Reliability of Grading Work in Mathematics

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

Source: *The School Review*, Apr., 1913, Vol. 21, No. 4 (Apr., 1913), pp. 254-259
Published by: The University of Chicago Press
Stable URL: https://www.jstor.org/stable/1076246

RELIABILITY OF GRADING WORK IN MATHEMATICS

DANIEL STARCH AND EDWARD C. ELLIOTT
University of Wisconsin

The present article is a sequel to the recent investigation of grading work in English[1] which revealed rather wide variations and differences among teachers in evaluating the same examination paper. It has been urged that marks in determining the merit of language work would necessarily vary considerably because of the personal and subjective factors involved, and that the situation would be very different in an exact science such as mathematics. Pursuant to this suggestion we have made a similar investigation with a geometry paper. This paper was written as a final examination by a pupil in one of the largest high schools in Wisconsin. Plates of this answer paper were made and several hundred copies were printed upon foolscap, thus exactly reproducing the original in every detail.

QUESTIONS

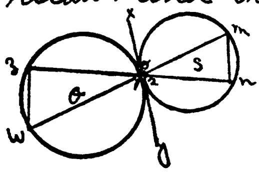
Choose 8, including one selected from 4, 6, and 8.

1. Two triangles having the three sides of one equal, respectively, to the three sides of the other, etc. Prove.
2. Prove that every point in the bisector of an angle is equally distant from the sides of the angle.
3. An angle formed by two intersecting chords is measured by, etc. Prove.
4. If the middle points of two opposite sides of a quadrilateral be joined to the middle points of the diagonals, the joining lines form a parallelogram.
5. To construct a mean proportional to two given lines. Explain fully.
6. AM is a chord of a circle, *xy* is a diameter perpendicular to AN and intersecting AM at O. XO is 10 in. and *ax* is 20 in. Find the diameter of the circle.
7. The ratio of the areas of two similar triangles is equal to, etc. Prove.
8. Find the area of a right triangle whose hypotenuse is 1 ft. 8 in. and one of whose legs is 1 ft.
9. The sum of the interior angles of a triangle is equal to, etc. Prove.
10. If two circles are tangent, and two secants are drawn through the point of contact, the chords joining the intersections of the secants and the circumferences are parallel.

[1] D. Starch and E. C. Elliott, "Reliability of the Grading of High School Work in English," *School Review*, XX, 442–57.

Answer to question X.

X If 2 Os are tangent and 2 secants are drawn through the point of contact the chords joining the intersection of the secants and the circumference are //.



Given the 2 ⊙ O and S tangent and the 2 secants zn and Wm connect their extremities and prove zW and mn are //. Draw the common tangent x y.

∠z is measured by ½ arc W O an inscribed ∠ is measured by ½ the intercepted arc.

∠1 is measured by ½ arc W O a tangent and a chord is measured by ½ its intercepted arc.

∠1 = ∠z. ∠m is measured by ½ arc n O ~~∠2 is meas~~ ∠ an inscribed ∠ is measured by one half its intercepted arc. ~~a line~~ ∠2 is measured by ½ arc n o a tangent and a chord is measured by ½ its intercepted arc.

∠1 = ∠2 because a line meeting a tangent at the point of tangency is ⊥ to it. z o is ⊥ to x y and n o is ⊥ to x y ∠1 = ∠2 rt ∠s.

∠1 = ∠z ∠1 = ∠n things = to the same thing are = to each other ∠z = ∠n. 2 st line cut by a transversal so the alternate interior ∠s are = the line are // ∴ zW and mn are //.

[Reduced about one-half]

A set of questions and a copy of the answer paper were sent to approximately 180 high schools in the North Central Association, with the request that the principal teacher in mathematics grade this paper according to the practices and standards of the school.

One hundred and forty papers were returned. Twelve had to be discarded because some of the data called for were not given.

