DOCUMENT RESUME

ED 084 156 SE 017 001

AUTHOR Townsend, Neal R.; Wheatley, Grayson H.
TITLE Frequency of Tests and Feedback of Test Results in Calculus Classes.
PUB DATE [73]
NOTE 10p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement; Calculus; *College Mathematics; Feedback; Instruction; *Mathematics Education; *Research; *Teaching Techniques; Testing
IDENTIFIERS Research Reports

ABSTRACT

A three-factor design was used to determine the effects of testing frequencies and feedback delays on college students' achievement in a beginning calculus course. Four test-period frequencies--daily quizzes (5-10 minutes), weekly quizzes (20-30 minutes), three midterm exams, or one midterm exam--were used. Two feedback-delay-levels for test returns were set at next class meeting or three-day delay on quizzes, and one week for exams. Subjects were blocked on ability level (SAT scores) yielding the third dimension of the design. A constructed achievement test with .78 reliability was used as a criterion measure; the test was multiple-choice. Data from an attitude measure and a chi-square test for differences in dropout proportions are reported as well as the results of the three-factor analysis of covariance. In general, classes given short daily quizzes had higher achievement scores and classes with the delay in feedback were significantly better than the class receiving results the next meeting. No attitude differences were found nor any interaction due to aptitude level. (JP)

Neal R. Townsend
Grayson H. Wheatley

FREQUENCY OF TESTS AND FEEDBACK OF TEST RESULTS IN CALCULUS CLASSES

Since classroom tests may be used not only to evaluate achievement, but also to teach by communicating course objectives and structure, by forcing students to review, and by motivating more serious study, it may be expected that the number of tests taken during a course and the manner in which results of tests are communicated to the students would have some effect on achievement. In a given subject and for a given level of student ability, one combination of test frequency may be more effective than another (Ammons, 1956; McKeachie, 1963).

More frequent testing has generally been found to have a favorable effect in mathematics classes (Schunert, 1951; Mach, 1963; Proger, 1968; Nystrom, 1969; Collins, 1971) and in psychology classes (Standlee and Popham, 1960; Feldhusen, 1964). While there are studies in which no significant differences were revealed (Curo, 1963; Selakovich, 1962), in no study has it been found that frequent tests have a detrimental effect.

The simple principle that knowledge of results facilitates learning is one of the few generalizations clearly supported by research on college teaching (McKeachie, 1963). Knowledge of results affects motivation, rate of learning, and level reached by learning. For every task and every state of learning, there is probably an optimum delay for feedback of

results (Ammons, 1956). There is some evidence (More, 1969; English and Kinzer, 1966) that short delayed information feedback enhances retention of higher order learnings.

The purpose of this study was to examine the effects on the achievement of beginning calculus students of four levels of test frequency and two levels of delay of test result feedback, taking into consideration the mathematics aptitude of the students.

Procedure

Subjects. Sixteen beginning analytic geometry and calculus classes during the Fall 1971 academic quarter at California State Polytechnic College were in the study. In each class, three aptitude subgroups were identified. Ten instructors taught the classes. Most of the 442 students in the classes were in their first year of college. Class sizes ranged from 11 to 35 with a mean of 28.

Design. The design of the study was a two-by-four-by-three factorial design with two observations per cell. The three fixed factors were D = delay of feedback of test results (two levels), T = frequency of tests (four levels), and A = aptitude (three levels). Two of the sixteen classes were randomly assigned to each of eight test frequency-feedback delay treatments, with the provision that an instructor having two classes had them assigned to different treatments. The criterion was mathematics achievement.

Variables. Mathematics achievement was measured by an achievement test constructed especially for this study. Concommitant variables were the mean class meeting time, and the mean class size. Other variables were the proportion of student withdrawals from class and the mean score on an attitude scale.

Treatments. The four levels of frequency of tests were:

T_1: 5-10 minute quiz each day; one midterm exam.

T_2: 20-30 minute quiz every fourth or fifth meeting; one midterm exam.

T_3: Three 30-50 minute midterm exams.

T_4: One midterm exam.

