

Reliability of the Grading of High-School Work in English

Author(s): Daniel Starch and Edward C. Elliott

Source: *The School Review*, Sep., 1912, Vol. 20, No. 7 (Sep., 1912), pp. 442-457

Published by: The University of Chicago Press

Stable URL: <https://www.jstor.org/stable/1076706>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *The School Review*

JSTOR

RELIABILITY OF THE GRADING OF HIGH-SCHOOL WORK IN ENGLISH

DANIEL STARCH AND EDWARD C. ELLIOTT
The University of Wisconsin

I. *Problem and method of investigation.*—The reliability of the school's estimate of the accomplishment and progress of pupils is of large practical importance. For, after all, the marks or grades attached to a pupil's work are the tangible measure of the result of his attainments, and constitute the chief basis for the determination of essential administrative problems of the school, such as transfer, promotion, retardation, elimination, and admission to higher institutions; to say nothing of the problem of the influence of these marks or grades upon the moral attitude of the pupil toward the school, education, and even life. The recent studies of grades have emphatically directed our attention to the wide variation and the utter absence of standards in the assignment of values. Dearborn pointed out in his investigation¹ the large inequalities in the standards of grading employed by different teachers. Of two instructors in the same department one gave 43 per cent of his students the grade of "excellent" and to none the grade of "failure," whereas the other gave to none of his students the grade of "excellent" and to 14 per cent the grade of "failure." The wide difference in this instance is no doubt due in part to a difference in the students and in the nature of the work, but largely to a difference in the standards of marking.

In order to determine precisely the personal equation in evaluating the work of pupils, it is necessary to eliminate all other causes of variation. The mere comparison of marks assigned by different teachers to their classes will not reveal this personal factor stripped of the other variable elements, such as difference in amount and kind of work covered by the class, emphasis upon different topics,

¹ W. F. Dearborn, *School and University Grades* (Bulletin of the University of Wisconsin, No. 368), p. 57.

differences in teaching ability, and differences in the pupils themselves. To discover the variability in judging the merits of an examination-answer paper it is necessary to have the same paper graded independently by different teachers.

Concretely, the problem of this investigation was to determine the range of variation and the reliability of the marks assigned by

5 Every parenthetical expression should be set off by a comma.

(5). There ~~are~~ are two corridors in this building, one above the other.

1. Every initial letter should be followed by a period.

(1) C. K. Rannon built the "Calumet K"

2 Every assertive sentence should be ended by period.

(2). The fact that the earth is round is undisputed.

II

In writing a paragraph, the following requirements should be observed;

- (a) Unity. (b) Coherence. (c) Emphasis.
(d) Proportion. (e) Parallel Construction.

Extract from paper A (reduced about one-half).

teachers to examination-answer papers. For this purpose, two examination-answer papers written by two pupils at the end of the first year's work in English,¹ together with the questions, were obtained from one of the largest high schools in Wisconsin. English papers were chosen for this first test because it was thought that personal differences were more apt to appear in the estimating

¹ Similar investigations with answer papers in geometry and history are now under way.

of merits of a language paper; and also because the findings with English might possibly have a certain significance for the grading of work in foreign languages. Plates of these answer papers were made, and several hundred copies were printed upon the same kind

- 4) The comma is used in an answer, as:
Yes, I will come.
- 5) Use the comma to make thought clear.
- 6) The period is used at the end of a sentence.
as, The day is warm.
- 7) The period is used in abbreviations, as,
for Captain use capt.

II a paragraph should be:

- a) coherent
 - b) unity
 - c) ~~narration~~
 - d) proposition
 - e) emphasis.
- and contain

Extract from paper B (reduced about one-half).

of paper (foolscap) as was used originally by the pupils. In this manner the handwriting, the errors and changes made by the pupils, the neatness, and so forth, were reproduced exactly as in the original papers.

EXAMINATION QUESTIONS

1. Give five rules for the use of the comma, two for the period, and illustrate each use by a sentence.
2. Give five requirements to be observed in the structure of a paragraph.
3. Write a brief business letter.
4. Define narration, coherence, unity; classify sentences rhetorically and grammatically. Illustrate or define.
5. Name all the masterpieces studied this year and name the author of each. (Answer: Irving's *Sketch Book*, Hawthorne's *Mosses from an Old Manse*,

Shakespeare's *Merchant of Venice*, Whittier's *Tent on the Beach*, Scott's *Ivanhoe*.¹)

6-10. Write a three-paragraph essay, narrative, descriptive, or both combined.

