Does Testing Type Matter? Comparing the Impact of Instant and Traditional Feedback Methods on Student Learning[1]

Dustin K. Grabsch[2] and Lori L. Moore
Texas A&M University
College Station, TX

Abstract

Recent research on student learning has suggested that immediate feedback provides better retention of information and, therefore, better student performance as compared with delayed feedback. Here, students completed four course exams during one semester that used response formats providing either delayed feedback or immediate feedback. Then, a quasi-experimental research design was utilized to compare student performance on course exams and a final exam considering testing type—delayed (traditional) feedback (Scantron sheets) versus instant feedback (immediate feedback assessment technique). The comprehensive final exam consisted of 50 items, with 20 items identical to questions from the previous four course exams. Scores on each of the four course exams did not differ (p=0.082, p=0.058, p=0.053, p=0.913) in their performance by test method. Results allow us to conclude that testing type and its associated feedback method have minimal positive effects on student performance.

Introduction

A necessary aspect of education is assessing student learning and providing feedback for improvement. One tool introduced to facilitate feedback and improvement is the teaching-testing machine proposed by Pressey (1926). The teaching-testing machine prompted the development of a number of instructional techniques that facilitate immediate feedback during multiple-choice questioning (Dihoff et al., 2003). The teaching-testing machine provided immediate feedback while also transforming the student from a passive gatherer of information to an active demonstrator of knowledge and skill (Skinner, 1958). Even with the broad use of the teaching-testing machine, educator feedback methods and their timeliness may not be identical in aiding student learning.

Traditional and instant feedback methods are discussed here as they are the primary assessment approaches utilized by educators today; however, much of the existing body of literature is rather dated. In the early 1960s, delayed or "traditional" feedback across brief intervals was suggested to promote the retention of meaningful material (Brackbill et al., 1962). This outcome was observed when feedback was delayed for 24 to 48 hours and retention intervals were lengthened up to seven days (Kulhavy and Anderson, 1972; O'Neill et al., 1976; Surber and Anderson, 1975). Generally, proponents of delayed feedback adhere to the interference perseveration hypothesis (Kulhavy and Anderson, 1972), which states that initial errors do not compete with to-be-learned correct responses if the correct response is delayed. The reason that interference does not affect student learning is because errors are likely to be forgotten and, thus, cannot impact retention.

The benefits of the delay-retention effect (DRE), commonly referred to as delayed feedback, are supported when comparing initial and retention test performances (Kulhavy and Anderson, 1972; Surber and Anderson, 1975). Although DRE has been unsupported in a few studies (Kippel, 1975; Newman et al., 1974; Phye and Andre, 1989), delayed feedback typically has suggested just as effective as immediate feedback. The traditional (delayed) feedback method remains a popular mode to assess student learning.

In contrast to traditional feedback methods, some educators also regularly employ instant feedback methods to assess student learning. Delayed feedback proponents recommend delays anywhere from one to two days or more (Brosvic and Epstein, 2007; Dihoff et al., 2004; Dihoff et al., 2003; Kulhavy et al., 1990) to seven days or more (Bruning et al., 2011; Robin, 1978) to facilitate the forgetting of errant responses and the learning of correct responses by students. Immediate feedback proponents, on the other hand, endorse students correcting an incorrect response and providing the correct response before exiting the test session (Brosvic and Epstein, 2007; Brosvic et al., 2006a; 2006b). In summary, instant and traditional feedback definitions are distinct in the fact that instant feedback relies on the learner's awareness of their errors and the correct response by the conclusion of the assessment period.

Some research on instant feedback has reported an increase in student learning retention compared to that observed for delayed feedback. Within the immediate feedback literature, it is difficult to compare results among studies because of noteworthy differences in the operational definitions used for immediate and delayed feedback (Brosvic and Epstein, 2007). For example, immediate feedback definitions range from the instantaneous presentation of the correct response (Epstein et al., 2003) to the provision of correct responses at the next weekly class meeting (Robin, 1978). A review of the literature to date suggests that no clear consensus exists to operationalize the intervals of time contained within

[1]The Institutional Review Board approved the study protocol, and all participants provided written informed consent prior to participation in the study.
[2]Agricultural Leadership, Education, and Communications, 979.862.9172, dgrabsch@tamu.edu.

traditional versus instant feedback methods. It is likely that even the median of the intervals presented above may represent considerably more prompt grading than that to which many students are accustomed.

