DOCUMENT RESUME

ED 129 754 SP 010 469

AUTHOR Wagener, Elaine H.
TITLE Use of Student-Composed Tests and Their Effect on the Attitudes and Task Performance of University Students.
PUB DATE [76]
NOTE 12p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Rating; Cognitive Measurement; Cognitive Processes; *College Students; Evaluation; *Grading; Higher Education; Learning Processes; Seminars; Statistical Analysis; *Student Attitudes; *Student Seminars; Task Performance; *Test Construction; Tests

ABSTRACT

The experiment described was designed to evaluate the possible effects on students in a small seminar of an evaluation system in which students were freed from some of the pressures of the conventional grading system and allowed to participate in the grading process, and to determine whether such participation would affect the acquisition of course content and attitudes of students toward that course. One of three randomly selected seminar groups was chosen to be the experimental group in which the grading process was altered. Pretesting was done covering course content to establish relative equality among the groups. An analysis of variance revealed no statistically significant difference in the group that constructed its own tests and assigned its own points for seminar participation and the two control groups, which were graded by a more traditional method. Chi Square analyses were performed on the attitude data revealing no significant differences in the groups on attitudes toward unit quizzes, course grading, or small group seminars. Although statistical analysis indicated no significant differences in attitudes in the groups, the study revealed that students felt the experience was valuable in learning to compose valid test questions, that it freed them from memorizing irrelevant details, released them from tension, and allowed a more receptive mindset for hearing and listening as well as allowing for increased teacher-student communication, rapport, and appreciation. (Author/JMF)

USE OF STUDENT-COMPOSED TESTS AND THEIR EFFECT ON THE ATTITUDES AND TASK PERFORMANCE OF UNIVERSITY STUDENTS

Elaine H. Wagener
Assistant Professor
Curriculum and Instruction
The University of Texas
Austin, Texas

USE OF STUDENT-COMPOSED TESTS AND THEIR EFFECT ON THE ATTITUDES AND TASK PERFORMANCE OF UNIVERSITY STUDENTS

Grading practices have remained at the status quo level in spite of many innovations in education. Teachers face many questions within themselves as to the validity or appropriateness of current grading procedures. Consequently, some new possibilities are being studied. In a report by Collins and Nickle (1) five hundred forty-four institutions of higher education were surveyed. Institutions replying seemed to be experimenting with different grading systems, using traditional forms within the student's major area and using non-traditional grading systems outside the major area of study. Many research studies seem to indicate that grading has an anxiety producing effect on students or in some way affected motivation which in many cases negatively affected task performance (2, 3, 4, 5). In a survey of university students, 65.8 percent felt that grades interfered with learning because of undue pressure on grades (6). This negative effect on the development of intrinsic and permanent intellectual interests and the failure of grades to motivate learning are emphasized in a report of evaluation practices at the University of California (7).

Several studies reported no significant differences in either attitude or performance when variations in grading practices were used--criterion referenced grading (8), contracting (9), or indiscriminately raising the grades of every student by one level in an experimental group of high school Spanish students (10). Other studies revealed positive results. Emerson (11) reported positive student attitudes toward a social science course and the grading method when a system of points

was used in evaluating objective tests; textbook assignments; and reviews of tapes, movies and articles. In a study by Good (12) of one hundred forty-seven high school biology students there was no significant difference at the .05 level in achievement, but self-graded students were more productive and had a significantly different level of aspiration. In response to this problem and with the desire to implement a more productive approach to grading, the study described in this paper was undertaken.

DESCRIPTION OF THE STUDY

This experiment was designed to evaluate the possible effects on students in a small seminar of an evaluation system in which students were freed from some of the pressures of the conventional grading system and allowed to participate in the grading process.

The students were primarily sophomores who were education majors in a southern university. The course was a required survey education course with an introduction to educational psychology, learning theory, educational legislation, local school politics, teacher organizations and unions, and current educational issues. The pretest and posttest which was given at one time to the fifteen seminar sections was fifty multiple-choice items based on the content of the above items. Information concerning these topics was placed on cassette tapes in an independent study lab. Bi-weekly quizzes which were also objective multiple-choice type questions were composed jointly by the five seminar leaders and given to all the students in the seminars. The bi-weekly quizzes covered information given on assigned tapes. Seminars were designed to provide discussion of questions which arose out of the tapes and to review questions from quizzes of the previous week.

Three of the fifteen seminar groups were chosen to participate in the experiment. The three groups were all conducted by the same seminar leader. The students were assigned to the groups randomly by computer. One of the three groups was assigned randomly to the experimental treatment.