PAPER A

I

English

Every introductory clause or phrase should be set off by a comma.

(1) In the morning, when it is light, we will go.

2. "But" is a word of opposition and should always be preceded by a comma.

(2) Many are called, but few are chosen.

3. Words, phrases, and clauses forming a series should be separated by a commas.

(3). The warm, humid air is oppressive.

4 A noun or pronoun in the direct address should always be followed by a comma.

(4). Mary, give me a book.

5 Every parenthetical expression should be set off by a comma.

(5). There are two corridors in this building, one above the other.

1. Every initial letter should be followed by a period.

(1). C. K. Bannon built the "Calumet K"

2. Every assertive sentence should be ended by period.

(2). The fact that the earth is round is undisputed.

II

In writing a paragraph, the following requirements should be observed,;

(a) Unity. (b) Coherence. (c) Emphasis. (d). Proportion. (e) Parellel Construction.

III

1002 Jenifer St.,
Madison, Wis.,
June 14th, 1910.

Funk and Wagnall's Co.,
44-60 E. 23d St., New York.

Gentlemen:

Enclosed please find draft for five dollars (\$5⁰⁰/₁₀₀), for which send to the above address the "Literary Digest" for one year. I should like to begin with the first July number.

Yours respectfully,

IV

(1). Narration is that form of discourse in which the writer or speaker presents a series of related events.

(2). Coherence is that principle of rhetoric which demands that ideas be presented in their natural and logical order.

¹ This answer, of course, was given with the questions sent out to the teachers.

(3). Unity is that principle of rhetoric which demands that one topic and one only be discussed in a paragraph, and one thought and one only be expressed in a single sentence.

(4). (a) Sentences are classified grammatically, as to form and use.

As to form, sentences are classified as simple, compound, complex, and compound-complex. A simple sentence is one consisting of one subject and one predicate, either or both of which may be compounded. A compound sentence is one consisting of two or more simple sentences. A complex sentence is one consisting of an independent clause, and one or more dependent clauses. A compound-complex sentence is one consisting of a compound sentence and a complex sentence.

As to use, sentences are classified as assertive, which states a fact; imperative, which commands or orders; interrogative, which asks a question; and exclamatory, which expresses a sudden outburst of feeling

(b). Sentences are classified rhetorically as loose, balanced, and periodic.

A loose sentence is one that may be brought to a close in two or more places and in each case make complete sense.

A balanced sentence is one consisting of two parts similar in structure, but often contrasted in meaning.

A periodic sentence is one that reserves the most important thought for the end of the sentence.

V

Masterpieces studied this year and names of authors are,

"Ivanhoe" by Sir Walter Scott.

"The Sketch Book" by Washington Irving.

"The Tent on the Beach" by John G. Whittier.

"Mosses from an Old Manse" by Hawthorne.

"Rob Roy" by Sir Walter Scott.

"Last of the Mohicans" by J. F. Cooper.

"The Man Without a Country" by Everett E. Hale.

"The Merchant of Venice" by Shakespeare.

VI-X

"An Irish Hall of Fame."

"An Irish Hall of Fame" will form part of the equipment of a new Boston College, now being erected by the Jesuit order. This hall with other buildings will be a memorial to Daniel O'Connell. (Daniel O'Connell, the great Irish Statesman was born in 1775. He took the side of the Catholics, claiming that they should equal rights with the Protestants of Ireland and that Ireland should have its own government. He secured the rights for the Catholics and kept on lecturing for free government. He was arrested, tried, and acquitted. His friends forsook him and he died of a broken heart in 1847).

The center hall of this building is one hundred and sixty feet high. Surrounding the hall are two corridors, supported by high Gothic arches, one above

the other. Opening onto these corridors are thirty-two minor apartments, representing the thirty two counties of Ireland. These rooms will be furnished as museums, illustrating the history of each county. It is also planned to have friezes representing the spiritual and material progress of the Irish race.

The largest room in this building is fitted with all the modern theatrical equipments, and has a seating capacity of two thousand. There are also many committee rooms in the building which makes it an ideal place for conventions. Great pains are being taken in selecting the architecture of this building, so that it will be in keeping with the high memorial purpose. The cost of building is two hundred thousand dollars.

