The Psychological Record, 2003, 53, 533-548

THE ROLE OF FEEDBACK DURING ACADEMIC TESTING: THE DELAY RETENTION EFFECT REVISITED

ROBERTA E. DIHOFF
Rowan University

GARY M. BROSVIC and MICHAEL L. EPSTEIN
Rider University

Students completed 5 quizzes during the semester using response formats that provided no feedback (word-processed answer sheet, Scantron form), delayed feedback (end of test, 24 hours), or immediate feedback while answering each test item. Administered 2 weeks after students had completed the 5th quiz, the final examination consisted of 50 items, with 10 items randomly selected from each quiz. Scores on each quiz, time to complete each quiz, and average study time per quiz did not differ as a function of response format. Students demonstrated the highest recall, the most accurate identification of initial responses, the most confidence in their answers, and the least amount of perseverative incorrect responding on those final examination items that were originally responded to when immediate feedback was provided. These same students demonstrated less recall, less identification accuracy, lower confidence in their answers, and more perseverative incorrect responding on those final examination items that were originally responded to when either end of test or delayed feedback had been provided. Students' self-reports assessing how response format affected learning, retention, and confidence were consistent with quantitative outcomes. The present results support prior demonstrations that combining immediate feedback with the opportunity to answer until correct not only assesses, but also teaches, in a manner that promotes the retention of course materials across the academic semester.

The introduction of the teaching-testing machine by Pressey (1926) prompted the development of numerous techniques through which immediate feedback during multiple-choice and alternative-choice questions could be delivered. The teaching machine described by Skinner (1958) not only presented immediate feedback, but also transformed the role of the student from a passive gatherer of information to an active demonstrator of skills and knowledge. Programmed instruction by teaching machines was, in part, intended to maintain vigilance during the testing process, the benefits of which have been

Address correspondence to Gary M. Brosvic, Department of Psychology, Rider University, Lawrenceville NJ 08648. (E-mail: Brosvic@Rider.edu).

demonstrated across a number of tasks (see Kritch & Bostow, 1998; Miller & Malott, 1997; Tudor, 1995). Information supplied during programmed teaching can range from partial to complete, but almost as a rule, the presentation of feedback has been both immediate and on an item-by-item basis. When partial feedback is provided, students are informed that responses are either correct or incorrect. When complete feedback is provided, corrective information is coupled with an answer-until-correct procedure. During recent years there has been a growing recognition that some conditions that promote performance during training interfere with retention, and that some conditions that interfere with performance during training promote retention (see Gick & Holyoak, 1987; Schmidt & Bjork, 1992).

In the early 1960s, Brackbill and her associates (Brackbill, Bravos, & Starr, 1962) demonstrated that delayed feedback across brief intervals promoted the retention of meaningful material. This outcome was also observed when feedback was delayed for 1 to 2 days and retention intervals were lengthened to 7 days (e.g., Kulhavy & Anderson, 1972; O'Neill, Rasor, & Bartz, 1976; Surber & Anderson, 1975). Proponents of delayed feedback generally adhere to the interference-perseveration hypothesis proposed by Kulhavy and Anderson (1972): Initial errors do not compete with to-be-learned correct responses if corrective information is delayed, because errors are likely to be forgotten and thus, they cannot interfere with retention. The superiority of delayed feedback, known as the delay-retention effect (DRE), was supported when Anderson and his associates compared the accuracy of responses on a retention test with the accuracy of responses on the initial test (e.g., Kulhavy & Anderson, 1972; Surber & Anderson, 1975). Although the delay-retention effect has not been supported in several studies (e.g., Kippel, 1974; Newman, Williams, & Hiller, 1974; Phye & Baller, 1970), delayed feedback has typically been as effective as immediate feedback.

Proponents of immediate feedback theorize that the earlier corrective information is provided, the more likely it is that efficient retention will result (Phye & Andre, 1989). The superiority of immediate feedback has been robustly demonstrated for the acquisition of verbal materials (Ammons, 1956) and motor skills (Anderson, Magill, & Seklya, 2001; Brosvic & Cohen, 1988), although Sassenruth (1972) contends that immediate feedback promotes proactive interference once participants commit themselves to an incorrect response. In theory, the amount of interference increases when participants must search repeatedly for a correct answer, and thus the amount of searching is a reasonable indicator that the learner neither knew the item initially nor acquired the correct response. However, Peeck and Tillman (1979) have presented convincing data that incorrect responses are not forgotten, and that they facilitate the acquisition of correct responses during feedback.

While there seems to be considerable agreement that feedback facilitates learning, there is little agreement as to what type of feedback is the most effective (Robin, 1978). Kulik and Kulik (1988) reported that

immediate feedback is more effective than delayed feedback for applied, but not laboratory, studies. Corrective feedback on objective examinations completed in the classroom, in the absence of computers, cannot be provided until the examination has been completed. In comparison, the conditions and equipment within the laboratory permit the immediate delivery of corrective feedback on an item-by-item basis. Until recently, the simple and practical means through which immediate feedback might be provided in the classroom in the absence of computers has not been available.

