The Journal of Educational Research

ISSN: 0022-0671 (Print) 1940-0675 (Online) Journal homepage: http://www.tandfonline.com/loi/vjer20

Effect of Verbal Feedback of Test Results upon Learning

Kenneth N. Wexley & Cash L. Thornton

To cite this article: Kenneth N. Wexley & Cash L. Thornton (1972) Effect of Verbal Feedback of Test Results upon Learning, The Journal of Educational Research, 66:3, 119-121, DOI: 10.1080/00220671.1972.10884423

To link to this article: http://dx.doi.org/10.1080/00220671.1972.10884423

THE JOURNAL OF EDUCATIONAL RESEARCH
(Volume 66, Number 3, November 1972)

Effect of Verbal Feedback of Test Results upon Learning

KENNETH N. WEXLEY and CARL L. THORNTON
The University of Akron

ABSTRACT

The learning effects of providing teacher verbal feedback of quiz results were studied using two classes of college students. The relationship between amount of learning as a result of feedback and certain characteristics of the students was also studied. After each quiz, students were given feedback on half of the test questions. Their final examination contained an equal number of feedback and nonfeedback items repeated from the previous quizzes. Students did significantly better on those items on which they had originally received feedback, even in the case where the feedback had been given 9 weeks prior to the final examination. These learning gains were found to be unrelated to students' age, grade point average, number of hours employed, and number of college credits accumulated. Theoretical and practical implications are discussed.

IT HAS BEEN established, both in the experimental laboratory and in classroom settings, that feedback is important for learning to occur (1, 4). Various learning theories emphasize that feedback must be "immediate" in order to be effective. Apparatus devices have been developed which instantaneously convey to the pupil that his response is either correct or incorrect.[1] Test feedback of this type has been found to be effective as measured by means of subsequent retesting. There is, however, no experimental evidence to show whether or not delay of feedback until the end of testing would lessen its utility. It should be noted that any loss in effectiveness of feedback due to delay of presentation may be compensated for by the increase in information that can be transmitted in the feedback itself. That is, verbal feedback at the completion of a test can convey reasons or explanations as to why a response was correct or incorrect rather than simply the right-wrong distinction provided by apparatus devices.

Zahorik (17), in a study of the types of verbal feedback given in the classroom, found that teachers tended to utilize very few types of feedback. Of the types they did utilize, very few contained cues as to the reasons why a response had merit. Of course, such verbal feedback is difficult to administer in large classes. Testing, however, offers a setting where teacher verbal feedback should be beneficial to all students by reinforcing correct answers and correcting errors. Such feedback requires only 10–20 minutes of class time and may be quite effective.

The purpose of this experiment was to determine if, in fact, verbal feedback enhances learning as measured by test performance 1 to 2 months after original administration of test items. The study also investigated whether there is a relationship between the effectiveness of teacher verbal feedback and certain student characteristics.

Experiment I

Subjects. The day class Ss were 169 undergraduate students (161 males, eight females) taking a course in psychology at a midwest university. They ranged in age from 18 to 31, with a median of 20.5 years. More than half were business majors and had completed one other psychology course. Sixty-three percent of them held outside jobs, working a median of 21 hours per week. The breakdown of Ss by college level was as follows: 28 percent freshmen; 49 percent sophomores; 17 percent juniors; and 6 percent seniors. The above data were obtained by means of a questionnaire administered during the first week of the term. Subjects were not told that they were participating in an experiment, with all data being obtained as an integral part of the routine testing of the class.

Apparatus. Test questions were individually presented on 35 mm. slides and projected for 45 seconds by means of a Kodak Carousel projector on an 8 x 10 foot screen.

Procedure. Subjects were given four quizzes during the term, each consisting of thirty-five multiple-choice questions. After each test, Ss were reshown eighteen test items. The instructor read aloud each question, gave the answer, and read a three or four sentence explanation as to why the answer was correct. The presentation of these eighteen feedback (FB) items took approximately 10 minutes. Subjects were told that a lack of time prevented the reviewing of the remaining seventeen nonfeedback (NF) items. The class period did, in fact, end about the time the instructor finished the FB items.

A final examination was given at the end of the term containing a set of thirty-eight FB

and thirty-eight NF items repeated from the previous quizzes. These two sets of items were matched using the procedure devised by Gulliksen (7) so that they formed parallel tests having equal means, variances, and intercorrelations.

Subjects were told during the course that the final examination would cover material from the entire term's work. They were not told, however, that items from their previous quizzes would reappear.

