12	4	10.44
2	6	14	15	7	12	11	5	15	6	9	13	4	9	8	9	16	8	15	8	14	16	9	10	12	7	10.32
3	9	15	13	8	9	11	7	14	9	9	11	5	9	9	8	13	9	11	8	11	14	9	9	9	6	
4	6	11	14	7	9	8	5	15	5	7	13	4	8	9	7	16	5	15	6	11	15	7	8	9	5	
5	5	13	14	8	9	9	4	12	8	7	13	4	8	9	9	16	8	11	6	13	14	8	9	9	6	9.12
6	7	14	13	8	9	11	6	12	8	9	16	5	9	8	9	16	8	9	7	13	16	8	9	12	6	
7	6	11	12	8	9	9	4	12	8	9	13	4	5	8	9	16	5	14	8	13	13	8	9	9	5	
8	6	13	12	8	9	11	5	12	7	7	14	5	8	8	7	16	5	11	6	14	16	7	12	9	5	9.32
9	7	15	15	8	11	12	5	16	10	11	16	7	9	9	10	17	7	15	9	16	17	9	11	12	7	
10	6	16	18	9	12	10	5	13	7	9	16	4	9	10	9	16	5	14	8	14	14	7	8	9	6	
11	8	16	15	6	9	9	5	15	7	8	16	4	9	11	8	16	8	13	8	12	15	7	9	11	6	
12	7	14	13	9	13	11	4	12	7	9	13	4	9	8	9	16	7	16	5	11	14	7	9	9	6	
13	7	13	12	7	9	10	5	12	7	8	13	4	8	9	8	13	7	11	9	13	17	9	12	10	7	
14	7	14	15	7	12	11	4	15	8	12	16	4	8	10	9	13	5	16	5	12	16	8	9	9	4	
15	6	15	12	6	9	11	4	13	8	9	15	4	8	9	9	16	6	11	5	13	14	7	9	9	5	
16	9	13	13	7	9	13	5	13	9	9	16	5	9	12	8	13	7	11	8	12	14	9	8	9	8	9.72
17	7	11	13	7	9	9	4	12	8	9	12	4	9	9	9	16	4	13	5	14	13	6	9	12	5	
18	8	14	16	7	9	10	6	14	11	10	14	5	8	12	9	16	7	14	8	11	14	9	10	13	6	
19	7	13	13	7	9	9	5	15	8	9	16	4	9	9	8	16	7	13	5	11	14	8	9	12	5	
20	6	11	12	5	9	9	4	12	5	7	13	4	7	9	9	16	8	11	7	11	13	7	9	9	5	
21	7	13	12	7	9	8	7	13	8	8	13	5	8	8	8	12	7	11	7	11	12	8	9	9	7	
22	7	16	13	7	9	10	5	16	7	8	13	4	8	9	8	16	7	13	8	12	13	8	9	9	7	
23	6	14	15	7	9	11	5	13	7	9	16	4	8	9	11	16	5	14	5	14	17	9	9	12	6	
24	6	16	14	9	12	11	5	12	8	9	16	4	9	11	9	16	8	11	7	11	14	8	9	8	5	
25 9.72	7	14	13	7	9	9	4	12	7	8	13	4	9	9	8	15	5	11	5	11	12	7	8	8	5	9.24
26 9.88	6	14	12	7	8	8	4	12	7	8	10	4	7	8	7	13	5	10	6	10	10	7	9	8	5	9.04
27	6	13	12	7	9	8	4	11	7	7	11	4	7	9	8	16	5	11	6	10	12	7	10	9	4	
28	6	16	13	8	11	11	4	13	6	7	14	5	7	9	9	16	5	13	6	13	16	7	10	10	5	
29 10.24	8	15	12	9	10	9	4	13	7	7	15	4	9	8	9	16	8	15	4	14	14	8	9	9	7	9.72
30	5	11	14	6	11	12	4	12	8	9	16	4	9	11	10	16	7	13	7	11	13	8	9	12	5	
31	9	16	12	9	12	11	4	13	11	11	16	5	9	12	10	16	7	14	7	14	15	7	9	9	6	
32 9.36	5	13	15	7	10	9	5	15	7	8	14	4	8	10	9	15	6	13	7	14	16	8	11	12	7	9.76

Section III.