For the past several years, Epstein and his colleagues have been refining a simple yet elegant technique, the Immediate Feedback Assessment Technique, or IF AT (Epstein et al., 2002). The IF AT form (see Figure 1) is a multiple-choice answer sheet with rows of rectangular

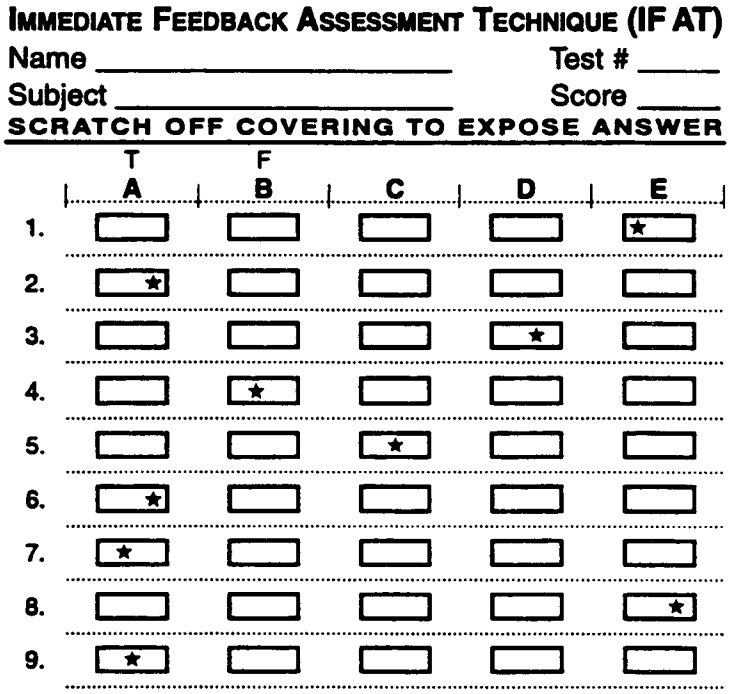

Figure 1. Sample portion of the Immediate Feedback Assessment Technique (IF AT) form. Patent is held by E3 Corporation.

answer spaces (e.g., A, B, C, D, E) that is nearly identical in layout to the ubiquitous machine-scored answer sheet available from Scantron Corporation. Participants scrape off an opaque, waxy coating covering 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 instant feedback that an incorrect choice was made. However,

rather than simply exiting the question, the student reviews the remaining response options, continues to respond until discovering the correct answer (a self-correction procedure), and thus exits each question with the correct answer. In the absence of such corrective feedback, the student would likely exit the question with the assumption that an incorrect response was actually correct; thus, a test procedure that does not employ immediate feedback likely promotes misconceptions rather than learning (Epstein et al., 2001).

In the present study we sought to examine the effects of varying delays in the provision of feedback on the retention of classroom materials presented during the academic semester. Feedback was provided after each response, at the end of the test, or after a 24-hr delay, and, for control purposes both a Scantron form and a word-processed answer sheet were used. Participants completed one semester quiz using each of the feedback and each of the control procedures, and then a final examination. These procedures enabled determination of the effects of varying delays in feedback across retention intervals of 2 to 10 weeks. After the final examination, the ability of participants to identify initial correct responses and errors was evaluated in a test of the interference-perseveration hypothesis.

Methods

Participants

Participants included 33 male and 62 female students enrolled in an undergraduate psychology course.

Materials

Six tests were prepared from a publisher-supplied test bank, each with 50 questions, and each question with four response options (i.e., A, B, C, D). Five of the examinations were administered during the semester and the sixth, the final examination, was cumulative and included 10 items randomly selected from each of the five quizzes completed during the semester.

Design and Procedures

A latin squares design was used so that participants were evenly distributed across the response formats and proceeded through them in a counterbalanced order. On the day after each quiz, all participants returned to the classroom during an open period for 30 minutes. During this time participants in the delayed feedback group reviewed the examination, the correct solutions, and their corrected answer sheets for 30 minutes. Participants in the other conditions were requested to read non-course materials until the end of the test period, and during the review process, all participants were required to remain silent and were not permitted to share their materials.

In the first control condition (traditional), answers were recorded with a pencil using a traditional word-processor generated answer sheet. In

the second control group (Scantron), answers were recorded with pencil on a Scantron form. In the end-of-test-feedback condition, answers were recorded with pencil on a Scantron form, and upon completion of the test, all writing implements were removed and participants were permitted to review the examination, the correct solutions, and their answer sheets for 30 minutes. Participants in the other conditions were requested to remain seated and to read non-course materials until the end of the test period. During this review process all participants were required to remain silent and were not permitted to share their materials. In the delayed-feedback condition, participants recorded their answers in pencil on a Scantron form and, on the following day, these participants reviewed the examination, the correct solutions, and their corrected answer sheets for 30 minutes. Participants in the other conditions were requested to remain seated and to read non-course materials until the end of the test period. During this review process all participants were required to remain silent and were not permitted to share their materials. In the immediate-feedback condition, answers were recorded using the IF AT form (E3 Corporation) which enabled participants to receive immediate affirming or corrective feedback and to respond until discovering the correct answer. The final examination was completed 2 weeks after the fifth quiz, and at the time, all participants used the word-processed answer sheet to record their answers. Scores on the final examination served as the primary measure of recall. After the final examination, participants reviewed each examination item and identified their initial responses, both correct and incorrect, and then rated confidence in the accuracy of their identifications on a 100-point scale ranging from 0 (no confidence) to 100 (complete confidence). Participants then completed a six-item questionnaire assessing satisfaction with and the relative benefits of the five-response formats; and for each item, they identified which of the five-response methods they preferred.