The two levels of delay of feedback of test results were:

D_1: All graded quiz and exam papers returned and discussed the next meeting after being taken.

D_2: Graded T_1 quizzes returned and discussed three meetings after being taken; graded T_2 quizzes returned and discussed the meeting before the next quiz; all exams returned and discussed one week after being taken.

All classes were administered the achievement test and the attitude scale at the end of the quarter.

Aside from the experimental treatment combination assigned and the content coverage imposed by the textbook and course outline, each instructor was free to teach and evaluate his students as he wished. To avoid a possible "Hawthorne" effect, students were not informed of the study.

Measuring instruments. The achievement test used as criterion measure was constructed as follows: A 59-item

multiple-choice test tailored to the textbook and content emphasis of the course was constructed and administered to 160 third and fourth quarter calculus students. The test was item-analyzed, "bad" items removed, and a reliable and valid 30-item test obtained. This test was administered in the last week of the quarter. The class mean of the students' scores was a measure of each classes' achievement. The mean KR-20 reliability coefficient of the achievement test was .78.

Scores on the mathematics portion of the College Entrance Examination Board Scholastic Aptitude Test were used to determine aptitude subgroups within each class. The 33rd and 67th percentiles of all SAT scores obtained were used to assign students in each class to high, middle, or low aptitude subgroups.

The Purdue master attitude scale, A Scale to Measure Attitude Toward Any School Subject (Remmers, 1960) was the instrument used to measure attitude. One of a series of Purdue University originated scales, it consists of a list of 17 statements. Students are directed to indicate those statements with which they agree. The student's score is the median scale value of those statements with which he agrees. The class mean of the students' scores was taken as the measure of the classes' attitude.

Analysis. A three-factor analysis of covariance was used to test for differences among cell means on the achievement test. A chi-square test for differences in dropout proportions

was used. Two-factor analyses of variance were used to test for differences among the cell means for each of the concommitant variables and for differences in attitude.

Results

Analyses of variance revealed that classes assigned the D_2 treatment met significantly later in the day with significantly fewer students than those in the D_1 treatment. Class time and size were accordingly used as covariates in the analyses of covariance for differences among cell means in mathematics achievement.

It was determined that the data provided by the achievement test satisfied all the homogeniety and linearity assumptions underlying analysis of covariance for which it was appropriate to test. The adjusted cell means of the achievement test scores are shown in Table 1.

The analysis of covariance revealed that only the F ratios for the mean effects of factors D, T, and A were significant. The highly significant F ratio for factor A (aptitude) implied that the aptitude grouping by Scholastic Aptitude Test was effective. The overall adjusted aptitude means were $A_1 = 55.63$, $A_2 = 62.29$, and $A_3 = 74.53$.

The significance of the F ratio for the main effect of factor T (frequency of tests) implied the existence of significant differences among the means of the four levels of that factor. To make comparisons between means, the Newman-Keuls procedure described by Winer (1962, p. 80) was followed.

Only the difference between T_1 = 70.49 and T_4 = 57.45 was significant at the .05 level.

Table 1: Adjusted Achievement Test Cell Means

	D_1			D_2			
	A_1	A_2	A_3	A_1	A_2	A_3	
T_1	58.56	62.16	69.21	75.92	68.77	88.31	70.49
T_2	46.92	56.82	70.67	54.83	65.98	81.88	62.85
T_3	50.46	56.26	76.16	57.47	78.02	76.52	65.81
T_4	47.78	50.23	63.63	53.11	60.06	69.86	57.45
	50.93	56.37	69.92	60.33	68.21	79.15	
		59.07			69.23		

The significance of the F ratio for the main effect of factor D (delay of test result feedback) implied the superiority of the D_2 treatment, since the adjusted means for the treatments were D_1 = 59.07 and D_2 = 69.23.

A chi-square test for differences in dropout proportions yielded a computed value of $\chi^2_{obs} = 4.05$. The tabular value for three degrees of freedom at the .05 level is $\chi^2_{.05} = 7.82$. Thus no significant differences in dropout proportions among the classes in the eight treatment groups were found.