Experiment II

Subjects. Experiment I was replicated by using ninety-two (eighty-six males, six females) different Ss enrolled in an evening class in psychology at the same university. They ranged in age from 17 to 48, with a median of 25.5 years. The majority were business majors and had taken one course in psychology. All held jobs, working a median of 40 hours per week. The sample consisted of 16 percent freshmen; 37 percent sophomores; 18 percent juniors; and 29 percent seniors. They were not told that they were part of a study.

Procedure. Procedures, including the apparatus, for this experiment were identical to experiment I, with the following exceptions: (a) a different instructor taught the evening class, (b) different test questions were used, and (c) the final examination contained thirty FB and thirty NF items repeated from three, rather than four, previous quizzes.

Results

The final examination yielded two scores, FB and NF, for each S. Each score was arrived at by summing the number of FB or NF items which an S had answered correctly. The corrected split-half reliability coefficients of these scores were computed for the day and evening classes. The four resultant coefficients ranged from .67 to .77 and are all significantly greater than zero ($p < .01$).

Table 1 presents matched-pair t-tests of the differences in means of the FB and NF scores. In both the day and evening classes, results indicate that Ss got significantly higher FB than NF scores on the final examination. Evidently, this type of teacher verbal feedback facilitates learning.

Table 1.—Comparisons Between FB and NF Scores in Day and Evening Classes

	Comparison		
Class	df	M_{diff}	t
Day	169	5.02	17.86**
Evening	92	2.63	7.11**

**p < .01, two-tailed test.

Table 2.—Average Proportions (P) of FB and NF Items in the Two Classes

Quiz No.	Condition	Average P on Quiz	Average P on Final	Difference
1	FB-d*	.63	.69	+.06
1	FB-e	.60	.69	+.09
1	NF-d	.62	.62	+.00
1	NF-e	.61	.58	-.03
2	FB-d	.63	.69	+.06
2	FB-e	.51	.62	+.11
2	NF-d	.59	.59	+.00
2	NF-e	.59	.57	-.02
3	FB-d	.61	.84	+.23
3	FB-e	.64	.79	+.15
3	NF-d	.74	.76	+.02
3	NF-e	.58	.56	-.02
4	FB-d	.58	.72	+.14
4	FB-e	—	—	—
4	NF-d	.53	.52	-.01
4	NF-e	—	—	—
Combined	FB	.59	.72	+.13
	NF	.59	.59	+.00

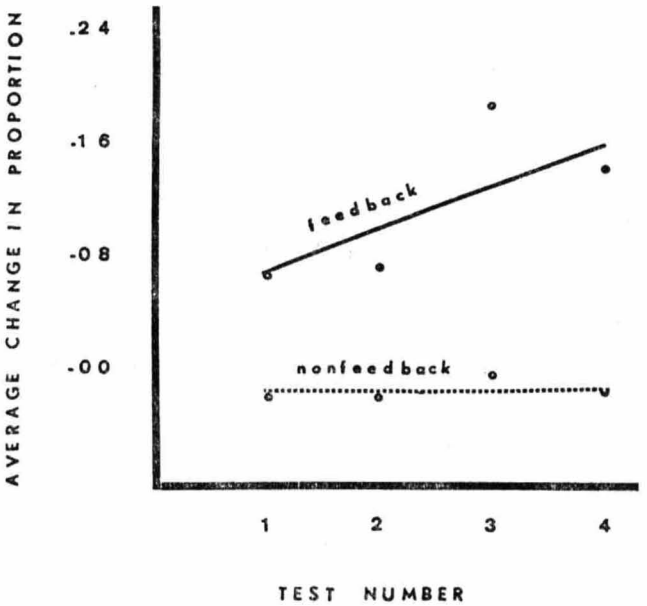
*d designates day; e designates evening.

Another way of tabulating the data is to average across test items, rather than across Ss' scores. The number of Ss correctly answering an item was divided by the total number of Ss to yield a proportion (P), an index of item difficulty. Next, the average P values for the FB and NF items on each quiz were obtained. The average P's were also determined for the same sets of items on the final examination. Table 2 shows the original P values, the final P values, and the differences between the two.[2] Since the differences in P values are quite similar for both classes, the results were averaged and presented in Figure 1.

It is evident from Figure 1 that there is some forgetting due to the time lag between each quiz and the final examination. That is, the difference between the FB and NF lines in Figure 1 decreases from quiz four back through quiz one. Despite the effect of this forgetting, the difference between FB and NF proportions for the first quiz is statistically significant ($t = 1.74$; $df = 18$; $p < .05$). Thus, the effects of feedback hold up well over the 9-week time interval.