Results

Although the IF AT method enables the assignment of partial credit (i.e., correct responding on the first attempt is assigned 100% of item credit whereas responding on the second, third, or fourth attempt could be assigned reduced percentages according to instructor discretion), this procedure was not used, and the results described below were based upon the accuracy of initial responses. There were no differences in any measure as a function of sex of participant, all $F < 1$, all $p > .5$; and also no differences in the amount of self-reported study time as a function of quiz, response format, or their interaction, all $F < .57$, all $p > .33$. Scores on the five semester quizzes did not differ as a function of quiz, response format, or their interaction, all $F < 1.14$, all $p > .16$. The average number of times that participants changed their first response across the five semester quizzes differed only as a function of response format, $F(4, 360) = 4.14$, $p = .004$, as seen in Figure 2. Scheffé comparisons indicated that

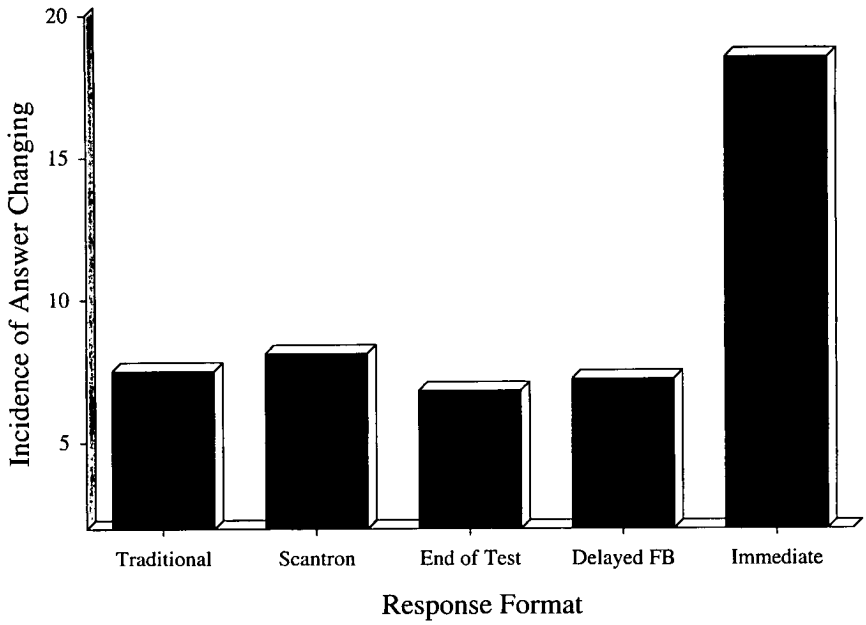

Figure 2. Number of answer changes on an initial exam as a function of response format.

participants receiving immediate feedback made the most changes, $p < .001$. Thus, performance on the quizzes was not affected by response format, and this outcome was anticipated because the beneficial effects of corrective information should not be observed until the final examination which served as the primary measure of retention.

Recall scores, averaged for the 10 items randomly selected from each of the five semester quizzes and used on the final examination, are presented in Figure 3 as a function of answer format. Recall scores differed only as a function of response format, $F(4, 360) = 120.25$, $p < .0001$. Scheffé comparisons indicated that recall scores were higher (a) for questions for which immediate feedback was provided and (b) when either end-of-test or delayed feedback was provided rather than when either a traditional word-processed answer form or a Scantron form was provided, all $p < .001$. These results demonstrate that provision of immediate feedback promoted a higher level of recall than provision of feedback that was delayed until either after a test or for 24 hours.

The mechanisms underlying the DRE appear to be related to the general beneficial effects of feedback, such as the correction of previously inaccurate assumptions and the reduction of inaccurate perseverative responding (Kulhavy & Anderson, 1972; Surber & Anderson, 1975). These two putative mechanisms were evaluated by a further analysis of responses to the items on the final examination. After completing the final examination each participant was requested to

