R E P O R T R E S U M E S

ED 014 132 EA 000 789

AN INVESTIGATION OF ACHIEVEMENT GRADING BASED ON SCHOLASTIC ABILITY DISTRIBUTION.

BY- MASON, GEOFFREY

PUB DATE FEB 67

EDRS PRICE MF-$0.25 HC-$0.36 7P.

DESCRIPTORS- *GRADING, GRADES (SCHOLASTIC), *ELEMENTARY SCHOOL TEACHERS, *INTELLIGENCE QUOTIENT, *ACHIEVEMENT RATING, *ELEMENTARY SCHOOL STUDENTS, GRADE 6, HYPOTHESIS TESTING, METHODOLOGY, ACHIEVEMENT TESTS,

IN A STUDY OF GRADING METHODS PRACTICED BY TEACHERS OF 103 SIXTH GRADE CLASSES, IT WAS FOUND THAT USE OF THE I.Q. SYSTEM AS A BASIS FOR ACHIEVEMENT GRADING PRODUCES APPROXIMATELY THE SAME RESULTS AS PERMITTING TEACHERS TO USE THEIR OWN JUDGMENT IN ASSIGNING ACHIEVEMENT GRADES. TWO HYPOTHESES WERE TESTED--(1) THE AVERAGE I.Q. LETTER GRADE OF A CLASS PROVIDES AN APPROPRIATE MIDPOINT FOR ACHIEVEMENT GRADING, AND (2) THE I.Q. LETTER GRADING SYSTEM PROVIDES SUITABLE HELP TO TEACHERS IN DETERMINING ACHIEVEMENT LETTER GRADES. COMPARISONS OF TEST LETTER GRADES WITH THE I.Q. MEANS FOR EACH OF THE CLASSES STUDIED REVEALED A CLOSER AGREEMENT BETWEEN I.Q. MEAN AND READING ACHIEVEMENT MEAN THAN BETWEEN I.Q. MEAN AND LANGUAGE, ARITHMETIC, AND SCIENCE ACHIEVEMENT MEANS. USE OF THE I.Q. LETTER GRADING SYSTEM TO DETERMINE THE DISTRIBUTION OF ACHIEVEMENT GRADES COULD NOT BE JUSTIFIED FOR ONE-THIRD OF THE CLASSES. NO SIGNIFICANT DIFFERENCES WERE FOUND IN A COMPARATIVE ANALYSIS OF THE TWO GRADING METHODS STUDIED--(1) RELIANCE LARGELY UPON THE TEACHER'S JUDGMENT, AND (2) USE OF I.Q. DISTRIBUTION AS A BASIS FOR ACHIEVEMENT GRADING. (JK)

An Investigation of Achievement Grading Based on Scholastic Ability Distribution

This is a paper on the age old problem of assigning fair or valid letter grades for school achievement. The study was done over a two year period in one school district in the Province of British Columbia. It is therefore questionable whether the findings can be generalized to any extent. The interest, then, should be in the approach rather than the specific results.

The questions to be investigated can be stated quite simple as

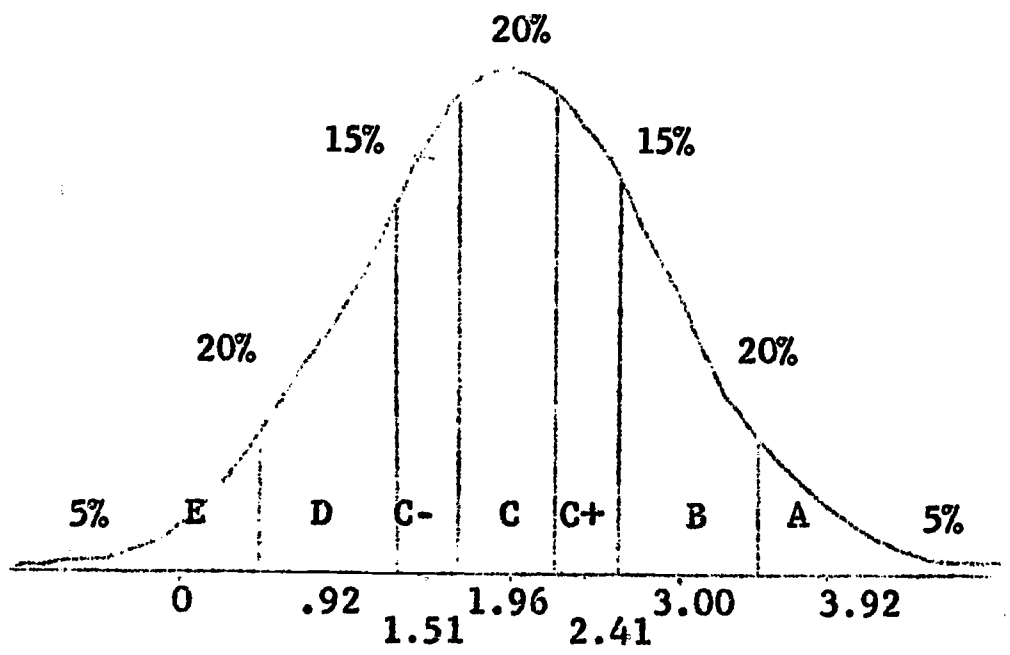
1. How can one assure some comparability between grades given by different teachers or by different schools?