Many feedback methods and techniques exist to provide students with assessment of their learning. Since the early 2000s, educational psychologists have been studying, refining, and validating an instant feedback procedure known as the immediate feedback assessment technique (IF-AT) (Dihoff et al., 2003, 2004; Epstein et al., 2002, 2003). These studies and others have documented positive student perceptions of the IF-AT. Feedback includes the general positive effect and the usefulness in aiding student learning (Cotner et al., 2008; DiBattista et al., 2004; Epstein and Brosvic, 2002). A growing undergraduate acceptance of the IF-AT for multiple-choice testing (DiBattista et al., 2004) has led to the exploration of its expanded use in science, technology, engineering, and mathematics (STEM) disciplines (Epstein et. al., 2002). Little to no research exists using the IF-AT instrument to aid student performance within the context of agricultural education and leadership.

The primary focus of the present study was to examine student learning using the IF-AT as compared to traditional Scantron sheets on multiple-choice course and final exams. Unlike prior studies that have examined specialized systems of instruction or utilized different research designs and samples, the specific objectives of the study included the following:

1. Compare the impact of testing methods on achievement on regular course exams
2. Determine if immediate feedback on regular course exams impacts achievement on a comprehensive final exam

Methods

Participants

The population of interest for this study consisted of students enrolled for the fall 2014 semester in an undergraduate agricultural leadership and development course focused on leadership theory at a land-grant university in the southern United States. Two-hundred students were enrolled in the course. The accessible population consisted of 179 of the 200 students who consented to participate in the study. Usable data were obtained from 163 of the 179 students, yielding a participation rate of 91%. Usable data created two groups for analysis based on testing method. Scantron sheets served as the traditional testing method (n=80), and the IF-AT served as the instant testing method (n=83).

Design and Procedures

The general format for the course comprised two components. (1) All students enrolled in the course attended a twice-weekly 50-minute lecture. (2) There was also a 60-minute weekly recitation requirement for the course, consisting of eight groups of 25 students each. The course required students to take four exams during the regular semester and a comprehensive final exam. The regularly scheduled course exams and the comprehensive final exam consisted of single-best-type multiple-choice questions, fill-in-the-blank questions, and short-answer questions. There were 20 multiple-choice questions on each regularly scheduled course exam and 40 multiple-choice questions on the comprehensive final exam. On the final exam, 20 of the questions were drawn from previous exams while 20 of the questions were new. The multiple-choice section of all of the exams constituted 80% of the total exam score.

To accomplish the objectives of this study, a quasi-experimental design was employed (Campbell and Stanley, 2015). Four of the recitation groups were selected randomly to complete the four regularly scheduled course exams using Scantron forms, while the remaining four recitation groups were selected randomly to complete the four regularly scheduled course exams on the IF-AT forms. The IF-AT form is a multiple-choice answer sheet with rows of rectangular answer spaces (e.g., A, B, C, D) that is nearly identical in layout to the pervasive, machine-scored answer sheet available from Scantron Corporation. Epstein Educational Enterprises (n.d.), the maker of the IF-AT, has described the form in this way: "*Participants scrape off an opaque, waxy coating that covers an answer space on the IF-AT form to record their answer . . . . If a symbol (e.g., a star) is printed beneath the covering, the student receives instant feedback that a correct choice was made; the absence of a symbol provides instantaneous feedback that an incorrect choice was made.*" (para. 5)

Rather than simply exiting the question, the student reviews the remaining response options, continues to respond until the correct answer is obtained (a self-correction procedure), and completes each question with the correct answer (Epstein Educational Enterprises, n.d.). For this study, credit for a correct response was provided only if the correct response was selected on the first attempt.

The four groups completing the exams using Scantron sheets received feedback approximately one week after the exams were graded. Feedback was provided by reviewing correct exam answers during the next class period. The four groups completing the exams on IF-AT forms, if used correctly, knew the correct answer to all of the multiple-choice questions prior to leaving the exam session. While correct answers were reviewed, Scantron and IF-AT students were not provided with copies of their exams to prepare for the final exam; however, study guides were provided to guide students to focus on key course concepts.

In order to determine the effect of testing type on student learning, five questions for each of the four exams were kept identical to questions the students would later experience on the comprehensive final exam. Notes in Table 3 link the exam question numbers and their corresponding questions on the final exam. We compared each of the four student exams, as well as their performance on the comprehensive final exam, by testing type using independent-sample t-tests.