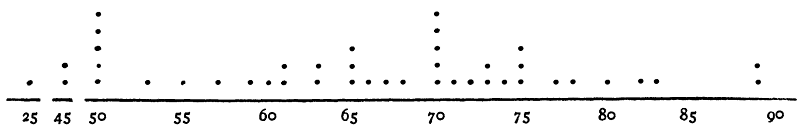


FIG. 1.—Passing grade 70. 43 schools. Median 67. Probable error 8.

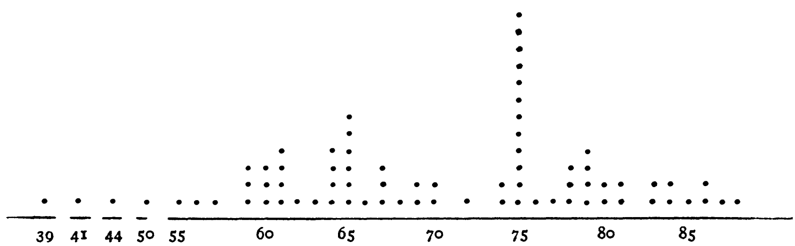


FIG. 2.—Passing grade 75. 75 schools. Median 70. Probable error 7.2.

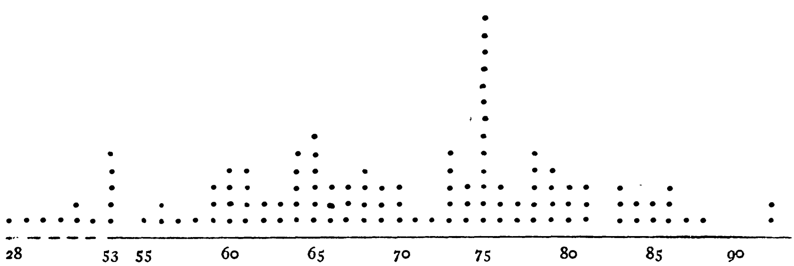


FIG. 3.—Passing grade 75. Marks assigned by schools whose passing grade is 70 are weighted by 3 points. Median 70. Probable error 7.5.

Of the remaining 138, 43 came from schools whose passing grade is 70, 75 from schools whose passing grade is 75, and 10 from schools whose passing grade is 80. The papers show evidence of having been marked with unusual care and attention. Separate grades and comments usually accompanied the answer to each question.

The grades thus assigned are represented by the distribution charts in Figs. 1, 2, and 3. The scheme of these charts is selfevident. The range of marks is indicated along the base line 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 70 was assigned by 5 teachers. The marks assigned by 10 schools whose passing grade is 80 are 72, 80, 83, 80, 58, 50, 50, 75, 73, 70.

Fig. 1 gives the values assigned by 43 teachers in schools whose passing grade is 70. Fig. 2 gives the values assigned by 75 teachers in schools whose passing grade is 75. The median indicates the central measure. It is roughly, but not exactly, equivalent to the average. It is used here in preference to the average because it represents more correctly the central tendency than the average would. The probable error is roughly, though not exactly, equivalent to the average amount of error or deviation of the mark from the median.

Fig. 3 is a composite chart showing the values assigned by the entire group of teachers. The values assigned by the teachers in schools whose passing grade is 75 are represented as in Fig. 1, while the values assigned by the teachers in schools whose passing grade is 70 are all weighted by three points because the medians of the two groups differ by that amount.

The investigation shows the extremely wide variation of the grades even more forcibly than our study of English marks. The distribution considered purely from the statistical standpoint is a normal distribution just like that of any set of mental or physical measurements. But the alarming fact is the wide range of the distribution.

A geometry paper was used because of the current assumption that a mathematical paper can be graded with mathematical precision. Our investigation shows that the marks of this particular geometry paper vary even more widely than the marks of either English paper used in the former study. The probable error of the geometry marks is 7.5 (Fig. 3), whereas the probable error of the English papers was 4.0 and 4.8 respectively.