After finding in Section II that reliable averages could not be obtained in all cases by combining the individual averages of 2,3,4 and 5 judges a study was made of averages obtained by random selection of individual judgments. As an illustration it was found that random selection, based upon gradings made by inexperienced graders, did not produce reliable results. The following cases may be taken as typical.

Table XIX may be read as follows: The first horizontal column at the top gives the series number of each sample graded and in the same vertical column in which the number of each sample occurs may be found the judgments on the sample. In each case there are 32 separate judgments on each sample or one each by 32 graders. The vertical column to the left gives the number of each grader. The graders were ranked in alphabetical order. In table XIX may be found the judgments of each of the 32 graders on the first 25 samples of Set I.

The averages of twenty five samples were obtained by random selection. These are shown by diagonal lines. To the right of the page, at the top, may be found the following averages: 10.44, 10.32, 9.12 and 9.32; and at the bottom 9.24, 9.04, 9.72 and 9.76. To the left may be found the averages, 9.72, 9.88, 10.24 and 9.36.

Conclusion. Averages obtained by random selection are not reliable.

Problem IV.

The problem. What is the minimum time of practice required in grading samples of handwriting before a variation low enough to be reliable can be secured?

(A) Individual variation.

Source of data. The material for this study consists of the first two gradings each of the sets graded in Study B counting the regradings of Set I as though they were made upon new sets. For instance, Set I, shown in Table VII was regraded on July 2 and 3. The individual variation will be computed as though it were a new set. This data is compiled in Table XX.

(B) Group Variation. The conclusions of Study B will be used and the second grading of Set III, Study A.

Personnel. Any material selected from either of the studies will be a cross section of the gradings taken from the entire group. This in Study A will be the gradings of five judges and of Study B of eleven judges.. These have been described in detail under the caption of "personnel" in Studies A and B.

Individual variation. The method of Study A was not effective in reducing individual variation but the method of Study B was helpful so an intensive study will be made of the individual variation of each of the eleven

judges on each of the first two gradings of the sets of samples taken in the order in which they were graded.

Table XX may be read as follows: The first horizontal column, reading from the left and top, gives the number of the sets taken in the order in which they were graded as Sets I, II the first regrading of Set I, Set IV etc. The variation between the first two daily averages of judge 1 on Set I, shown in the second horizontal column from the top, was .17 of a quality, the corresponding variation for judge 2 was .23 etc. The variation of judge 1 on the averages of the first two gradings of Set II was .09 of a quality, judge 2 had a corresponding variation of .33 etc.

Whenever an individual variation was found that was less than .20 of a quality it was indicated by drawing a red ring around the low variation. The increasing number of rings reading in the vertical columns from the left indicate the increasing number of judges securing low and trustworthy individual variation. A consistent grader throughout the study would have a succession of rings in the horizontal column in which his name appears. Judge 10 made the best record of the group. His variation, only on the first set, was too large to be reliable but after that it was trustworthy.

Table XX.

A summary of the arithmetical averages of each of the eleven judges on each of the first two gradings each of the sets of samples graded in Study B.

Set	I	II	I [*]	IV	V	VI	I	VII	VIII	I
Judge #1	(.17)	(.09)	(.11)	(.08)	(.20)	(.14)	.25	(.16)	(.20)	(.05)
2	.23	.33	(.13)	(.16)	(.14)	(.08)	(.03)	(.13)	(.06)	(.09)
3	(.18)	---	.31	(.02)	.22	.30	(.04)	(.15)	(.07)	(.05)
4	.72	(.02)	.64	(.16)	(.06)	.78	.31	(.01)	(.11)	.31
5	(.03)	(.03)	(.00)	.33	(.19)	.25	(.18)	(.13)	(.03)	(.03)
6	(.09)	(.04)	.29	.24	(.09)	.47	(.03)	.24	.27	(.01)
7	.30	.26	.58	(.14)	(.02)	.29	.21	(.05)	(.02)	.28
8	.37	.83	.69	.34	.39	(.10)	.32	.51	.41	(.12)
9	.41	.25	(.05)	(.20)	(.20)	.59	(.09)	.33	.29	(.08)
0	.30	(.14)	(.08)	(.03)	(.05)	(.14)	(.07)	(.00)	(.01)	(.00)
1	.26	(.01)	.46	(.09)	.71	(.00)	(.02)	(.16)	(.07)	(.15)

* Set I was regraded on three different occasions.