In order to conduct the comparison on the comprehensive final exam, student performance was transformed into two new dummy variables to account for the previous corresponding exam performance. First, a simple dummy variable was created to capture incorrect responses on the exam but a correct response on the identical final exam question. This variable is of most interest to the researchers as this would likely suggest student learning from the exam to comprehensive final. A zero indicated incorrect on the final and a one indicated correct response on the final. Two-tailed t-tests were conducted on the dummy variable for each of the twenty final exam questions by test type.

Second, another dummy variable was created to capture different variations of student performance between the exam and the comprehensive final. Table 1 summarizes the creation of this second dummy scale variable ranging from 1 to 4. It is important to note that transformed variables 2 and 4 both could indicate no difference in student learning. To illustrate, if a student response were the same on the course exam and the comprehensive final exam, the learning could have occurred before the first exam and may or may not have been connected to testing method. It may also have affirmed the learning to reinforce performance on the final. Thus, transformed variable 3 is the most likely to be connected to testing method.

Table 1. Final Exam Performance Based on Course Exam Performance

Course Exam Responses	Final Exam Reponses	Transformed Variable
Correct (0)	Incorrect (1)	1
Incorrect (1)	Incorrect (1)	2
Incorrect (1)	Correct (1)	3
Correct (0)	Correct (0)	4

Results and Discussion

Objective one of this study sought to compare the impact of testing method on achievement on regular course exams. The overall performance of each exam was compared by test type. Independent-sample t-tests for each exam revealed no statistical differences as a function of test type (see Table 2), although exams two and three approached significance (p=0.058 and p=0.053, respectively). Thus, the results of objective one led the researchers to conclude that testing type alone has no impact on performance on a single test.

Table 2. Independent-Sample T-Test of Course Exam Performance by Testing Method

Course Exam	t	p	95% CI
Exam 1	–0.251	0.802	[–3.770, 2.919]
Exam 2	1.911	0.058	[–0.111, 6.742]
Exam 3	1.951	0.053	[–0.036, 5.836]
Exam 4	0.110	0.913	[–2.972, 3.322]

The results indicated similarities between exams one and four and similarities between exams two and three. Exam one scores could be lower due to students familiarizing themselves with the instructor's expectations and the types of exam questions therefore resulting in a lower exam average. Exam four scores may also be lower as the course content near the end of the semester is distinct from the previous content and may be more challenging for students to learn. Exam four also occurred following a fall break period where students may not have prepared or studied as much for this exam. Exams two and three performance is similar. This may be explained by the course content being shared can be expected and occurred during the regular rhythm of the semester uninterrupted by breaks.

Objective two sought to determine if immediate feedback on regular course exams impacted achievement on a comprehensive final exam. Independent-sample two-tailed t-tests were conducted to examine the difference between instant and traditional feedback testing methods in relation to respondents' performance on each of the 20 comprehensive final exam items. The independent-sample two-tailed t-test was fist conducted on the simple dummy variable that was created to capture incorrect responses on the exam but a correct response on the identical final exam question.

Table 3 contains the analysis of the t-tests on the simple dummy variable. The test revealed a statistically significant difference between instant and traditional testing methods for seven final exam questions, including: 9 (t=1.397, df=161, p<0.01), 19 (t=-1.801, df=161, p<0.00), 20 (t=-1.128 , df=161, p<0.02), 23 (t=-2.578, df=161, p<0.00), 29 (t=-2.715, df=161, p<0.00), 32 (t=-1.405, df=161, p<0.005), and 34 (t=1.282, df=161, p<0.01). A statistically significant number of IFAT users answered more comprehensive final exam questions correctly after incorrect responses on their preceding exams for final exam questions 9 (M=0.125) and 34 (M=0.229) compared to scantron users. Scantron users answered a statistically significant number of more final exam questions correctly after incorrect exam responses on final exam questions 19 (M=0.163), 20 (M=0.125), 23

Table 3. T-Test by Simple Incorrect on Exam then Correct on Comprehensive Final Exam Question