PAPER B

English

- I 1) All parenthetical expressions should be set off by commas. as; George Washington, the first President of the United States, was the father of his country.
- 2) All introductory clauses and phrases should be set off by commas: as, Fido, the dog that ran away, is found
- 3) All words in apposition should be set off by the comma, as; Queen, the horse, is dead.
- 4) The comma is used in an answer, as; Yes, I will come.
- 5) Use the comma to make thought clear.
- 1) The period is used at the end of a sentence. as, The day is warm.
- 2) The period is used in abbreviations. as, for Captain use capt.
- II A paragraph should be:
- Coherent.
and contain
 - Unity
 -
 - proportion
 - emphasis.
- A paragraph may be developed in the following ways
- by definition and explanation
 - example and illustration
 - comparison and contrast
 - narrative.
 - Proof.

504 Wis. Ave
Madison, Wis
June 14, 1910.

III Marshall Field & Co.

Dear Sirs,—

Please send to above address five yards of material to match the enclosed sample:

Yours Truly.

IV

Narration is that form of discourse in which the writer or speaker gives a series of related events.

Coherence is that principle of rhetoric which demands that ideas be presented in their natural and logical order.

Unity is that principle of rhetoric which demands that one thought, and one only be discussed in a paragraph, and that one idea, and one only be expressed in a sentence

Sentences are classified gramatically, as to form and use: as to use they are classified as, Imperative, declaritive, exclamatory and interrogative.

An imperative sentence is one that gives a command

A declarative sentence is one that asserts.

An Interrogative sentence is one that asks a question.

An exclamatory sentence contains an exclamation, as, Alas, the bird is dead.

As to form, they are classified as Simple, complex and compound.

A simple sentence contains a subject and predicate, either, or both of which may be compound.

A complex sentence is one which contains one Independent and two or more dependent clauses.

A compound sentence contains two or more simple sentences.

Sentences are classified rhetorically as; loose, periodical and balanced.

A loose sentence is one which may be brought to a close in two or more places, and in each case make complete sense. A periodical sentence is one which retains the most important part until the last.

A balanced sentence, contains two parts, which are alike in construction, but contrasted in meaning

V

Mosses of an Old Manse, Hawthorne

Sketch book, Irving

Tent on the beach, Whittier.

Last of the Mohicans, Cooper.

John Halifax, Mulock

Merchant of Venice, Shakespeare.

Ivanhoe }
Kenilworth } Scott.

VI-X

My puppy.

My puppy is an affectionate little animal, about three months old. He is white, with so many little black spots on him that many people think that he should have been named "Freckles", but we named him "Patsy" instead.

He is very bright, and learns quickly.; When we first got him, he made a great comotion in the dining room, if he did not get something to eat when we were eating, but now he knows better and lies quietly under the table all

through the meal, without even chewing our shoe strings. The only time that he gets cross, is when he runs into the back yard of the people who live next to us, Then I have to go. after him because he finds meat and such things that the people have put out for their dog, but are not good for a puppy, He does not like this because he is very fond of meat.

At first the "next door" dog, did not like Patsy, and was very jealous of him, because he did not receive so much attention, as he had before. But now they are becoming fast friends and are together every minute. Patsy has given us many a scare, for, when we have been sitting on the front porch, he has run off and we have been sure that he was lost, or run over by an automobile, and gone to look for him, but on our return we have always found him dozing comfortably on the bach mat, or playing with Brandy, which is the name of the "next door" dog.

A set of questions and a copy of each of the two answer papers were sent to each of two hundred high schools in the North Central Association, with the request that the principal teacher of first-year English grade these two papers according to the practices and standards of the school.

One hundred and fifty-two out of the two hundred papers were returned. Five were discarded either because they had not been graded numerically or because they did not give the passing grade of the school. In two schools the passing mark was 60 and in three it was 80. These five also were discarded from the tabulations because of the obvious difficulty of evaluation. Of the remaining one hundred and forty-two schools, fifty-one had a passing mark of 70 and ninety-one had a passing mark of 75.

That the grading was done carefully is evident from the fact that, with a few exceptions, separate marks and comments were given upon the answer to each question.

The grades assigned by the one hundred and forty-two teachers can best be represented by the distribution curves on the following pages.

II. *Explanation of charts.*—The range of possible marks is indicated along the base line of each chart and the number of times each grade was given is indicated by the number of dots above that grade. Thus in Fig. 1 the grade 80 was given by two teachers and the grade 85 by four teachers. The two papers are designated as A and B.