It was hypothesized that the experimental students' attitudes toward the seminars and the testing process would be more positive if they participated in the construction of the bi-weekly quizzes and graded their own participation in the seminars. It was also hypothesized that the scores of experimental subjects on a pretest and posttest which evaluated knowledge of material available on cassette tapes would compare favorably with the scores of subjects in the control groups. Acceptance of the hypothesis would seem to indicate that the removal of grading incentives did not adversely affect task performance.

The two control groups were subjected to a grading system in which tests were taken bi-weekly and two points were given for each test passed at a mastery level of eighty percent or above. The quizzes covered content presented in tapes available in the independent study lab and assigned weekly. They were constructed by the staff of seminar leaders and the professor in charge of the project. Quizzes could be taken again for half credit if they were not passed satisfactorily the first time.

The experimental group listened to the tapes in the listening lab at whatever time they preferred. The first quiz was given in the same manner as in the control groups. However, the students were told that on the tapes that followed they would write their own questions which would be answered by all seminar members. It was explained that this was an attempt to free them from listening to the tapes "in order to answer questions on a quiz." They were told that this process would perhaps enable them to listen to assimilate information which might be of value

to them. Each student was to listen to all four or five tapes but signed up to write two questions on one specific tape. Each student was to use the questions on the first quiz as a model for his or her questions. They were instructed to try not to ask trick questions but to compose questions which were related to what they considered to be the most important information on the tape.

Each question written by a student was to be answered by all other students in the seminar. After the evaluation period, the questions were to be discussed both to clarify content and to consider problems involved in writing appropriate questions which would be clear and unambiguous.

Each student who came prepared with his or her questions and participated in the evaluation and discussion was given the right to evaluate his own participation in the entire process and assign himself up to six points for the three quizzes. This part of the design was included to relieve students of the pressure of answering questions which might be unclear or poorly constructed and place the emphasis on the content rather than on grades.

If a student was absent or came to class unprepared, he would automatically take the standard quiz used in the other seminars and be graded on the usual basis.

Students in all three seminars were given a 50-item multiple-choice pretest based on questions similar to those used on the four quizzes. They were given this same test again at the last seminar. A comprehensive evaluation form was completed by all students taking this course and this evaluation form was completed by all students taking this course and this evaluation was used as a basis for comparing the attitudes of the experimental group with the control groups. In addition, each student

in the experimental group was asked to make comments on the experimental process.

RESULTS AND DISCUSSION

The Subjects/Groups X test analysis of variance indicates that there is a significant pretest vs. posttest main effect ($p < .01$) as might be predicted in most courses of study. However, the Between Groups main effect and the Groups by Trials interactions were not significant (Table 1) indicating that the student participation in the preparation of the bi-weekly quizzes and self-evaluation of their participation in weekly seminars neither affected their performance negatively or positively. These findings indicated that this process was neither superior nor inferior to traditional grading methods. Content was retained equally well by students who were studying for a test and students who were not operatin, nder the test-pressure syndrome.

TABLE 1

Ss/Gps X Tests ANOVA

Source of Variations	df	SS	MS	F
Between Ss	38	838.48	---	
Between Gps $\bar{X}$	2	30.33	15.16	< 1
Ss/Gps	36	808.15	20.72	
Within Ss	39	2736.50	---	
Overall Tests	1	2272.32	2272.32	196.28*
Groups x Tests	2	47.41	23.70	2.04
(SS x Tests)/Gps	36	416.76	11.57	
Total	77	3574.98	---	

*$p < .01$

The final attitude evaluation form contained eighteen items measured on a five point Likert scale ranging from Unsatisfactory (1) through Adequate (3) to Very Satisfactory (5). Since this experiment was designed to measure difference in attitude toward grading, testing, and seminar sessions, these will be the only comparisons reported.

It was hypothesized that the attitudes of the experimental group would be significantly higher than the control groups. Chi Square analyses were performed to ascertain if the experimental treatment did indeed affect the attitude of students concerning unit quizzes, course grading, or small groups seminars (Tables 2, 3, and 4).