	Scantron Form		IF-AT					
Variable	M	SD	M	SD	T	p	95% CI	Cohen's d
Final Q1[4]	0.01	0.11	0.01	0.11	-0.026	0.959	[-0.034, 0.033]	
Final Q3[2]	0.11	0.32	0.13	0.34	0.387	0.439	[-0.082, 0.122]	
Final Q4[2]	0.09	0.28	0.13	0.34	0.914	0.067	[-0.052, 0.142]	
Final Q8[1]	0.40	0.49	0.55	0.50	1.982	0.257	[0.001, 0.308]	
Final Q9[3]	0.00	0.00	0.02	0.15	1.397	0.005	[-0.010, 0.058]	0.179
Final Q15[3]	0.10	0.30	0.12	0.33	0.415	0.407	[-0.077, 0.118]	
Final Q18[1]	0.11	0.32	0.10	0.30	-0.335	0.504	[-0.111, 0.079]	
Final Q19[4]	0.16	0.37	0.07	0.26	1.982	0.000	[-0.189, 0.009]	0.281
Final Q20[2]	0.13	0.33	0.07	0.26	-1.128	0.024	[-0.145, 0.040]	0.201
Final Q22[4]	0.11	0.32	0.13	0.34	0.387	0.439	[-0.082, 0.122]	
Final Q23[4]	0.08	0.27	0.00	0.00	-2.578	0.000	[-0.132,-0.017]	0.414
Final Q24[2]	0.08	0.27	0.10	0.30	0.484	0.333	[-0.066, 0.109]	
Final Q25[1]	0.05	0.22	0.06	0.24	0.284	0.570	[-0.061, 0.081]	
Final Q29[4]	0.02	0.15	0.14	0.35	-2.715	0.000	[-0.196,-0.031]	0.446
Final Q32[3]	0.05	0.22	0.01	0.11	-1.405	0.005	[-0.091, 0.015]	0.230
Final Q33[3]	0.03	0.16	0.02	0.15	-0.037	0.941	[-0.049, 0.047]	
Final Q34[1]	0.15	0.36	0.23	0.42	1.282	0.010	[-0.043, 0.200]	0.205
Final Q37[3]	0.10	0.30	0.07	0.26	-0.628	0.209	[-0.114, 0.059]	
Final Q38[2]	0.04	0.19	0.02	0.15	-0.493	0.324	[-0.067, 0.040]	
Final Q39[1]	0.04	0.19	0.04	0.19	-0.046	0.927	[-0.060, 0.057]	

Note: Superscript notations relate to the course exam (i.e., 1, 2, 3, or 4) where the identical final exam question was first posed to respondents.

(*M*=0.075), 29 (*M*=0.138), and 32 (*M*=0.050). Therefore since scantron users may have demonstrated more student learning on five final exam questions compared to their IF-AT user counterparts who did so on two final exam questions.

Next, the t-test analysis was reviewed that compared the four types of student performance (see Table 1) and each of the final exam questions. The test revealed a statistically significant difference between instant and traditional testing methods for final exam questions 1 (t=2.183, df=161, p<0.05) and 4 (t=2.723, df=161, p<0.05). On final exam question 1, IF-AT respondents (*M*=3.92, *SD*=0.474) performed significantly higher than that of the Scantron form users (*M*=3.66, *SD*=0.927). Exam question 4 also showed IF-AT users (*M*=2.98, *SD*=1.18) performing at a higher statistically significant score than Scantron users (*M*=2.48, *SD*=1.17). Final exam question 20 was shown to approach statistical significance (t=1.923, df=161, p<0.05). If final exam question 20 were to report significance, it too would suggest IF-AT users (*M*=2.23, *SD*=0.941) performing at a higher score than Scantron users (*M*=1.963, *SD*=0.818). Thus, 2 out of 20 possible items resulted in a statistically different outcome on performance by testing method. Table 4 presents the results of the independent t-test for each of the 20 final exam questions, as well as the effect size using Cohen's d for tests, which resulted in statistical significance. These results suggest that immediate feedback on regular course exams does not increase achievement on a comprehensive final exam.