Judge 2, a grader of previous experience, made a very good record. He varied too much on the first two sets but after that his variation became less and was trustworthy during the remaining portion of the study.

Certain graders were quite erratic. For instance 4,6,7,8 and 9. At no point in the study did every member of the group secure a trustworthy variation. On the last regradings of Set I but two failed to secure the desired low variation but the group closely approximated this number on Sets IV, and V, the second regrading of Set I,VII and VIII.

Conclusions. The individual variation was so erratic that the minimum time of practice for individual variation, in order to reduce it within trustworthy limits, cannot be told from this study.

This table, taken as a whole, shows that individual variation may be reduced by independent practice without any instruction in the use of the scale. It also shows that some judges become more proficient in the use of the scale than others.

While it does not prove the fact, we may infer that some graders may never learn to use the scale with any degree of consistency and reliability.

Other judges, as number 1, may be able to secure and maintain low individual variation and the results be trustworthy by making a small allowance for individual variation. It will also be noted on Table XX that five judges, 2,3,5,10 and 11, had an individual variation that was reliable on the second regrading of Set I and maintained it throughout the remainder of the course. Just how long this would have continued cannot be told from this study.

Group variation: The methods of Study B did not tend to reduce group variation so no data will be evaluated from that study. The lowest group range obtained in Studies A and B was in Study A, Set III, the second grading.

Set III. Study A.

Date		11/6-10	
Student	# 1	10.51	11.13
..	# 2	11.19	11.02
..	# 3	11.12	11.22
..	# 4	11.15	10.89
..	# 5	11.14	11.39
Average		11.02	11.13
A.D.		.206	.140
% variability		1.87	1.26
Group range		.68	.50

The group range being .50 of a quality. Each of the judges had made 1209 gradings on samples of handwriting previous to the second grading of Set III and may therefore be said to be practiced in the use of the scale.

.50 of a quality is as much as the difference between the averages of two successive grades in handwriting and is not trustworthy.

Conclusion. The minimum time of practice, to reduce group variation, cannot be told from this study.

Problem V.

Section I.

The object of this problem is (a) to determine if variation may be reduced by many regradings upon the same samples and (b) to determine if variation is reduced more effectively by gradings made (1) close to each other in point of time as on successive days or by gradings (2) separated by intervals of weeks.

The problems. What will be the effect upon variation of many regradings upon the same set of samples (1) separated from each other in point of time and (2) many gradings not separated from each other in point of time but graded on successive days?

Section II

The problem. What will be the effect upon variation of many regradings on the same set of samples separated from each other in point of time?

Personnel and source of data. Set I was graded eleven times during the summer of 1914 by the eleven judges of Study B. For a more detailed description of the data and personnel see Study B, Section I.

In this problem only the first two gradings of Set I, which were made by inexperienced graders, and the last two, the tenth and eleventh gradings made after seven

weeks of practice, were used to draw conclusions from. The last two represent the last two of many regradings on the same samples made by experienced judges.

Method. (a) Individual variation: Any individual variation of less than .20 of a quality was considered trustworthy and the fact was indicated by a ring drawn around the variation as (.18) shown in Table XXI.

(b) Group variation. The group variation will be determined by a comparison of the average group range on the first two gradings of Set I with the average group range on the last two gradings.

Individual variation. On the first two gradings of Set I, see Table XXI, but four judges, 1, 3, 5, and 6, came within the desired limits of variation which is .20 of a quality. On the last two gradings, the tenth and eleventh, after seven weeks of practice judges number 4 and 7 were the only exceptions in the attempt to secure the desired low individual variation.

The ratio of the consistent judges on the first two gradings of Set I to the number engaged in gradings was four to eleven or .36. The ratio of the consistent judges on the last two gradings of Set I to the number grading was nine to eleven or .82. A comparison of the two ratios obtained reveals the fact that the number of consistent judges was more than doubled after seven weeks of practice.

Table XXI.

A summary of the individual averages of the eleven judges of Study B on the first, second, tenth and eleventh grading of Set I.