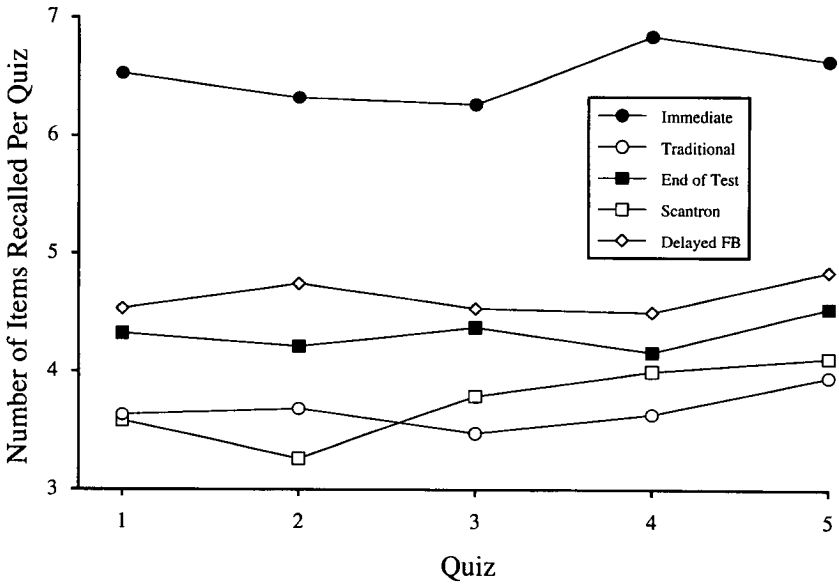

Figure 3. Mean recall scores on the five examinations as a function of response format.

identify those final examination items that they recalled answering incorrectly on the initial examination, and then to report the degree of confidence in their identifications. These identifications could be made only when feedback was provided, and thus participants' responses when the traditional and Scantron forms were provided are not included.

The percentage of items correctly identified by participants as having been answered incorrectly on their initial administration is presented in Figure 4 as a function of the response format provided when the items were initially administered. These percentages differed only as a function of response format, $F(4, 360) = 40.09$, $p < .0001$. Scheffé comparisons indicated the percentage of correctly identified initial errors was (a) highest when immediate feedback was provided and (b) higher when delayed rather than the end-of-test feedback was provided, all $p < .001$. The percentage of correctly identified initial responses is presented in Figure 5 as a function of response format when the items were initially administered. These percentages differed only as a function of response format, $F(4, 360) = 58.01$, $p < .0001$. Scheffé comparisons indicated the percentage of correctly identified initial responses was (a) highest when immediate feedback was provided and (b) higher when delayed rather than end-of-test feedback was provided, all $p < .001$.

Confidence ratings for identifications are presented in Figure 6 as a function of response format. Confidence levels differed only as a function of response format, $F(4, 360) = 5.92$, $p = .0003$. Scheffé comparisons indicated that confidence levels were highest when immediate feedback was provided, $p < .001$. The results of these analyses indicate that the

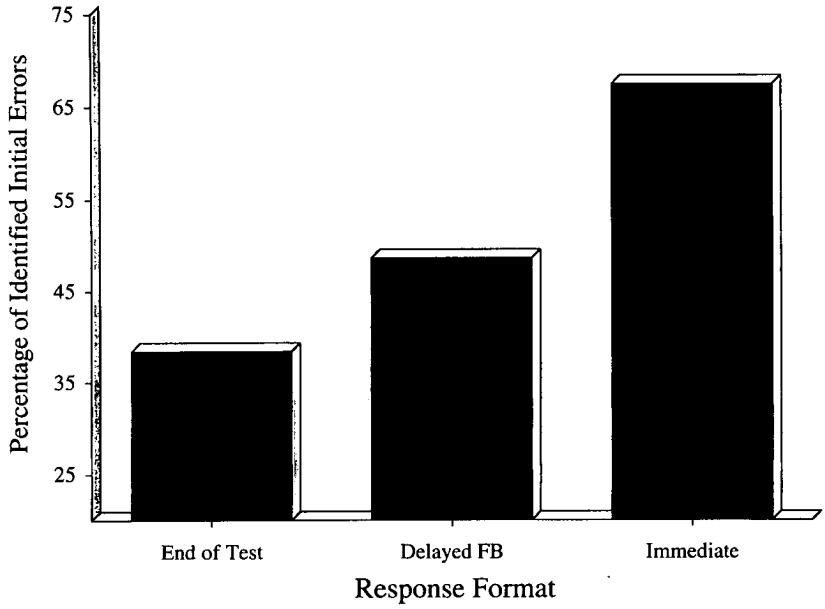

Figure 4. Percentage of items correctly identified by participants as having been answered incorrectly when immediate feedback, end-of-test feedback, or delayed feedback was provided.

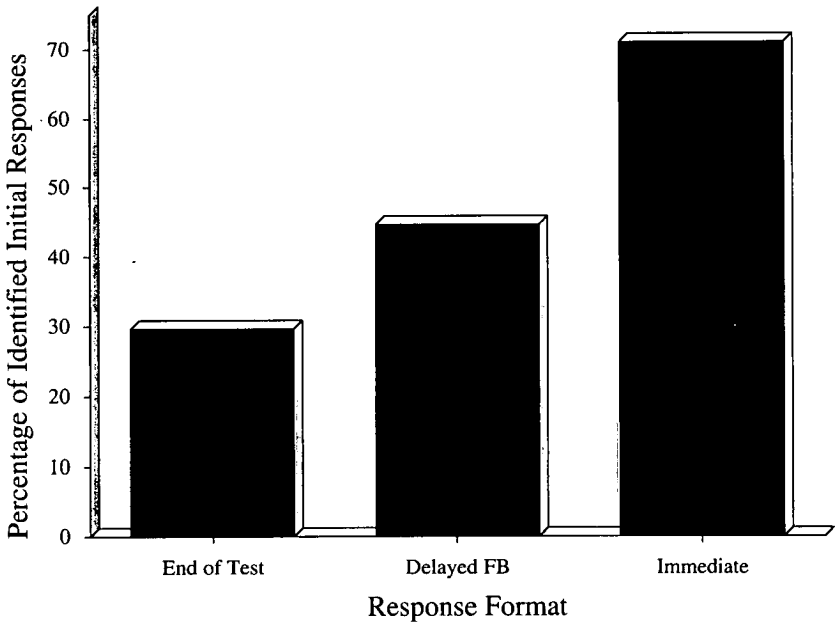

Figure 5. Percentage of correctly identified initial responses on the initial administration when immediate feedback, end-of-test feedback, or delayed feedback was provided.