Fig. 1 gives the values assigned by ninety-one teachers to paper A in schools whose passing grade is 75. Fig. 2 gives the values

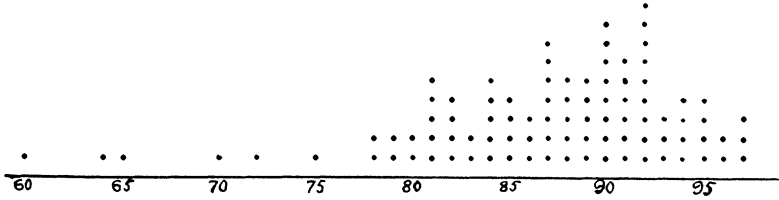


FIG. 1.—Paper A. Passing grade 75. Median 88.3

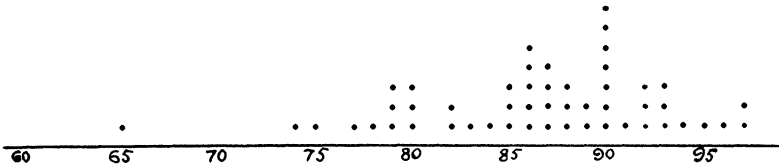


FIG. 2.—Paper A. Passing grade 70. Median 87.2

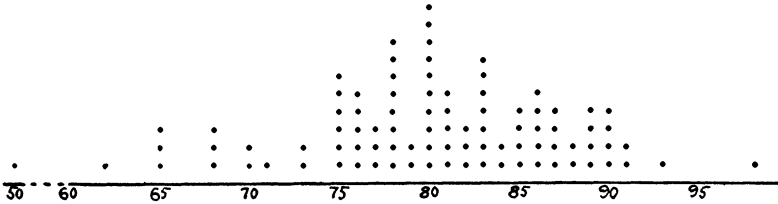


FIG. 3.—Paper B. Passing grade 75. Median 80.4

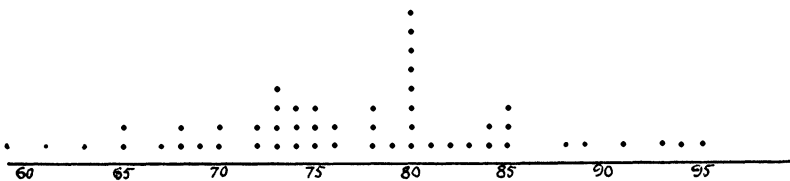


FIG. 4.—Paper B. Passing grade 70. Median 78.8

assigned by fifty-one teachers to the same paper in schools whose passing grade is 70. It will be noted that there is a difference of

only 1.1 points between the medians¹ of the two groups, although there is a difference of 5 points between the passing grades.

Figs. 3 and 4 represent the marks given by the same two groups to paper B. Here again there is a difference of only 1.6 points between the medians. Paper B is considerably poorer. It is judged 7.9 points lower by the first group of teachers and 8.4 points lower by the second group.

Fig. 5 is a composite chart showing the values given to paper A by the entire one hundred and forty-two teachers. The values

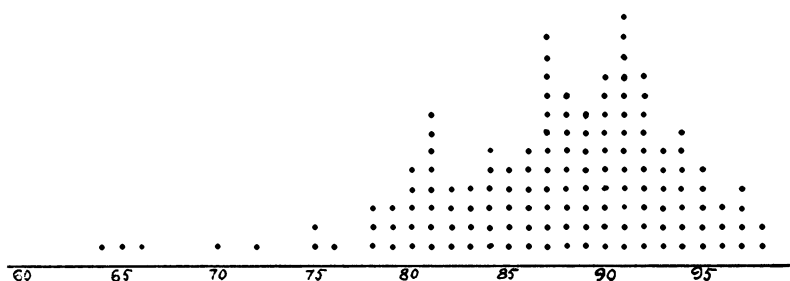


FIG. 5.—Paper A. Passing grade 75. Median 88.2. Marks assigned by schools whose passing grade is 70 are weighted by one point.

assigned by teachers in schools whose passing grade is 75 are represented as in Fig. 1, while the values assigned by teachers in schools whose passing grade is 70 are all raised one point, because the difference between the medians is approximately one point. Thus the combination of the two sets of schools in one chart is entirely fair.

