

1954

Increasing the Objectivity of Final Marks in Multiple-Section Courses

Charles O. Neidt
University of Nebraska

Copyright ©1954 Iowa Academy of Science, Inc.

Follow this and additional works at: <https://scholarworks.uni.edu/pias>

Recommended Citation

Neidt, Charles O. (1954) "Increasing the Objectivity of Final Marks in Multiple-Section Courses," *Proceedings of the Iowa Academy of Science*, 61(1), 394-401.
Available at: <https://scholarworks.uni.edu/pias/vol61/iss1/51>

This Research is brought to you for free and open access by the Iowa Academy of Science at UNI ScholarWorks. It has been accepted for inclusion in Proceedings of the Iowa Academy of Science by an authorized editor of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.

Increasing the Objectivity of Final Marks in Multiple-Section Courses

BY CHARLES O. NEIDT¹

During the past forty years the reliability of course marks assigned by teachers has received considerable research attention and much speculation from investigators. A review of the literature relating to this topic reveals that researchers have long recognized the importance of evaluating this characteristic of course marks, but have been restricted in their research endeavor by lack of appropriate methodological techniques. Chauncey and Fredriksen (1, p 97) have expressed both the importance of one aspect of the general problem and the need for more adequate methodology in the following statement:

"In any serious research program aimed at improving the prediction of academic success in college, another important first step is to determine the adequacy of the criteria of academic success. What is the reliability of a course grade? What is the reliability of an average of course grades? Which courses tend to be reliably graded and which unreliably graded? . . . To none of these questions have the experts in measurement yet found satisfactory answers . . . Too little effort has been made to gather the kind of data which would make possible a thorough-going analysis of the situation in any given situation. An important contribution can be made by the expert in measurement who begins his evaluation of admission procedures by first determining the essential characteristics of the criteria they attempt to predict."

The attenuating effect of unreliability in the criterion is an important consideration in constructing test batteries for the prediction of college marks. Even the most effective preregistration test batteries seldom account for more than half of the total variation in average course marks. Just how much of the remaining variance unaccounted for by present tests is attributable to error and how much is attributable to student behavior untapped by the test battery cannot be answered at the present time.

Odell (2, p. 713) has indicated the paucity of research evidence relating to this topic in the following statement:

"No significant data exist which would indicate directly the

¹The author is indebted to the following instructors who participated in this study: Dr. Louis W. Kazienko, Mankato State Teachers College; Dr. John MacRae, University of Omaha; Dr. Jack H. Monderer, Vocational Counseling Service Inc., New Haven Connecticut; Mr. William P. McDougall, Caldwell Public Schools, Caldwell, Idaho.

reliability of the mark based on a whole term or semester's work.

Such evidence as is available, from studies of marks given by the same teacher to the same individual in successive courses, from studies between marks actually received and those students think they should receive, and from other sources seem to justify the conclusion that the usual reliability of semester marks is indicated by coefficients of correlation of .70 to .80 or even as high as .80 to .90."

Other than the attenuating effect of unreliability on predictive effectiveness discussed by Chauncey and Fredriksen, there is another aspect of the reliability of course marks which is important to departments offering several sections of the same course which are taught and evaluated by different instructors. This aspect of the problem of course mark reliability, usually referred to as objectivity, is the extent of agreement among instructors in assigning final marks to students in the various sections. It should be recognized that objectivity, as here used, is only one aspect of reliability, but is a contributing factor to reliability. Reliability may be defined as the extent to which observations are free from compensating errors. One type of compensating error is that error arising from an evaluation of the same behavior or performance by different judges or observers. It is this freedom from variation among the observations of several judges which is usually termed objectivity. If the reliability of course marks is to be adequately assessed, such an assessment must include an evaluation of objectivity.

It is the purpose of this paper to describe a method for assessing objectivity of final course marks, to present some evidence resulting from the use of the method and to identify tentatively some factors related to objectivity. It is clearly recognized that the method and the evidence to be presented do not answer the imposing questions raised by Chauncey and Fredriksen, but it is felt that the implications will be of value in the search for evidence regarding the reliability of course marks at the college level.

STATEMENT OF THE PROBLEM

It was the purpose of the study reported here to determine the extent of agreement among the final course marks assigned by five instructors to 31 students enrolled in one section of an undergraduate educational psychology course at the University of Nebraska.