An analysis of variance was used to test for differences in the mean scores on the attitude scale. The cell means for the attitude scale are given in Table 2. None of the F ratios obtained in the analysis was significant, thus no significant differences in attitude were found among the classes in the eight treatment groups.

Table 2: Cell Means for Attitude

	T_1	T_2	T_3	T_4	
D_1	7.50	7.52	7.68	7.72	7.60
D_2	7.44	7.56	7.83	7.69	7.63
	7.47	7.54	7.76	7.70	7.62

Conclusions

Classes to which short daily quizzes were assigned had higher achievement in calculus, as measured by the achievement test, than classes given other test frequency treatments; significantly higher than those given only a midterm exam. All other differences among the means for the four levels of frequency of tests were not significant. This study thus provides additional evidence for the effectiveness of short frequent tests on achievement.

Classes in the test feedback delay treatment D_2 (long delay) had significantly higher achievement test adjusted mean scores than those in treatment D_1 (short delay). This result is interesting and somewhat surprising. The "optimum delay" suggested by Ammons (1956) apparently is longer than one day for college freshman calculus students. Such students are apparently able to mediate the two time-disconnected events of taking a test and getting back the test results over a longer period advantageously. The discussion of test results after an intervening period during

which other topics are studied apparently serves as an effective review and enhances learning, at least for freshman calculus students.

The fact that there were no significant interaction effects revealed in the analysis of covariance permits the conclusion that students' aptitudes need not be a prime consideration in the determination of a particular technique of frequency of tests and feedback of test results to be used with beginning calculus students. It must be noted that this conclusion necessarily applies only to the very restricted (by college entrance and course enrollment requirements) aptitude range of the subjects in this study.

The various levels of test frequency and feedback of test results did not result in significant differences in attitude, as measured by the Purdue master attitude scale employed in this study, nor were there significant differences among the proportions of students who withdrew from the classes.

References

Ammons, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. Journal of general psychology, 1956, 54, 279-299.

Collins, K. M. An investigation of the variables of Bloom's mastery learning model for teaching junior high school mathematics. Unpublished doctoral dissertation, Purdue University, 1971.

Curo, D. M. An investigation of the influence of daily pre-class testing on achievement in high school American history classes. Unpublished doctoral dissertation, Purdue University, 1963.

Dixon, W. J. (Ed.) Biomedical computer programs, X-series supplement. Berkeley: University of California Press, 1969.

English, R. A., & Kinzer, J. R. The effects of immediate and delayed feedback on retention of subject matter. Psychology in the schools, 1966, 3, 143-147.

Feldhusen, J. F. Student perception of frequent quizzes and post-mortem discussion of tests. Journal of educational measurement, 1964, 6, 51-54.

Mach, G. M. A comparative study of student performance in an intermediate calculus class as a result of different evaluation programs. Unpublished doctoral dissertation, Purdue University, 1963.

McKeachie, W. J. Research on teaching at the college and university level. In N. L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963.

More, A. J. Delay of feedback and the acquisition and retention of verbal materials in the classroom. Journal of educational psychology, 1969, 60, 339-342.

Nystrom, N. K. An experimental study to compare the relative effects of two methods of instruction on learning of intermediate algebra. Unpublished doctoral dissertation, Arizona State University, 1969.

Proger, B. B. The relationship between four testing programs and the resultant achievement and test anxiety levels of high- and low-previous achievement sixth-grade arithmetic students. Unpublished doctoral dissertation, Lehigh University, 1968.

Remmers, H. R. Manual for the Purdue master attitude scales. W. Lafayette, Indiana: University Book Store, 1960.

Schunert, J. The association of mathematical achievement with certain factors resident in the teacher. Journal of experimental education, 1951, 219-238.

Selakovich, D. An experiment attempting to determine the effectiveness of frequent testing as an aid to learning in beginning college courses in American government. Journal of educational research, 1962, 1, 178-180.

Standlee, L. S., & Popham, W. J. Quizzes contribution to learning. Journal of educational psychology, 1960, 12.

Townsend, N. R. The relationship of frequency of tests and delay of feedback of test results to achievement in first quarter analytic geometry and calculus. Unpublished doctoral dissertation, Purdue University, 1972.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.