2. Can a system be introduced which will help teachers grade more validly?

In 1939 Ross in his Measurement in Today's Schools suggested that the I.Q. distribution within a single class should provide a satisfactory basis for awarding letter grades for achievement in various school subjects. This suggestion was adopted enthusiastically by many schools in British Columbia and has been used widely and uncritically for the last twenty years.

The procedure used by British Columbia schools is a simple one. First, the I.Q.'s of all the children of a given grade in the school district are collected and norms established. The I.Q.'s are then letter graded so that the top 5% are called "A" I.Q.'s, the next 20% "B" I.Q.'s, the next 15% "C+" I.Q.'s and so on, following the percentage breakdown shown in Figure 1.

Figure 1



The I.Q.'s of the children in a given class are then converted to letter grades using the district norms, and a distribution is made of these letter grades. For example, it may be found that in a given class there are 2 "A" I.Q.'s, 6 "B" I.Q.'s, 5 "C+" I.Q.'s and so on.

In grading for, say, arithmetic achievement the teacher now uses the same letter grade distribution. In this example, the top two achievers in arithmetic would receive A grades, the next 6 B grades, the next 5 C+ grades, and so on.

In an earlier study I pointed up the large disparity in any given class between the distribution of the I.Q. letter grades and the distribution of the achievement letter grades obtained from district norms.

The results of this study were countered by the argument that while the distribution of the letter grades for I.Q. and for achievement may not be in close correspondence, nevertheless the use of the I.Q. system at least ensures that the achievement grades are grouped around a mid-point that is appropriate for the class. Secondly, because I.Q. results are readily available in most school districts it is held that this system is therefore of great help to teachers. It is with the validity of these statements that this study is concerned.

The study is divided into two parts. The first attempts to examine the hypothesis that the average I.Q. letter grade of a class provides an appropriate mid-point for achievement grading. The second, the hypothesis that the I.Q. letter grading system provides suitable help to teachers in awarding achievement letter grades.

The first question then is "To what extent does the mean of the I.Q. letter grades correspond with the mean of the achievement letter grades of a class, if both are based on sets of norms obtained from all the children of a given grade in a school district?"

To answer this question the results of 103 classes on 4 tests of achievement were examined. These classes formed the grade 6 population of one school district over a period of two years.

The achievement tests in language, reading, arithmetic computation and science were constructed by the staff of the school district superintendent. The I.Q.'s were obtained from the Otis Self-Administering Test of Mental Ability.

Thus each class was given a battery of tests which were letter graded on the basis of the results of all the children in Grade 6 of the school district. It was now possible to take the mean of the I.Q.'s for each class and compare it with the mean of the achievement in each subject.

To do this a value was given to the mid-point of each letter grade interval according to its position in the normal curve. The zero point was then set at the mid-point of E. The successive z values for mid-points of the letter grades are shown under the base line of the curve in Figure 1.

Results

The results are given in Table 1. The left hand column gives the step intervals showing the size of the difference between the mean of the I.Q. letter grades and the mean of the achievement letter grades for a class. The columns headed Language, Reading, Arithmetic Computation and Science show the frequency with which differences of various magnitudes occurred.

Table 1

Frequency of Various Deviations from the Scholastic Aptitude Letter Grade Mean of the Letter Grade Means in Language, Reading, Arithmetic Computation and Science for 103 Classes

Size of Deviation	Language	Reading	Arithmetic Computation	Science
0 - .10	33	45	23	16
.11 - .20	23	31	31	33
.21 - .30	20	19	23	22
.31 - .40	4	6	4	13
.41 - .50	4		12	6
.51 - .60	17		2	2
.61 - .70	2	2	4	2
.71 - and over	0		4	9
	103	103	103	103

As can be seen from the table there is a much closer agreement between the mean for I.Q. and the mean for reading achievement than for the other subjects. If one were to arbitrarily take a deviation of up to .30 (or very roughly half a letter grade in the B to D range) as an acceptable difference for practical purposes between the achievement and I.Q. means for a class, then 8 of the 103 classes fell outside this range for reading, whereas 27, 26 and 32 classes fell outside for language, arithmetic and science respectively. In other words, whereas practically all fell within a deviation of .30 for reading only 2/3 to 3/4 fell within this range for the other subjects. Therefore, one might argue that the mean of the I.Q. is reasonably valid as a mid-point for reading achievement but suspect for grades in the other three subjects. The justification of the use of the I.Q. letter grade distribution as providing an appropriate mid-point cannot be maintained for 1/3 of the classes. However, the question still remains whether the I.Q. system is better than nothing.