Table 4. T-Test by Four Different Student Performances on Comprehensive Final Exam Question

	Scantron Form		IF-AT					
Variable	*M*	*SD*	*M*	*SD*	T	*p*	95% CI	Cohen's d
Final Q1[4]	3.66	0.93	3.92	0.47	2.210	0.031	[0.024, 0.482]	0.344
Final Q3[2]	3.20	1.16	3.28	1.11	0.450	0.653	[-0.273, 0.434]	
Final Q4[2]	2.48	1.17	2.98	1.18	2.720	0.007	[0.138, 0.864]	0.427
Final Q8[1]	3.41	0.74	3.36	0.60	-0.490	0.087	[-0.259, 0.156]	
Final Q9[3]	3.96	0.34	3.94	0.36	-0.405	0.686	[-0.137, 0.086]	
Final Q15[3]	2.90	1.22	2.98	1.20	0.407	0.685	[-0.300, 0.452]	
Final Q18[1]	3.53	0.98	3.63	0.88	0.696	0.232	[-0.186, 0.389]	
Final Q19[4]	3.24	1.08	3.07	1.16	-0.941	0.348	[-0.512, 0.181]	
Final Q20[2]	1.96	0.82	2.23	0.09	1.930	0.056	[-0.007, 0.540]	
Final Q22[4]	3.34	1.11	3.31	1.14	-0.138	0.891	[-0.372, 0.324]	
Final Q23[4]	3.74	0.76	3.82	0.72	0.707	0.481	[-0.147, 0.310]	
Final Q24[2]	2.64	1.38	3.01	1.28	1.790	0.075	[0.038, 0.787]	
Final Q25[1]	3.64	0.90	3.65	0.86	0.095	0.787	[-0.260, 0.286]	
Final Q29[4]	3.20	1.12	3.25	1.23	0.288	0.774	[-0.311, 0.417]	
Final Q32[3]	3.81	0.64	3.71	0.86	-0.828	0.409	[-0.336, 0.138]	
Final Q33[3]	3.84	0.63	3.77	0.75	-0.588	0.557	[-0.289, 0.152]	
Final Q34[1]	2.23	1.15	2.46	1.05	1.350	0.410	[-0.107, 0.573]	
Final Q37[3]	3.46	1.02	3.49	1.03	0.237	0.813	[-0.280, 0.357]	
Final Q38[2]	3.89	0.50	3.90	0.48	0.208	0.835	[-0.127, 0.169]	
Final Q39[1]	3.74	0.81	3.90	0.43	1.630	0.106	[-0.007, 0.540]	

Note: Superscript notations relate to the course exam (i.e., 1, 2, 3, or 4) where the identical final exam question was first posed to respondents.

The present study was undertaken to examine the impact of testing type on student learning and performance. Students completing the four course exams on IF-AT forms, and therefore receiving immediate feedback, did not perform significantly better on the whole as those completing the exams using traditional Scantron sheets when considering performance on the final exam. Collectively, the present results demonstrate that scantrons may have demonstrated more student learning when using a simple dummy variable on five out of seven final exam questions. When considering more student performance possibilities, immediate feedback may promote recall in two comprehensive final exam questions, but not to a significant level that would warrant the cost associated with the immediate feedback instrument.

This study's IF-AT results are similar to those observed in prior studies, despite considerable variation in the definition of immediate feedback (i.e., feedback provided immediately after a response to feedback provided by the end of the day) and the use or nonuse of an answer-until-correct process, partial (50% of test items) to complete feedback. The researchers do acknowledge that in order for immediate feedback to improve performance, the method needs to be utilized properly. Anecdotal evidence suggests that some students did not use the forms correctly to receive immediate feedback before leaving the test-taking environment. Thus, future studies ought to provide more extensive instructions to participants using the IF-AT instrument in order to elicit the benefits of immediate feedback.

Research should be conducted to determine and assess other appropriate methods of incorporating the IF-AT within agricultural and leadership education. The IF-AT provides individualized performance feedback during the testing process, regardless of the size of a lecture course. The results observed in other studies using the IF-AT may provide small-group learning experiences, opportunities for partial credit, or other alternatives to multiple-choice examination.

Summary

The present study primarily assessed the impact of testing method on student achievement and learning. While some promising outcomes were identified, additional investigation is warranted. To these ends, the IF-AT is presented to the larger community of educators for continued validation and development.

Literature Cited

Brackbill, Y., A. Bravos and R.H. Starr. 1962. Delay-improved retention of a difficult task. Journal of Comparative and Physiological Psychology 55: 947-952. DOI: 10.1037/h0041561.

Bruning, R.H., G.J. Schraw and R.R. Ronning. 2011. Cognitive psychology and instruction. 5th ed. Prentice-Hall, Inc.

Brosvic, G.M. and M.L. Epstein. 2007. Enhancing learning in the introductory course. The Psychological Record 57(3): 391-408. DOI: 10.1007/BF03395584.