Fig. 6 represents a similar combination of the marks of paper B. In this chart, however, the marks assigned by the teachers in schools whose passing grade is 70 are raised two points, because here the difference between the medians is approximately two points.

These composite charts show more clearly just where the bulk of the grades lie. They also bring out the fact that the range of marks is approximately the same for a small group of teachers as

¹ The median is the grade which is as often above as below the grade assigned. It is roughly equivalent to the average. It is used here because it represents the central tendency more accurately than the average would.

for a large group. In fact, almost any ten marks picked out at random will cover nearly the entire range of variation.

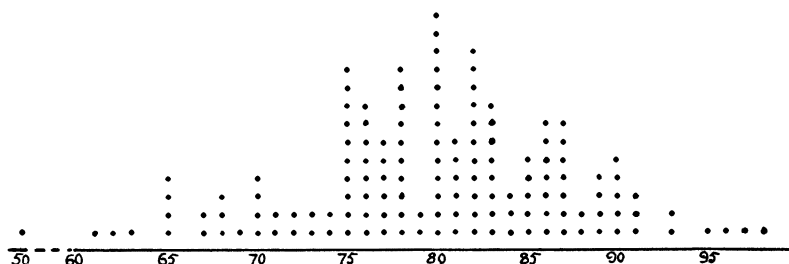


FIG. 6.—Paper B. Passing grade 75. Median 80.2. Marks assigned by schools whose passing grade is 70 are weighted by two points.

Figs. 7 and 8 represent the marks given to the two papers by eighty-six students in the course on the teaching of English in the

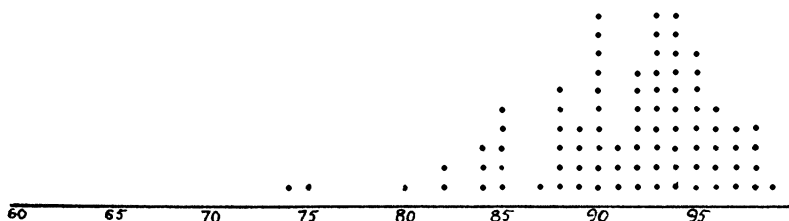


FIG. 7.—Paper A. Assumed passing grade 70. Median 92.4. Marks assigned by students in the course on the teaching of English, at the University of Wisconsin.



FIG. 8.—Paper B. Assumed passing grade 70. Median 84.5. Marks assigned by the students in the course on the teaching of English, at the University of Wisconsin.

University of Wisconsin. Only a few of these persons had had teaching experience. For the sake of definiteness, 70 was assumed as the passing grade. The range of variation is approximately the

same for the grades assigned by the students as for the grades assigned by the teachers. A difference worth noting, however, is that the students graded more leniently than the teachers. The median of the marks assigned by the teachers to paper A is 87.2 (Fig. 2), whereas the median of the marks assigned by the students is 92.4. The medians for paper B are 78.8 (Fig. 4) and 84.5, respectively. The students thus graded the first paper 5.2 points, and the second paper 5.7 points, higher than the teachers.

Figs. 9 and 10 represent the marks given to the two papers by ninety-eight students in the course on educational measurements

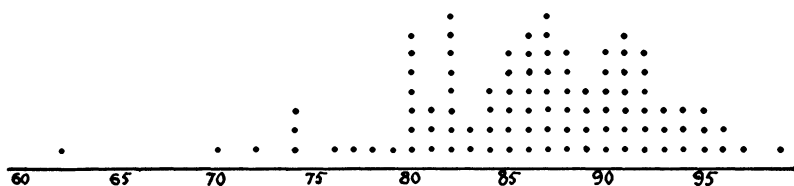


FIG. 9.—Paper A. Median 86.7. Marks assigned by the students in the course on educational measurements, at the University of Chicago.

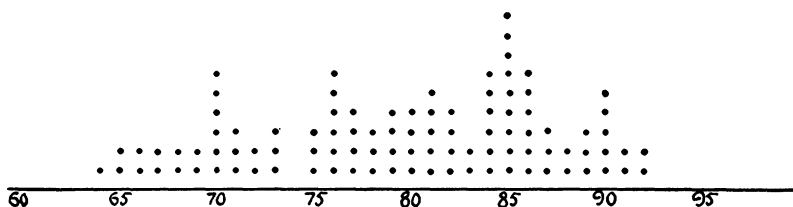


FIG. 10.—Paper B. Median 80.5. Marks assigned by the students in the course on educational measurements, at the University of Chicago.