Number of grading	1st	2nd	Ind. var.	10th	11th	Ind. var.
Judge # 1	9.51	9.34	(.17)	9.35	9.30	(.05)
2	9.59	9.26	.23	9.21	9.30	(.09)
3	10.03	9.85	(.18)	9.32	9.37	(.05)
4	10.34	<u>9.62</u>	.72	9.34	<u>9.03</u>	.31
5	8.97	9.00	(.03)	8.71	8.68	(.03)
6	9.65	<u>9.74</u>	(.09)	9.51	9.50	(.01)
7	10.19	10.49	.30	8.71	<u>8.99</u>	.28
8	<u>9.85</u>	9.48	.37	8.71	8.59	(.12)
9	9.39	9.80	.41	8.69	8.61	(.08)
10	9.02	8.72	.30	9.68	9.68	(.00)
11	10.76	11.02	.26	8.81	8.96	(.15)
Group average	9.76	9.67		9.09	9.09	
Group range	1.79	2.30		.99	1.09	
A. D.	.436	.292		.335	.307	
% of variability	4.46	3.02		3.68	3.37	

Conclusion. Many regradings on the same set of samples separated from each other in point of time reduces individual variation.

Group variation. The average group range on the first two gradings of Set I was 2.04 qualities. The average group range on the last two gradings of Set I, after seven weeks of continuous practice, was 1.04 qualities. This is a reduction of 1.00 of a quality.

Conclusion. Many regradings on the same set of samples separated from each other in point of time reduces group variation but in this case the average range on the last two gradings was too large to be trustworthy.

Graph II shows the wide group variation.

Graph II.

Graph showing variability based upon the daily
arithmetical averages of the eleven judges of Study I.

Grading 1st 2nd 3rd 4th 5th 6th 7th 8th 9th 10th 11th

PLATE A - FOR SALE BY THE UNIVERSITY CO-OPERATIVE STORE, COLUMBIA, MO.

DEPARTMENT OF ENGINEERING, UNIVERSITY OF MISSOURI

The problem. What is the effect upon variation of many regradings on the same set of samples not separated from each other in point of time but graded on successive days?

Personnel and source of data. Sets IV and V were graded five times each by the five judges of Study A and the individual averages on these gradings furnished the data for Table XXII..

The material graded, personnel and method of grading had been described in detail under Study A.

Frequency of grading. Each set was graded five times on successive days. A month elapsed between the fifth grading of Set IV and the first grading of Set V. During this time no grading of samples of handwriting was done by the members of the class.

Individual variation. On the first two gradings of Set IV three judges, 1, 4 and 5, were consistent. . On the fourth and fifth gradings of Set IV all the judges were consistent.

That this did not happen accidentally may be seen by observing the next set graded by the same judges which was Set V. On the first two gradings judges 2, 3 and 5 were consistent while on the fourth and fifth gradings all were consistent.

Table XXII.

A summary of the individual averages of the first two and the fourth and fifth gradings on (a) Set IV and (b) Set V.

Number of grading	1st	2nd	Ind. Var.	4th	5th	Ind. var.
Judge # 1	10.13	10.22	(.09)	9.91	9.86	(.05)
2	9.55	10.08	.53	9.52	9.56	(.04)
(a) 3	10.12	9.54	.58	9.52	9.55	(.03)
4	10.06	10.10	(.04)	10.11	10.07	(.04)
5	9.36	9.27	(.09)	9.71	9.62	(.09)
Group range	.77	.95		.59	.52	

Number of grading	1st	2nd	Ind. Var.	4th	5th	Ind. var.
Judge # 1	7.89	7.40	.49	7.99	7.99	(.00)
2	7.71	7.62	(.09)	7.83	7.81	(.02)
(b) 3	8.04	8.16	(.12)	8.07	8.16	(.09)
4	9.09	8.69	.40	8.37	8.24	(.13)
5	8.59	8.69	(.10)	8.63	8.81	(.18)
Group range	1.38	1.29		1.80	1.00	

Conclusion. Many regradings upon the same samples close to each other in point of time is effective in reducing individual variation.

It is an interesting fact that but one of the graders that was consistent on the first and last two gradings of Set IV, judge number 5, was consistent on the corresponding gradings of Set V.