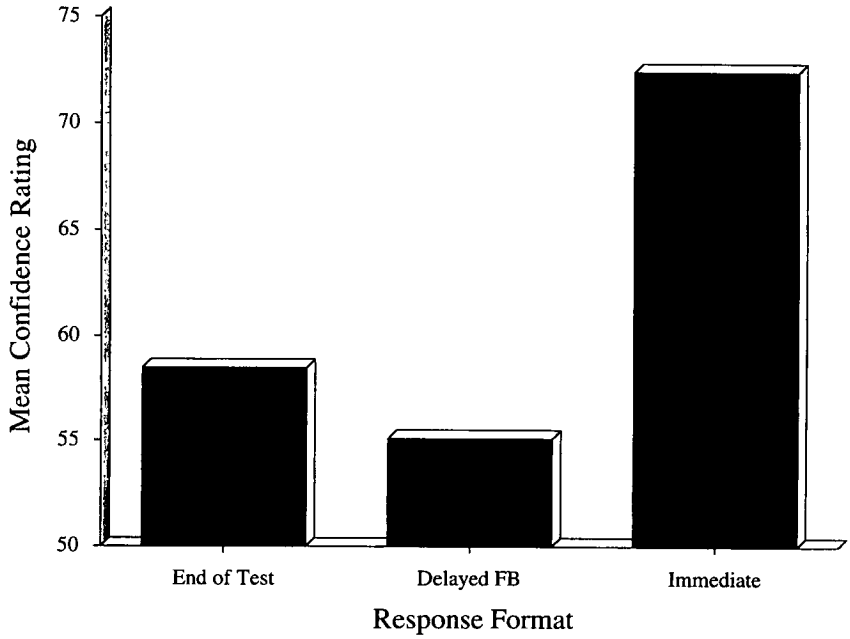

Figure 6. Mean confidence ratings when identifying initial responses on the final examination as a function of response format.

provision of immediate feedback enabled participants to recall more of their initial responses and to be more confident in the identification of their responses. These outcomes were significantly less robust for the same participants when either end-of-test or delayed feedback was provided.

Reductions in inaccurate perseverative responding were evaluated by determining the conditional probabilities of correct responding on the second (final examination) and the first (initial quiz) administration of each item. A review of the conditional probabilities presented in Table 1

Table 1

Conditional Probability (in percentages) of Final Examination Outcomes Given Initial Test Outcomes By Response Format

	Traditional Form	Scantron Form	End-of-Test Feedback	Delayed Feedback	Immediate Feedback
Correct on Final/ Correct on Initial	38.52	35.85	50.31	53.79	73.25
Correct on Final/ Incorrect on Initial	21.98	22.81	29.07	32.86	56.67
Incorrect on Final/ Correct on Initial	61.48	64.15	49.69	46.21	26.75
Incorrect on Final/ Incorrect on Initial	78.02	77.19	70.93	67.14	43.33

suggests that the hypothesis that inaccurate perseverative responding is reduced by delayed feedback cannot be supported. The likelihood of responding correctly on the first and second administration of an item differed only as a function of response format, $F(4, 360) = 12.56$, $p < .0001$. Scheffé comparisons indicated that the value of this conditional probability was (a) highest when immediate feedback was provided and (b) higher when either end-of-test or delayed feedback rather than either a traditional or a Scantron form was provided, all $p < .001$. The likelihood of responding correctly on the second administration of an item after having responded incorrectly on its initial administration differed only as a function of response format, $F(4, 360) = 8.07$, $p < .0001$. Scheffé comparisons indicated that the value of this conditional probability was (a) highest when immediate feedback was provided and (b) higher when either end-of-test or delayed feedback rather than when a traditional form or Scantron form was provided, all $p < .001$. The likelihood of responding incorrectly on the second administration of an item after having responded correctly on its initial examination administration differed only as a function of response format, $F(4, 360) = 21.87$, $p < .0001$. Scheffé comparisons indicated that the value of this conditional probability was (a) highest when either a traditional or a Scantron form was provided and (b) higher when either end-of-test or delayed feedback rather than immediate feedback was provided, all $p < .001$. The likelihood of responding incorrectly to the same item on both administrations differed only as a function of response format, $F(4, 360) = 15.54$, $p < .0001$. Scheffé comparisons indicated that the value of this conditional probability was higher when a traditional form, a Scantron form, or end-of-test feedback rather than immediate feedback was provided, all $p < .001$.