The second hypothesis that the I.Q. letter grade system provides suitable help to teachers in awarding achievement letter grades was now examined. The question to be answered is whether teachers can grade more validly if freed from the restraints imposed by the I.Q. letter grade system.

In the month of February objective tests constructed by the staff of the school district superintendent in arithmetic computation and reading were given to all grade 6 classes. Each of the 41 teachers who had at least 2 years experience teaching the grade was asked to administer the tests to his class, mark the tests, then make an estimate of the appropriate letter grade to be awarded to each paper. A test of scholastic ability was also given at this time and returned to the central office for marking.

The tests from all the grade 6 classes of the district were now collected and norms established in the same way as described in the first part of the study.

In other words each raw score was turned into a letter grade based on the grade 6 population of the district. These letter grades provided the criterion against which to judge the relative efficacy of the two methods of awarding grades which were under scrutiny.

The two methods were:

A. To rely largely on the teacher's judgment;

B. To use the I.Q. distribution as a basis for the achievement grading.

Method 1, which is designated the teacher's judgment method, was as follows:

1. The teacher marked the papers of his class - these were objective tests.
2. He was given the mean I.Q. letter grade value of his class. This was all the information supplied to him.
3. He decided, from what he knew of the work of the class and the I.Q. mean, on an appropriate average around which to base his achievement grading.
4. He allocated grades taking into account the gaps in the raw score distribution and the average he desired to maintain.

Method 2 involved the application of the I.Q. letter grade distribution of the class to the raw scores for achievement.

This was done by the experimenter on the basis of the I.Q. letter grades obtained from the central office.

Thus two sets of letter grades for achievement were available for each of the 41 classes: one based on the teacher's judgment, the other on the I.Q. distribution in the class.

A criterion against which to judge each set of grades also existed as each raw score had been letter graded in the central office on the basis of norms established for all the grade 6 children in the district.

In order to compare each of the two methods with the criterion, deviation scores were computed in the following way:

1. The z values previously assigned to the letter grades were rounded off so that 3.92 became 4, 2.41 became 2.5 and so on. The values are shown at the bottom of Figure 2.

Figure 2

Class 17												
Criterion	A	A	A	B	B	B	B	C+	C+	C+	C+	
Grade by I.Q.	A	B	B	B	B	B	B	B	B	C+	C+	etc.
Deviation Score		1	1					½	½			
Criterion	A	A	A	B	B	B	B	C+	C+	C+	C+	
Teachers' Judgment	A	A	B	B	B	C+	C+	C+	C+	C	C	etc.
Deviation Score			1			½	½			½	½	

A = 4 B = 3 C+ = 2½ C = 2 C- = 1½ D = 1 E = 0

2. Each student's grade by the teacher's judgment method was compared with the criterion (his grade on the district norms) and a deviation score calculated by taking the absolute value of the difference between the letter grade values. The deviation score for each student was summed for the whole class. Figure 2 portrays the method of arriving at the deviation scores.

3. The procedure was repeated for the letter grades obtained by the I.Q. distribution method. In other words, the I.Q. distribution of letter grades was applied to the raw scores and these grades were compared with what would have been obtained by the use of the district norms.

Results

The results were as follows: The mean deviation score per class for reading using the teacher's judgment was 7.6 ($s = 4.5$) compared with 8.2 from applying the I.Q. distribution system. For arithmetic the respective means were 6.8 against 7.4. Neither difference between the means is significant.

The teachers were also considered individually. When using their own judgment for the reading grades, of the 41 teachers 19 were superior and 15 inferior to the application of the grade by I.Q. system in the ability to predict the criterion grades. For arithmetic the figures were 21 superior and 17 inferior. The difference in favour of the teacher's judgment is not significant. These results are shown in Table 2.

Table 2

Number of Teachers Demonstrating Superior Judgment
by Use of Own Judgment and by I.Q. System

	Reading	Arithmetic
Own Judgment Superior	19	21
I.Q. System Superior	15	17
No difference	7	3
	41	41

When the teachers were divided into groups designated "Superior" and "Inferior Judgment" there was no significant difference in the mean amount of teaching experience of the groups. Neither were there sex differences in the ability to judge achievement.

In other words grading by the use of the I.Q. system produced approximately the same results as permitting the teachers the use of their own judgment. The factors which enabled approximately half the teachers to do better and the other half worse when freed from the restrictions of the I.Q. letter grade distribution have not yet been identified.

Geoffrey Mason
University of Victoria

February, 1967