(Summer Quarter, 1911) in the University of Chicago. This class was composed almost entirely of superintendents, principals, and teachers, a few of whom were not accustomed to grading English papers. The large majority, however, were experienced in marking papers, and regarded themselves capable of evaluating them with considerable accuracy. Their results are of special interest because of the fact that the standards for marking in public school systems, to a large extent, are set and scrutinized by the superintendents and principals. Nevertheless, the medians and the ranges of their

marks are almost identical with those in Figs. 1 to 6. Consequently, their judgments are no less variable than those of the teachers.

III. *General results.*—1. The first and most startling fact brought out by this investigation is the tremendously wide range of variation. Teachers usually state, when asked about differences in marking, that the grades of the same paper assigned by different teachers might differ at the most 10 points. It is almost shocking to a mind of more than ordinary exactness to find that the range of marks given by different teachers to the same paper may be as large as 35 or 40 points.¹ In Fig. 6 one-half of the marks lie between 83 and 92, and in Fig. 7 one-half lie between 75 and 85.

The two papers, A and B, were marked 80 and 75 respectively by the teacher under whom the pupils had taken the course. The passing grade in this school is 70. The medians of the marks given by the other schools whose passing grade is 70 are 87.2 and 78.8 (Figs. 2 and 4). Hence the teacher of these pupils marked their papers considerably lower than the teachers in the other schools.

Another interesting fact is the bearing of these data upon the question of promotion and retardation. The pupil who wrote paper B, the poorer of the two, received from his teacher a mark 5 points above the passing grade, whereas twenty-two out of the one hundred and forty-two teachers (Figs. 3 and 4) did not give a passing grade to this pupil. Therefore it may be easily reasoned that the promotion or retardation of a pupil depends to a considerable extent upon the subjective estimate of his teacher.

Even the standard of a given teacher is more or less variable and indefinite. It is sometimes said that even if different teachers give different marks, yet the relative estimates of the papers will be the same. To what extent is it true that if a teacher gives a high grade to one paper he will give a correspondingly high grade to the other, or vice versa? The curves in Fig. 11 give a definite answer to this question. The points along the horizontal axis represent the individual teachers and the points along the vertical axis the grades. The points at which the two curves intersect a given ordinate are the marks given by the same teacher to the two papers.

Evidently if it were true that a given teacher marked both

¹ The wide range of these marks is perhaps due, but we believe only to a small extent, to the differences in method of teaching and in the emphasis and importance placed by different teachers on different aspects of English.

papers correspondingly high or low, the two curves would be parallel. In general both have a descending direction, but the variations are by no means uniform. Thus paper B is marked on the average 8 points lower than paper A, yet nineteen of the one hundred and forty-two teachers marked it higher than paper A, and twenty-three marked it 15 or more points lower than paper A.

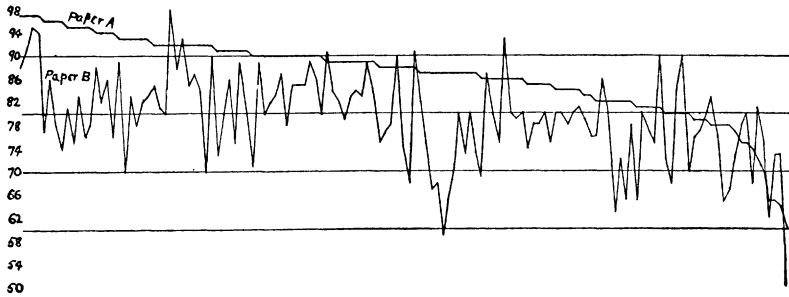


FIG. 11.—Showing individual variability of teachers in grading paper A and paper B.

The range of marks of paper B is somewhat wider than that of paper A, because it is a poorer paper, leaving more room for variation in the upper marks. Paper A is nearer to the 100 mark and so is more limited on that side of the distribution surface.