Group variation. The average group range on the first two gradings of Set IV, Table XXII(a), was .86 while the average group range on the last two gradings, the fourth and fifth, of the same set was .55 of a quality. .55 of a quality is more than the average difference between two successive grades in penmanship.

The corresponding average group ranges on Set V, Table XXII (b) was 1.33 and 1.00. This reduction is too small to be significant.

The average group range on Set V is much higher than on Set IV but lack of practice may partially account for this since a period of four weeks elapsed from the last grading of Set IV and the first grading of Set V.

Conclusion. Many regradings upon the same samples close to each other in point of time does not produce trustworthy results in reducing group variation.

Summary of Problem V. Individual variation. Many regradings on the same set of samples not separated from each other in point of time is more effective in reducing individual variation than many regradings separated from each other in point of time.

On the last two regradings of Sets IV and V of Study A every member of the group had a trustworthy variation while on the last two regradings of Set I of Study B not all the members of the group had trustworthy individual variation.

Group variation. Many regradings upon the same set of samples, whether made close to each other in point of time or removed, does not produce trustworthy results in reducing group variation.

Bibliography

- Ayres, L.P.-- A Scale for Measuring the Quality of Handwriting of School Children. Russell Sage Foundation, 1912.
- Bagley, W.C.--The Need of Standards for Measuring Progress and Results. N.E.A. 1912: 634-659.
- Charters, W.W.- A Study of the Thorndike Handwriting Scale. (An unpublished manuscript).
- Davenport, E.-Principles of Breeding. Ginn and Co.
- Elderton, W.P. and Ethel M. --Primer of Statistics. Macmillan Co.
- Freeman, F.N.--Some Issues in the Teaching of Handwriting. Elementary School Teacher, Vol.12, pp.1-7, 53-59.
- Freeman, F.N.-Problems and Methods of Investigation in Handwriting. Journal of Educational Psychology. Vol.3 pp.181-190.
- Freeman, F.N.-The Teaching of Handwriting. Houghton Mifflin Company, 1914.
- Gray, C.T.---The Training of Judgment in the Use of the Ayres Scale for Handwriting. Journal of Educational Psychology, February, 1915.
- Kelly, J.K.--Teachers' Marks. Teachers College Record, No.66.
- King, I., and Johnson, H.--The Writing Abilities of the Elementary and Grammar School Pupils of a School System Measured by the Ayres Scale. Journal of Educational Psychology, Vol.3, 514-520.
- Manuel, H.T.-The Use of an Objective Scale for Grading Handwriting. Elementary School Journal. Jan. 1915.
- Starch, D.---The Measurement of Handwriting. Journal of Educational Psychology. Vol.4. pp.445-464. Vol.6. pp.106-114.
- Thorndike, E.L.-Handwriting. Teachers College Record, 1912, 1914.

Thorndike, E. L. -- Social and Mental Measurements.

Tompson, Mary E. -- The Psychology and Pedagogy of Writing.

Whipple, G. M. -- Manual of Mental and Physical Tests.
Warwick and York, Baltimore.

*Approved
to Atlanta*

010-100734528

RECEIVED
JUL 7 1916
UNIV. OF MO.

378.7M71
XH946

Hurt1916ScanSpecs.txt

MU Libraries
University of Missouri --Columbia

MU Theses and Dissertations (Pre-1920)

Local identifier Hurt1916

Capture information

Date captured	09192016
Scanner manufacturer	Zeutschel
Scanner model	OS 15000
Scanning system software	Omni scan v. 12.4 SR4 (1947) 64-bit
Optical resolution	600 dpi
Color settings	grayscale, 8 bit and color, 24 bit
File types	tiff

Source information

Format	Book
Content type	Text
Source ID	010-100734528
Notes	Pages typed and single-sided. Text is light on graphs. Title page has signature and perforated property stamp. Call number written on page 1. Seven pages are foldout graphs. Inside front cover has two labels pasted in. Pocket slip pasted in at end. Inside back cover has stamp, barcode, and call number.

Derivatives - Access copy

Compression	Tiff compressed with LZW before conversion to pdf
Editing software	Adobe Photoshop CS5
Resolution	600 dpi
Color	Grayscale and color
File types	pdf
Notes	Grayscale pages cropped, canvassed, and images brightened. Color pages cropped and canvassed. Blank pages removed.