A little analysis, however, will show the absurdity of assuming greater precision in evaluating a mathematical paper than in evaluating a language or any other kind of paper. While it is true that there can be no difference of opinion as to the correctness of a demonstration, yet there are countless ways in which the demonstration may be worked out, involving the succession of the steps, the use of theorems and definitions, the neatness of the drawings, and most of all the relative value of each particular demonstration or definition in the evaluation of the paper as a whole. Obviously the complication of factors is as intricate in one sort of paper as in another.

Why the marks of this particular paper vary even more widely than those of the English papers is to be sought in the fact that this geometry paper allowed of two fairly distinct ways of evaluation. The form, make-up, and appearance of the paper were of decidedly poor quality. Some teachers entirely disregarded these elements while others imposed a heavy penalty upon the paper on their account. In many instances this was indicated by the comments on the papers. But even this difference in viewpoint alone does not explain the extremely high or extremely low marks. For example, one teacher gave the paper a mark of 50 and said that he had deducted 4 points for spelling. Another marked it 45 and stated that he had made no deduction for poor form. Still another one marked it 75 including a penalty for form, or 85 excluding a penalty for form. Furthermore the amount that was subtracted for careless make-up ranged from 3 points in the case of one teacher to 13 points in the case of another.

It is therefore fully evident that there is no inherent reason why a mathematical paper should be capable of more precise evaluation than any other kind of paper. In fact, the greater certainty of correctness or incorrectness of a mathematical demonstration or definition may even contribute slightly to the wider variability of the marks, because the strict marker would have less occasion to give the pupil the benefit of the doubt.

In the next place, the criticism might be offered that the wide variation of the marks is due to the fact that the paper was graded by schools scattered over a large area, each one having a different standard of attainment. A propos of this point we may note that the school from which the paper was obtained has five teachers of geometry, each of whom graded the paper independently as follows: 70, 65, 60, 70, and 59, average 64.8, mean variation 4.2.

In a large high school in Ohio the paper was graded by four teachers of geometry as follows: 76, 75, 67, and 61, average 69.8, mean variation 5.8. In both of these schools the passing grade is 70.

Finally we may raise the question: How much variation is there in the marks assigned to the answer of any individual question? Sixty-two of the returns contained marks for the answer to each separate question. Forty-nine were graded on a scale of 0 to $12\frac{1}{2}$, and thirteen on a scale of 0 to 10. The marks of the latter given to

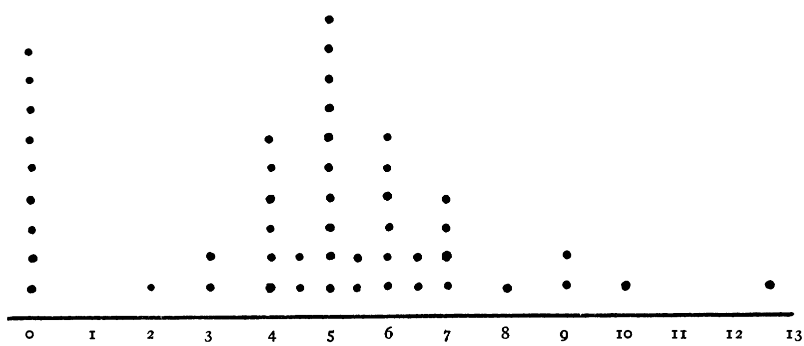


FIG. 4.—Grades of answer to question ten. Median 5.1. Probable error 1.1

the answer for question ten, which is reproduced at the beginning of this paper, were as follows: 5, 5, 0, 0, 5, 3, 4, 2, 3, 2, 5, $6\frac{1}{2}$, 5, average 3.5, mean variation 1.7.

Fig. 4 exhibits the distribution of the values assigned to the answer for question ten by the 49 teachers who graded on the scale of 0 to $12\frac{1}{2}$. The median is 5.1 and the probable error is 1.1. If we transpose this probable error into terms of the usual scale of 0 to 100 by multiplying it by 8, we obtain a probable error of 8.8. This is nearly the same as the probable error of all the marks in Fig. 3, namely 7.5. Hence we see that the marks of the answer for a single question of the paper vary about as widely as those of the entire paper.