DESCRIPTION OF EXPERIMENTAL CONDITIONS

The course utilized for this investigation was a first-semester

course in educational psychology. This course represents the first half of a two-semester sequence of courses ordinarily taken during the second semester of the freshman year and the first semester of the sophomore year. In general, the content of the first course includes units involving the study of motivation; emotional, social and mental development; and mental hygiene. The second-semester course is concerned primarily with the study of learning and the measurement of aptitude and achievement. At the University of Nebraska there are from twelve to fourteen sections of these two courses offered each semester. Enrollment in the various sections is customarily kept at thirty-five, but occasionally circumstances have forced the sections to vary from 15 to 60.

Each class meets three times a week with the same instructor. The sections are usually conducted on a combination lecture-discussion basis. All tests but one half of the final examination are constructed by the individual instructors.

The students involved in this investigation were 31 freshmen and sophomores who had been assigned to a one o'clock section of educational psychology when they registered for the course. Assignment to this section depended upon a particular student's schedule and the time at which he registered. Instructors for the sections were not announced until registration was complete.

The instructors who assigned marks to the class of 31 students were members of the staff of the Department of Educational Psychology and Measurements at the University of Nebraska. One instructor was a senior staff member and four were advanced graduate students. Three of the graduate students held half-time instructorships and the fourth was a full-time instructor. Three were within one year of completing their Ph.D. degrees and one was a candidate for the M.A. degree. Three of the graduate student instructors had previously taught this same course at least once, and all of them were teaching one section of the same course themselves during the semester of the investigation.

All five instructors were participating in a staff seminar on the teaching of educational psychology at the time of this investigation. The seminar met once a week for one and one half hours. Activity of the seminar consisted of outlining the course content, constructing the departmental section of the final examination, constructing a syllabus for student use and discussing methods of teaching the two courses. The seminar was under the direction of a senior staff member and attendance of all persons teaching either of the two previously-described courses was required.

The system of reporting final course marks at the University of Nebraska involves a nine-point numerical scale ranging from one to nine. A value of one connotes failure and nine is the highest possible mark. Although the system is patterned after the stanine score method of reporting test performance, no attempts are made to force the percentage of assigned marks to correspond to the normal curve.

PROCEDURE

Five instructors attended every session of the class in beginning educational psychology held throughout the semester. Although the class was taught entirely by one instructor, the other four had seating charts and kept whatever record of class attendance and discussion they desired. At the start of the semester it was announced to the class that frequently observers would be sitting in the back row; that these observers were also teachers of other sections of the same course; and that the class would be conducted as if the observers were not present.

Before the investigation, the instructors agreed not to discuss the performance of the students among themselves and to evaluate this class in a way as nearly comparable as possible to that which they would use if they were teaching the class. Each instructor made his evaluation independently and in no instance did the instructors compare results until after the investigation was completed.

Five twenty-minute quizzes, two one-hour examinations, twenty abstracts of library references, a case study, and the scores from a two-hour final examination were evaluated by the five instructors. Each instructor combined the evidence from these sources, plus that from attendance and discussion, if he had recorded such evidence, according to any system he desired.

The test items were selected from a pool of items covering the entire course. When a quiz or examination was to be given, each instructor independently selected the items he wanted administered to the class and made his indications to a secretary who recorded the items selected by any given instructor. Many of the items chosen by the instructors overlapped, but each test contained some items which were unique to one or more instructors. In using essay type items, each unique item was mimeographed on a half sheet of paper. Following the administration of the tests, the secretary scored those objective items designated by each instructor and returned the distribution based on his items to him, and distributed the responses to the essay-type items in the same way.

Thus each instructor did not know how the students responded to any items other than those which he had designated. In the case of responses to the essay-type items used by more than one instructor, no marks were made by the instructors on the papers which the other instructors were also to evaluate.

At the end of the semester, each instructor turned into the secretary the distribution of final course marks which he had prepared for the class. Obviously, the distribution of final marks assigned by the instructor who had taught the class was the one officially reported to the registrar. This distribution was filed before any comparisons were made.

Table 1
Final Marks Assigned by Five Instructors to 31 Students in a
Class of Educational Psychology

Student No.	Instructor				
	I	II	III	IV	V
1	8	9	8	8	8
2	4	5	4	4	3
3	5	5	5	5	5
4	6	6	6	6	5
5	5	5	5	4	4
6	8	9	8	8	8
7	8	8	9	8	9
8	6	6	6	7	5
9	7	7	7	7	6
10	6	6	6	6	5
11	5	5	5	5	5
12	4	4	4	3	4
13	4	4	4	3	4
14	5	4	5	5	4
15	4	4	4	4	3
16	6	6	6	7	5
17	7	6	7	7	6
18	8	7	8	7	7
19	8	8	8	9	7
20	2	2	2	3	2
21	2	2	2	2	2
22	7	7	8	8	6
23	5	5	5	5	5
24	4	4	5	4	4
25	7	6	7	6	5
26	5	6	5	5	5
27	1	1	1	1	1
28	6	6	5	7	5
29	6	6	6	6	5
30	3	3	2	2	2
31	8	8	9	8	7

RESULTS

In Table 1 are shown the distributions of final course marks assigned by the five instructors to the 31 students. The similarity of the marks assigned by the instructors is striking. In five instances, the instructors all agreed as to the mark assigned. For 13 students there was agreement among four of the instructors and the fifth deviated by no more than one unit. In no case was there a deviation of more than two units among the five instructors. The degree of similarity among the course marks is shown by the coefficients of correlation in Table 2.