The hypothesis that delayed feedback reduces inaccurate perseverative responding was further examined by secondary review of the conditional probabilities of responding incorrectly on both administrations of an item. Responses on the second administration of these items were dichotomized into the categories of repeating the same incorrect answer selection (see Figure 7) or making a different but also incorrect answer. The percentages for each category differed only as a function of response format, all $F(4, 360) > 31.78$, all $p < .0001$. Scheffé comparisons indicated that the percentage of participants repeating the same incorrect response was (a) lowest when immediate feedback was provided, (b) lower when delayed rather than end-of-test feedback was provided, and (c) lower when either end-of-test or delayed feedback rather than a traditional or a Scantron form was provided, all $p < .001$.

Participants' evaluations of each response format were expressed on a brief questionnaire that assessed overall perceptions of each response format upon the conclusion of the final examination. The results of simple contingency table analyses (see Table 2) indicated that immediate feedback was the response format that was preferred, promoted the most learning and retention, facilitated the most involvement in the testing process, and corrected the most initially inaccurate assumptions, all $\chi^2(4) > 66.84$, all $p < .0001$. The traditional answer form was rated as the easiest to complete, all $\chi^2(4) > 24.16$, all $p < .001$.

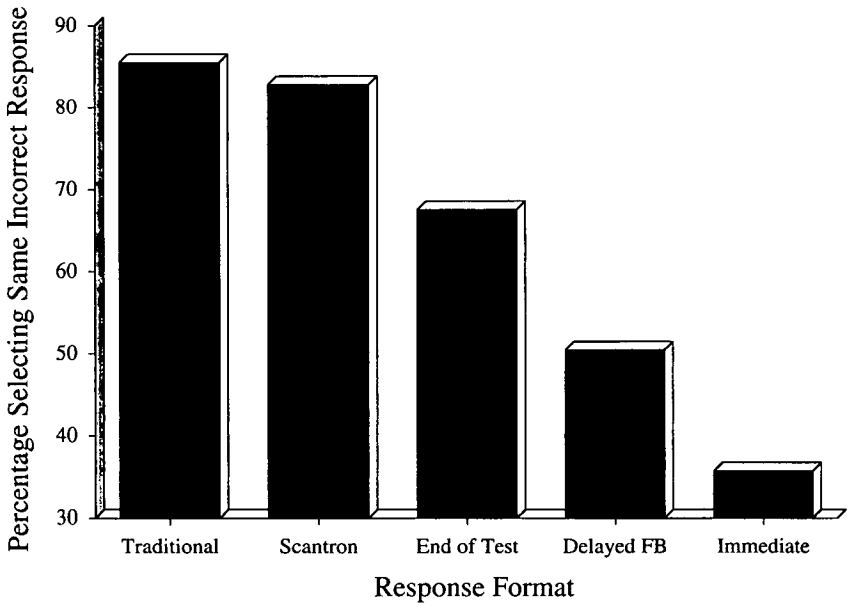

Figure 7. Percentage of participants selecting the same incorrect answer as a function of response format.

Table 2

Posttest Measures Assessing Perceptions of Each Response Format

	Traditional Form *n*	Scantron Form *n*	End-of-Test Feedback *n*	Delayed Feedback *n*	Immediate Feedback *n*	Chi Square
Preferred Response Format	5	6	8	12	64	134.73
Promotes Learning	7	5	9	14	60	112.95
Promotes Retention	5	5	8	12	65	140.95
Involvement in Testing Process	7	9	11	18	50	66.84
Corrects Initially Incorrect Assumptions	8	6	12	18	51	71.79
Ease of Completion	35	25	7	5	23	34.11

Discussion

The within-subjects design of the present study enabled evaluation of the performance of the same participants across five response formats that varied in the presentation and timing of corrective feedback. The response format that promoted greatest retention, greatest confidence, and greatest accuracy at identifying initial responses (correct and incorrect) was immediate feedback coupled with the opportunity to answer-until-correct; these results are consistent with participants' self-reports of the benefits of this response format. This combination, immediate feedback and answering-until-correct, is typically conducted only in laboratory studies; however, the Immediate Feedback Assessment Technique (IF AT) now enables this combination for regular classroom assessments that include multiple-choice and alternative-choice questions.

The IF AT embodies the theoretical and practical foundations of the teaching-testing machines described by Pressey (1926) and Skinner (1958), and it transforms the passive gatherer of information into the active demonstrator of skills and knowledge. In recent studies, Epstein and his colleagues reported that the IF AT form was easily used by (a) students without known learning difficulties in first to eighth grade (Epstein et al., 2002), through college (Epstein et al., 2001, 2002), (b) junior high school students classified with mild mental retardation to acquire multiplication facts (Epstein et al., 2002), (c) junior high school students (Epstein, Brosvic, Dihoff, Lazarus, & Costner, 2003) classified with mild mental retardation to acquire life-skill materials, and (d) preschool children with developmental delays studying academic readiness materials (Epstein et al., 2003). In each of these studies, the provision of immediate corrective feedback promoted greater retention and a greater correction of initially inaccurate strategies than when the identical stimuli were completed in the absence of immediate feedback.