Professor Jacoby of Columbia University made an investigation similar to the present one but on a much smaller scale. He submitted for grading eleven astronomy papers to six professors of astronomy. The marks on a scale of 10 were as follows:¹

	A	B	C	D	E	F
1.....	9.0	9.0	8.5	7.2	9.0	7.3
2.....	7.0	6.6	7.0	5.9	6.0	6.5
3.....	9.0	9.0	8.8	7.2	8.0	8.0
4.....	10.0	9.4	9.9	8.0	10.0	9.2
5.....	7.0	6.2	6.7	5.8	7.0	5.9
6.....	10.0	9.8	9.6	7.6	10.0	9.5
7.....	6.0	5.8	6.3	4.6	7.0	5.4
8.....	9.0	9.3	9.7	8.0	9.0	8.8
9.....	8.0	5.7	9.0	6.7	10.0	8.7
10.....	10.0	8.5	9.1	6.2	9.0	9.0
11.....	9.0	9.0	9.5	6.1	8.0	9.0
Average	8.5	8.3	8.6	6.7	8.5	7.9

¹ H. Jacoby, "Note on the Marking System in the Astronomical Course at Columbia College, 1909-1910," (*Science*, XXXI, 819).

Professor Jacoby makes the following comment on this table: "Making due allowance for this circumstance in the case of Professor D,¹ there is a very close accord in the marks given by the various professors." "It would appear . . . that the marking system is more precise than some critics would have us believe. Possibly this may be due to the fact that astronomy is an exact science."

Such an interpretation as this can scarcely be made of the above table. Even when we omit Professor D's marks, the range of the marks given to paper nine is 4.3, to paper one 1.7, to paper seven 1.6, to papers ten and eleven 1.5, and so on. Such wide variations certainly impeach the reliability of the marks. The range in the astronomy marks is on the whole not so large as in the English marks. This is due partly, perhaps, to the fact that astronomy is an exact science, and partly to the fact that the astronomy papers were graded by only six examiners, who were more expert in their field than the teachers of English were in theirs.

Just the opposite of Professor Jacoby's conclusion would be the more correct interpretation, at least of the English grades. Marks are far less precise than the majority of teachers and pupils believe.

The probable error² of grades calculated on the basis of the data in Figs. 5 and 6 is approximately 4.5. This means that the individual marks deviate on the average 4.5 points from the median or average of the entire group of marks. The fact of such a large probable error shows the absurdity of marking to the fractional part of one point, as was done in quite a number of papers. One paper was even graded as fine as 79.9. The probability is that at best any one teacher's mark is 4 or 5 points from the true mark, if the average mark given by a large number of teachers may be regarded as the true mark.

2. As already stated above, students without teaching experience mark more generously than teachers.

¹ Professor D seems to have regarded 5.0 instead of 6.0 as the passing mark. Hence his marks are somewhat lower than the others.

² The probable error is a statistical term which roughly is equivalent to the average amount of error. It is an index of the reliability or variability of a set of measurements.

3. Small high schools tend to mark somewhat more leniently than large high schools, but the range of variation is the same. Figs. 12 and 14 show the marks given to papers A and B respectively

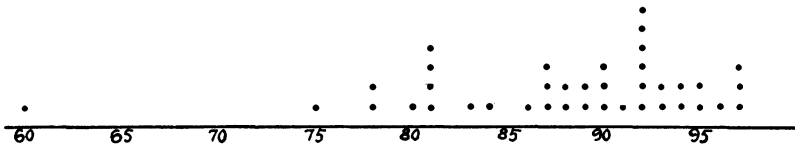


FIG. 12.—Paper A. Small schools. Passing grade 75. Median 89.5. Marks assigned by schools whose passing grade is 70 are weighted by one point.

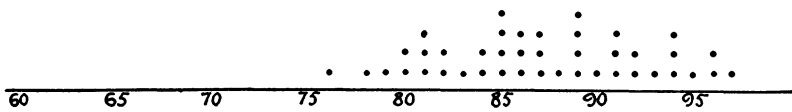


FIG. 13.—Paper A. Large schools. Passing mark 75. Median 86.8. Marks assigned by schools whose passing grade is 70 are weighted by one point.

by high schools with one hundred and fifty pupils or less. Figs. 13 and 15 represent the marks assigned to papers A and B respectively by high schools with one hundred and fifty-one pupils or



FIG. 14.—Paper B. Small schools. Passing grade 75. Median 82.0. Marks assigned by schools whose passing grade is 70 are weighted by two points.



FIG. 15.—Large schools. Passing grade 75. Median 80.3. Marks assigned by schools whose passing grade is 70 are weighted by two points.

more. The smaller schools graded paper A 2.7 points higher and paper B 1.7 points higher than the larger schools. Although this difference is small, it appears in both papers, and probably indicates a real tendency among smaller schools to grade more liberally.