Table 2

Intercorrelations Between Marks Assigned by Five Instructors to 31 Students

Instructor	Instructor				
	I	II	III	IV	V
I		.97	.99	.95	.95
II			.95	.94	.95
III				.94	.91
IV					.91
V					

In Table 3 is shown the analysis of variance of the five distributions of marks. Although the correlations between the pairs of distributions reported in Table 2 are very high, the F-value for instructors indicates that the distributions vary significantly.

Table 3

Analysis of Variance of the Marks Assigned by Five Instructors to 31 Students

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Instructors	4	8.93	2.23	10.14**
Individuals	30	569.34	18.98	86.27**
Unexplained	120	26.27	.22	
Total	154	604.54		

To identify more accurately the distribution or distributions of marks accounting for the significant F-value, the inexperienced instructor's marks (Instructor V) were kept separate and the distributions assigned by the four experienced instructors were pooled. The data classified according to this basis were analyzed and the F-value of 37.96 shown in Table 4 was found. The inexperienced instructor's marks were significantly lower than the marks assigned by the experienced instructors.

Table 4

Analysis of Variance Between the Inexperienced Instructor's Marks and the Pooled Marks of the Four Experienced Instructor's

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Instructors	1	8.73	8.73	37.96**
Individuals	30	569.34	18.98	86.27**
Unexplained	123	26.47	.23	
Total	154	604.54		

When the inexperienced instructor's marks were disregarded, and the four distributions of marks assigned by the experienced instructors were analyzed, no significant difference was found among the instructors. The results of this analysis are shown in Table 5. These results should be considered only as tentative, however, since part of the data were ignored in making the analysis.

Table 5

Analysis of Variance of the Marks Assigned by Four Experienced Teachers to 31 Students

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Instructors	3	.10	.03	.06
Individuals	30	448.00	14.93	31.10**
Unexplained	90	42.90	.48	
Total	123	491.00		

DISCUSSION

It was previously pointed out that objectivity is only one aspect of the reliability of course marks, hence the coefficients of correlation reported in Table 2 should not be construed as reliability coefficients. Rather the interpretation should be that under the conditions prevailing in this investigation, it is possible to achieve close agreement among marks assigned by instructors. Certainly a repetition of the investigation under other circumstances would be desirable.

If another such study were made, it would be of value to ascertain the abjectivity of the different types of evidence upon which the final course marks are based. The results of the present investigation were restricted to a single final course mark.

Another consideration for further use of the method is the size of the sample of the instructors. Inferences made on the basis of

five instructors are limited. Desirable as it would have been to have involved more instructors in the present study, however, a larger number who could contribute the extensive time and effort required for this type of investigation was unavailable.

Careful examination of the conditions under which this experiment was performed and of the results suggests certain inferences regarding factors relating to the objectivity of final course marks. These inferences are offered as tentative and further research is advocated for their verification.

1. Evidence from this investigation supports the hypothesis that homogeneous instructors tend to be more similar in the final marks assigned to students than less homogeneous instructors when they are classified according to teaching experience.

2. Frequent tests contribute toward greater objectivity of course marks assigned by instructors.

3. Frequent discussion of the course content and the teaching methodology appropriate for the course contribute toward greater objectivity.

4. It is possible to achieve a high degree of objectivity of course marks even though individual instructors are free to construct their own evaluation devices and weigh evidence for the final course mark in the manner most satisfactory to them.

SUMMARY

Five instructors who had attended each class session of a course in educational psychology independently assigned final marks to the 31 students in the class. Analysis of the marks indicated close correspondence among the marks assigned by the instructors, although some variation attributable to different amounts of teaching experience was found.

References Cited

1. Lindquist, E. F., (ed) *Educational Measurement*, American Council on Education, Washington, D. C., 1951.
2. Monroe, Walter S. (ed) *Encyclopedia of Educational Research*, Rev. Ed., The MacMillan Company, New York, 1950.

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY AND MEASUREMENTS,
UNIVERSITY OF NEBRASKA
LINCOLN 8, NEBRASKA