The results observed when the IF AT was used to provide immediate feedback 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), the use or nonuse of an answer-until-correct process, partial (50% of test items) to complete feedback, and stimulus materials (fictional materials to classroom concepts). In spite of these considerable differences, the provision of immediate feedback improved the formation of classes and class relations by undergraduates (Adams & Fields, 1999), increased undergraduates' knowledge of biological concepts presented using the PLATO system (Dempsey & Litchfield, 1993), enhanced the test-taking performance of fifth graders (Hanna, 1976) and seventh and eighth grade students completing an achievement test (Hanna & Long, 1979), and increased the general course performance of junior high school and college students (Beeson, 1973). In related studies, immediate and delayed feedback have been reported to be

equally effective (see Robin, 1978). The concerns noted above for the considerable variation in procedures employed in studies including immediate feedback are equally applicable to studies including delayed feedback. Despite these concerns, immediate and delayed feedback have increased the class performance of graduate students (Beck & Lindsey, 1979), the performance of college students on a general ability test (O'Neill et al., 1976), and accuracy during acquisition (Anderson et al., 2001). Similarly, Peeck and Tillman (1979) reported that permitting fifth grade children to review their original answer sheets influenced neither the effectiveness of feedback nor their accuracy at identifying initial responses at the time of retesting.

The Peeck and Tillman (1979) results do not support the interference-perseveration hypothesis position that the forgetting of an initial response is a prerequisite to benefit from feedback, and this outcome is at odds with results reported in some studies in which delayed feedback was more effective than immediate feedback. In these latter studies, the forgetting of initial mistakes coupled with increased attention to test items reviewed during delayed feedback have been presented as two of the causal mechanisms responsible for enhancing the test performance of high school students (Kulhavy & Anderson, 1972; Surber & Anderson, 1975), enhancing the formation, attainment, and transfer of concepts by undergraduates (Schroth, 1995), increasing the performance of undergraduates on a general information test (Webb, Stock, & McCarthy, 1994), and enhancing the multiple-choice performance of fifth graders (Peeck, van Den Bosch, & Kreupeling, 1985). However, participants in the present study were more likely to identify their initial responses, both correct and incorrect, when immediate feedback was provided. Indeed, participants were significantly more accurate at identifying their initial responses when immediate feedback was provided than when either end-of-test or delayed feedback was provided.

It is possible that the corrective information provided by immediate feedback might deter participants from committing to a response, and instead, might create a state of disequilibrium that is resolved by assimilation of the correct response. This hypothesis is supported by the secondary analysis of the conditional probabilities of responding incorrectly on the first and second administrations of test items. Recall that responses on the second administration were dichotomized into the categories of selecting the same incorrect answer and selecting a different but also incorrect answer, and that the percentage of participants making the same incorrect response was lowest when immediate feedback was provided. Another alternative explanation is that immediate feedback increases the depth at which corrective information is processed, as demonstrated by Lhyle and Kulhavy (1987). In that study, students were presented with feedback represented by either intact words or words in scrambled format (Study 1), or in either scrambled or unscrambled format with the requirement that the word was to be either unscrambled or written on a response sheet (Study 2). The results of

each study were consistent with a levels-of-processing effect, and the answer-until-correct process included in the present study likely produced an analogous effect.

When compared to the other four response formats included in the present study, immediate feedback coupled with an answer-until-correct process requires the student to review the question, to consider why an initial response was incorrect, to review the remaining options, to discriminate between the remaining answers, and to continue to respond until the correct response is discovered. The active nature of this process is analogous to a level-of-processing task, and it is likely that the variety of activities during encoding promotes the correction of initially inaccurate strategies. A test of this latter hypothesis is currently in progress in our laboratory in which one half of the students were provided with IF AT forms to use with practice tests prior to completing in-class examinations while the other half were provided with practice tests and the traditional (i.e., word-processed) answer form used in the present study. These procedures were reversed on every other test, and the results of preliminary analyses indicate that the performance of students was increased only when IF AT forms were available. It appears that the immediate feedback provided several days before an examination promoted the same retention benefits on classroom examinations that have been observed in our prior studies (Epstein et al., 2001, 2002). This outcome suggests that the proactive use of immediate feedback enables the selection of correct responses during future assessment situations, and we are currently examining if the integration of immediate feedback via the IF AT during training sessions (e.g., social skills for adolescents and adults, discrete-trial training for paraprofessionals) enables participants to make more appropriate responses.

Collectively, the present results demonstrate that immediate feedback promotes recall, the most accurate identification of initial responses, increases confidence in answers, and reduces perseverative incorrect responding. Considerably less robust outcomes were observed when feedback was provided at either the end of a test or after a 24-hour delay, although as noted in prior studies, both forms of delayed feedback promoted more learning than when control procedures were used. These findings are of particular interest because these comparisons were made within subjects, and after experience with each response format. Immediate feedback coupled with an answer-until-correct process was not only the most effective but also the most preferred. Until recently, this combination was rarely available in the classroom, but since the development and validation of the IF AT, immediate, self-corrective feedback is now readily available.

References

ADAMS, B. J., & FIELDS, L. (1999). Effects of unreinforced conditional selection trainings, multiple negative comparison training, and feedback on equivalence class formation. *The Psychological Record*, 49, 685-702.