TABLE 2

Unit Quizzes

	1	2	3	4	5	Ss
Experimental Group	.5 .67	.5 1	3.5 3.33	3.5 4	4.5 4	13
$Control_1$	.5 .67	1.1 1	3.5 3.33	4 4	4 4	13
$Control_2$	.5 .67	1 1	3.5 3.53	4.5 4	3.5 4	13
	2	3	10	12	12	

$X^2 = .6554$ ^8df

TABLE 3

Course Grading

	1	2	3	4	5	Ss
Experimental Group	.5 .33	.5 .67	1.5 1.67	3.5 3.33	7 7	13
$Control_1$	.5 .33	.5 .67	1.5 1.67	3.5 3.33	7.5 7	13
$Control_2$	.5 .33	.5 .67	1.5 1.67	3.5 3.33	7.5 7	13
	1	2	5	10	21	

$X^2_{8df} = .3888$

TABLE 4

	1	2	3	4	5	Ss
Experimental Group	0 0	.5 .33	1.5 2.67	4.5 4	6.5 6	13
$Control_1$	0 0	.5 .33	3.5 2.67	2.5 4	6.5 6	13
$Control_2$	0 0	.5 .33	2.5 2.67	4.5 4	4.5 6	13
	0	1	8	12	18	

$X^2_{8df} = 2.2180$

all X^2 are not significant

However, the X^2 analyses indicated no significant difference at the .05 level between groups in their attitudes on these three items. (It should be noted that the Yates correction for continuity was used due to the small expected frequencies). The hypothesis concerning student attitudes was therefore rejected.

CONCLUSIONS

Although statistical analysis indicated no significant differences in attitude[s] of the three groups, student comments concerning the experimental procedure revealed their perceptions about the advantages of this method. They felt it was valuable in learning to compose valid test questions, freed them from memorizing irrelevant details, released them from tension and allowed a more receptive mindset for hearing the tapes. Another benefit received was increased teacher-student communication, rapport and appreciation.

Long before John Holt wrote How Children Fail the stance regarding grading as a motivational device has been questioned. Teacher training institutions have been plagued with teaching one philosophy with regard to the grading process, particularly in elementary education, and have continued to incorporate procedures which contradict that instruction. The idea that without grades students would fail to learn or would cease expending maximum effort and time in preparation has been an overriding factor in the decision to retain grading.

This study would seem to indicate that students learn as much when emphasis is placed on content other than grades even though their attitudes toward the learning process are not significantly affected by participation in the grading process. Perhaps John Holt's position would bear a more careful consideration.

"...any tests that are not a personal matter between the learner and someone helping him learn but were given instead to grade and label for someone else's purposes...are illegitimate and harmful" (9).

REFERENCES

1 Collins, Janet R. and K. N. Nickel. "A Study of Grading Practices in Institutions of Higher Education," Resources in Education, 10 (March, 1975), p. 62 (ED097-846).

2 Briggs, Francis M. "Grades--Tool or Tyrant? Research on High School Grading," Journal of Educational Research, 52 (April, 1964), pp. 280-284.

3 DeRidder, L. M., and T. L. Hawk. "A Comparison of the Performance of Pre-Graded Students with Grade-Motivated Students," Journal of Educational Research, 56:10 (July - August, 1963), pp. 548-550.

4 Bain, Philip T., Loyde W. Hales and Leonard P. Rand. "An Investigation of Some Assumptions and Characteristics of the Pass-Fail Grading System," The Journal of Educational Research, 67 (November, 1973), pp. 134-136.

5 Thayer, Robert E. "Do Low Grades Cause College Students to Give Up?" Journal of Experimental Education, 41 (Spring, 1973), pp. 71-73.

6 Burke, Ronald J. "Student Reactions to Course Grades," The Journal of Experimental Education, 36:4 (Summer, 1968), pp. 11-13.

7 Miller, Stuart. "Report on Methods of Evaluating Students at the University of California, Berkeley," Resources in Education, 8 (September, 1973), p. 126, ED076-631.

8 Williams, Reed G., Martin J. Pollack and Nancy A. Ferguson. "Differential Effects of Two Grading Systems on Student Performance," Journal of Educational Psychology, 67 (April, 1975), pp. 253-258.

9 Yarber, William L. "A Comparison of the Relationship of the Grade Contract and Traditional Grading Methods to Changes in Knowledge and Attitude," The Journal of School Health, 44 (September, 1974) pp. 395-398.

10 Marshall, Jon C. and Diana L. Christensen. "Leniency in Marking: It's Effects on Student Motivation and Achievement," Education, 93 (February, 1973), pp. 262-265.

11 Emerson, George H. "Revised System of Evaluation," Resources in Education, 11 (January, 1976), p. 126, ED111 465.

12 Good, Wallace Martin, "Self-Graded and Teacher Graded Achievement in a BSCS High School Biology Course," Resources in Education, 8 (March, 1973), p. 93, ED069-494.

13 Holt, John. What Do I Do Monday? New York: E. P. Dutton and Company, 1970, p. 264.