Brosvic, G.M., R.E. Dihoff, M.L. Epstein and M.L. Cook. 2006a. Feedback facilitates the acquisition and retention of numerical fact series by elementary school students with mathematics learning disabilities. The Psychological Record 56(1): 35-54.

Brosvic, G.M., M.L. Epstein, R.E. Dihoff and M.J. Cook. 2006b. Acquisition and retention of Esperanto: The case for error correction and immediate feedback. The Psychological Record 56(2): 205-218. DOI: 10.1007/BF03395545.

Campbell, D.T. and J.C. Stanley. 2015. Experimental and quasi-experimental designs for research. Ravenio Books.

Cotner, S.H., B.A. Fall, S.M. Wick, J.D. Walker and P.M. Baepler. 2008. Rapid feedback assessment methods: Can we improve engagement and preparation for exams in large-enrollment courses? Journal of Science Education & Technology 17(5): 437-443. DOI: 10.1007/s10956-008-9112-8.

DiBattista, D., J.O. Mitterer and L. Gosse. 2004. Acceptance by undergraduates of the immediate feedback assessment technique for multiple-choice testing. Teaching in Higher Education 9(1): 17-28. DOI: 10.1080/135625103 2000155803.

Dihoff, R.E., G.M. Brosvic and M.L. Epstein. 2003. The role of feedback during academic testing: The delay retention effect revisited. The Psychological Record 53(4): 533-548. DOI: 10.1007/BF03395451.

Dihoff, R.E., G.M. Brosvic and M.L. Epstein. 2004. Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. The Psychological Record 54(2): 207-231. DOI: 10.1007/BF03395471.

Epstein, M.L. and G.M. Brosvic. 2002. Students prefer the Immediate Feedback Assessment Technique. Psychological Reports 90(3): 1136-1138. DOI: 10.1177/003329410209000315.2.

Epstein Educational Enterprises. n.d. What is the IF-AT? Retrieved December 31, 2018. http://www.epsteineducation.com/home/about/.

Epstein, M.L., G.M. Brosvic, K.L. Costner, R.E. Dihoff and A.D. Lazarus. 2003. Effectiveness of feedback during the testing of preschool children, elementary school children, and adolescents with developmental delays. The Psychological Record 53(2): 177-195. DOI: 10.1007/BF03395439.

Epstein, M.L., A.D. Lazarus, T.B. Calvano, K.A. Matthews, R.A. Hendel, B.B. Epstein and G.M. Brosvic. 2002. Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. The Psychological Record 52(2): 187-201. DOI: 10.1007/BF03395423.

Kippel, G.M. 1975. Information feedback, need achievement and retention. The Journal of Educational Research 60(7): 256-261. DOI: 10.1080/00220671.1975.10884766.

Kulhavy, R.W. and R.C. Anderson. 1972. Delay-retention effect with multiple-choice tests. Journal of Educational Psychology 63(5): 505-512. DOI: 10.1037/h0033243.

Kulhavy, R.W., W.A. Stock, T.E. Hancock, L.K. Swindell and P.L. Hammrich. 1990. Written feedback: Response certitude and durability. Contemporary Educational Psychology 15(4): 319-332. DOI: 10.1016/0361-476X(90)90028-Y.

Newman, M.I., R.G. Williams and J.H. Hiller. 1974. Delay of information feedback in an applied setting: Effects of initially learned and unlearned items. The Journal of Experimental Education 42(4): 55-59. DOI: 10.1080/00220973. 1974.11011494.

O'Neill, M., R.A. Rasor and W.R. Bartz. 1976. Immediate retention of objective test answers as a function of feedback complexity. The Journal of Educational Research 70(2): 72-75. DOI: 10.1080/00220671.1976.10884955.

Phye, G.D. and T. Andre. 1989. Delayed retention effect: Attention, perseveration, or both? Contemporary Educational Psychology 14(2): 173-185. DOI: 10.1016/0361-476X(89)90035-0.

Pressey, S.L. 1926. A simple apparatus which gives tests and scores – and teaches. School and Society 23: 373-376.

Robin, A.L. 1978. The timing of feedback in personalized instruction. Journal of Personalized Instruction 3: 81-87.

Skinner, B.F. 1958. Teaching machines. Science 128(3330): 969-977. DOI: 10.1126/science.128.3330.969.

Surber, J.R. and R.C. Anderson. 1975. Delay-retention effect in natural classroom settings. Journal of Educational Psychology 67(2): 170-173. DOI: 10.1037/h0077003.