ANDERSON, D. I., MAGILL, R. A., & SEKLYA, H. (2001). Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. *Journal of Motor Behavior*, 33, 59-66.

AMMONS, R. B. (1956). Effects of knowledge of performance: A survey of theoretical formulation. *Journal of General Psychology*, 54, 279-299.

BECK, F. W., & LINDSEY, J. D. (1979). Effects of immediate information feedback and delayed information feedback on delayed retention. *Journal of Educational Research*, 72, 283-284.

BEESON, R. O. (1973). Immediate knowledge of results and test performance. *The Journal of Educational Research*, 66, 224-226.

BRACKBILL, Y., BRAVOS, A., & STARR, R. H. (1962). Delay-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology*, 55, 947-952.

BROSVIC, G. M., & COHEN, B. (1988). The horizontal-vertical illusion and knowledge of results. *Perceptual and Motor Skills*, 67, 463-469.

DEMPSEY, J. V., & LITCHFIELD, B. C. (1993). Feedback, retention, discrimination error, and feedback study time. *Journal of Research on Computing in Education*, 25, 303-326.

EPSTEIN, M. L., BROSVIC, G. M., DIHOFF, R. E., LAZARUS, A. D., & COSTNER, K. L. (2003). Effectiveness of feedback during the testing of preschool children, elementary school children, and adolescents with developmental delays. *The Psychological Record*, 53, 177-195.

EPSTEIN, M. L., EPSTEIN, B. B., & BROSVIC, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports*, 88, 889-894.

EPSTEIN, M. L., LAZARUS, A. D., CALVANO, T. B., MATTHEWS, K. A., HENDEL, R. A., EPSTEIN, B. B., & BROSVIC, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52, 187-201.

GICK, M. L., & HOLYOAK, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications*. California: Academic Press.

HANNA, G. H. (1976). Effects of total and partial feedback in multiple-choice testing upon learning. *The Journal of Educational Research*, 69, 202-205.

HANNA, G. M., & LONG, C. A. (1979). Effect of answer until correct testing on reliability. *Perceptual and Motor Skills*, 49, 464-466.

KIPPEL, G. M. (1974). Information feedback, need achievement and retention. *The Journal of Educational Research*, 60, 256-260.

KRITCH, K. M., & BOSTOW, D. E. (1998). Degree of constructed-response interaction in computer-based programmed instruction. *Journal of Applied Behavior Analysis*, 31, 387-398.

KULHAVY, R. W., & ANDERSON, R. C. (1972). Delay-retention effects with multiple choice tests. *Journal of Educational Psychology*, 63, 505-512.

KULIK, J. A., & KULIK, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58, 79-97.

LHYLE, K. G., & KULHAVY, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology*, 79, 320-322.

MILLER, M. L., & MALOTT, R. W. (1997). The importance of overt responding in programmed instruction even with added incentives for learning. *Journal of Behavioral Education*, 7, 497-503.

NEWMAN, M. I., WILLIAMS, R. G., & HILLER, J. H. (1974). Delay of information feedback in an applied setting: Effects of initially learned and unlearned items. *Journal of Experimental Education*, 42, 55-59.

O'NEILL, M., RASOR, R. A., & BARTZ, W. R. (1976). Immediate retention of objective test answers as a function of feedback complexity. *The Journal of Educational Research*, 70, 72-75.

PEECK, J., & TILLMAN, H. H. (1979). Learning from feedback: Comparison of two feedback procedures in a classroom setting. *Perceptual and Motor Skills*, 48, 351-354.

PEECK, J., VAN DEN BOSCH, A. B., & KREUPELING, W. J. (1985). Effects of informative feedback in relation to retention of initial responses. *Contemporary Educational Psychology*, 10, 303-313.

PHYE, G. D., & ANDRE, T. (1989). Delay-retention effect: Attention, perseveration, or both? *Contemporary Educational Psychology*, 14, 173-185.

PHYE, G., & BALLER, W. (1970). Verbal retention as a function of informativeness and delay of informative feedback. *Journal of Educational Psychology*, 61, 380-381.

PRESSEY, S. L. (1926). A simple device 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.

SASSENRUTH, J. M. (1972). Effects of delay of feedback and length of post feedback interval on retention of prose material. *Psychology in the Schools*, 9, 194-197.

SCHMIDT, R. A., & BJORK, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207-217.

SCHROTH, M. L. (1995). Variable delay of feedback procedures and subsequent concept formation transfer. *The Journal of Educational Psychology*, 122, 393-399.

SKINNER, B. F. (1958). Teaching machines. *Science*, 128, 969-977.

SURBER, J. R., & ANDERSON, A. C. (1975). Delay-retention effect in natural classroom settings. *Journal of Educational Psychology*, 67, 170-173.

TUDOR, R. M. (1995). Isolating the effects of active responding in computer-based instruction. *Journal of Applied Behavior Analysis*, 28, 343-344.

WEBB, J. M., STOCK, W. A., & MCCARTHY, M. T. (1994). The effects of feedback timing on learning facts: The role of response confidence. *Contemporary Educational Psychology*, 19, 251-265.