Correlations between the Ss' gain scores and various student characteristics were computed separately for the two classes. Student characteristics studied were: age, number of college credits accumulated, number of hours employed, and

Figure 1.—Average Change in Proportions for the Four Quizzes



GPA. All correlations were nonsignificant ($p > .05$), indicating that the learning gains due to feedback are unrelated to any of the above student characteristics.

Discussion

The results confirm the assumption that posttest verbal feedback does facilitate learning. Thus, the list of teacher verbal feedback used by Zahorik (17) can be profitably increased by teachers by including this type of feedback in their repertoire.

For posttest verbal feedback to be effective, the test itself must be moderately difficult. Obviously, if 90–100 percent of the class correctly answers most of the items, the feedback has instructional value for only a small number of students. It should be noted in Table 2 that average P values ranged from .53—.74, representing moderate difficulty. This is probably a factor contributing to the success of the posttest feedback in this experiment.

The results of this study suggest that feedback does not have to be instantaneous to be effective. In this experiment, a delay of about 10 to 30 minutes was involved between the presentation of a test item and the verbal feedback. Despite this time lag, improved performance was found on a retest given more than 2 months later. These results run counter to the traditional caution against delaying feedback if it is to be effective (1, 2, 4).

This study investigated the effects of knowledge of quiz results on memory of factual information. Future research could explore the effects of this method on other types of learning such as application, synthesis, analysis, and interpretation of facts and principles (5). Despite these research questions that are as yet unanswered, it appears evident that the relatively straight forward and inexpensive verbal presentation of feedback after an examination is an effective means of increasing the amount of information the student will acquire.

FOOTNOTES

1. Typical of the apparatus designed for this purpose are: punchboards (2, 3, 8, 9, 15, 16), chemically treated answer sheets (11, 12), Pressey's multiple-choice devices (10, 13, 14, 15), and other devices which are more complex (6, 16).
2. The reason for the missing data in Table 2 is that on the fourth quiz in the evening class students insisted on reviewing all the test items. Rather than interfere with the class for the purpose of research, students were given feedback on all items. The fourth quiz was, therefore, deleted from the study.

REFERENCES

1. Ammons, R. B., "Knowledge of Performance, Survey of Literature, Some Possible Applications, and Suggested Experimentation," WADC Technical Report 54–14, Wright Air Development Center, 1954.
2. Angell, G. W., "Effect of Immediate Knowledge of Quiz Results on Final Examination Scores in Freshman Chemistry," *The Journal of Educational Research*, 42:391–394, 1949.
3. Angell, G. W.; Troyer, M. E., "A New Scoring Test Device for Improving Instruction," *School and Society*, 67:84–85, 1948.
4. Annett, J., "The Role of Knowledge of Results in Learning: A Survey," U.S. NAVTRADEVCEN Technical Document Report 342–343, 1961.
5. Bloom, B. S. (ed.), *Taxonomy of Educational Objectives*, Longmans Green, New York, 1956.
6. Briggs, L. T., "Two Self-Instructional Devices," *Psychological Reports*, 4:671–676, 1958.
7. Gulliksen, H., *Theory of Mental Tests*, Wiley, New York, 1961, pp. 208–210.
8. Jensen, B. T., "An Independent-Study Laboratory Using Self-Scoring Tests," *The Journal of Educational Research*, 43:134–137, 1949.
9. Jones, H. L.; Sawyer, M. O., "A New Evaluation Instrument," *The Journal of Educational Research*, 42:381–385, 1949.
10. Little, J. K., "Results of Using Machines for Testing and for Drill Upon Learning in Educational Psychology," *The Journal of Experimental Education*, 3:45–49, 1934.
11. Peterson, J. C., "A New Device for Teaching, Testing, and Research Learning," *Transactions of the Kansas Academy of Science*, 33:41–47, 1930.
12. Peterson, J. C., "The Value of Guidance in Reading for Information," *Transactions of the Kansas Academy of Science*, 34:291–296, 1931.
13. Pressey, S. L., "A Simple Apparatus Which Gives Tests and Scores—and Teaches," *School and Society*, 23:373–376, 1926.
14. Pressey, S. L., "A Machine for Automatic Teaching of Drill Material," *School and Society*, 25:549–552, 1927.
15. Pressey, S. L., "Development and Appraisal of Devices Providing Immediate Automatic Scoring of Objective and Concomitant Self-Instruction," *Journal of Psychology*, 29:417–447, 1950.
16. Skinner, B. F., "The Science of Learning and the Art of Teaching," *Harvard Educational Review*, 24:86–97, 1954.
17. Zahorik, J. A., "Classroom Feedback Behavior of Teachers," *The Journal of Educational Research*, 62:147–150, 1968.