The Psychological Record, 2005, 55, 401-418

EFFICACY OF ERROR FOR THE CORRECTION OF INITIALLY INCORRECT ASSUMPTIONS AND OF FEEDBACK FOR THE AFFIRMATION OF CORRECT RESPONDING: LEARNING IN THE CLASSROOM

GARY M. BROSVIC, MICHAEL L. EPSTEIN, and MICHAEL J. COOK
Rider University

ROBERTA E. DIHOFF
Rowan University

Participants completed 5 classroom examinations during which the timing of knowledge of results (no feedback: Scantron form; delayed feedback: end-of-test, 24 hour delay; immediate feedback: educator, response form) and iterative responding (1 response, up to 4 responses) were manipulated. At the end of the semester, each participant completed a 100-item final examination which included 10 items randomly selected from each classroom examination, plus 50 entirely new items. Neither the source of feedback nor the number of responses permitted influenced performance on classroom examinations but both factors interacted significantly to enhance the final examination performance of participants provided with immediate feedback and iterative responding. The correction of initially inaccurate strategies by combining immediate feedback with iterative responding was not differentially effective as a function of information source: educator or the Immediate Feedback Assessment Technique (IF AT) form. For these participants, response identification accuracy, confidence ratings, and retention were higher and inaccurate perseverative responding was lower. Performance on the final examination permits the preliminary quantification of how immediate feedback coupled with iterative responding, when used during classroom examinations that contain items that will be repeated on a cumulative final examination, not only assesses student knowledge but also teaches in a manner that promotes the retention of course materials.

Many of the earliest studies conducted in the psychological sciences were dedicated to examining changes in the performance of learners provided with information (i.e., feedback) that either affirmed a correct response or corrected an error (e.g., Thorndike, 1913, 1927). This corrective information, viewed initially within an associationistic framework as a contingent event, strengthened correct responses

Address correspondence to Gary M. Brosvic, Department of Psychology, Rider University, Lawrenceville, NJ 08343. (E-mail: Brosvic@Rider.edu).

through positive reinforcement and weakened incorrect responses through nonreinforcement. This mechanistic perspective emphasized minimizing errors, but provided neither the means for, nor an insight into, their correction. Indeed, the efficacy of errors and their inclusion within the learning processes did not enter mainstream psychology until the ascendance of information processing within which errors were seen as valuable resources that learners could use to evaluate their understanding of test materials and to correct their initially inaccurate assumptions (e.g., Kluger & DeNisi, 1996; Kulhavy & Stock, 1989).

One of the most divisive points in the literature on feedback concerns the optimal timing of feedback following the learner's response: immediately, or after the elapse of a prescribed interval. The definitions of immediate feedback range from seconds after a response (Epstein, Epstein, & Brosvic, 2001) to the next weekly meeting of a class (Robin, 1978). Definitions of delayed feedback range from the end of a test (Dihoff, Brosvic, & Epstein, 2003, 2004) to delays of 7 days (Bruning, Schraw, & Ronning, 1999; Robin, 1978). Even though the considerable overlap of these definitions precludes the direct numerical comparison of experimental outcomes, the rationales advanced for preferring either immediate or delayed feedback are more exclusive than the intervals that define the timing of feedback. Proponents of immediate feedback recommend the correction of errant responses and the acquisition of the correct response before exiting a test problem or test session (Epstein et al, 2001). In comparison, proponents of delayed feedback recommend the imposition of a delay of 24 to 48 hours to facilitate the forgetting of errant responses and the acquisition of correct responses in the absence of the interference that immediate feedback on an item-by-item basis generates (Kulhavy & Stock, 1989). Reports that immediate feedback is more effective than delayed feedback during applied, but not during laboratory studies (Gick & Holyoak, 1987) perhaps accounts for why the literature on the effects of the timing of feedback, shows approximately an even number of reports for the superiority of immediate over delayed feedback, and for delayed over immediate feedback.

The classroom is the one testing environment that has been included in both applied and laboratory studies, and in which immediate and delayed feedback have been directly compared. Typically, the provision of immediate feedback on examinations completed in the classroom has required the use of either computer-assisted instruction (CAI) or the assignment of one evaluator per learner. In the absence of computers, course-specific software, or programming skills, the delivery of feedback is delayed—if it is provided at all. In the laboratory, the use of CAI enables the presentation of supporting materials, branching for additional instruction and assessment, and the measurement of response times to index the processing of test materials. Until recently, the simple and practical means by which immediate feedback might be provided in the classroom in the absence of computers has not been available. The tool that has been refined and validated in our classrooms and laboratories is the Immediate Feedback Assessment Technique, or IF AT[1] (Epstein et al., 2001, 2002).

The IF AT manifests the theoretical and practical foundations of the teaching-testing machines described by Pressey (1950) and Skinner (1968), transforming the passive receiver of information into the active demonstrator of skills and knowledge. The IF AT form (see Figure 1) 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 ubiquitous machine-scored answer sheet

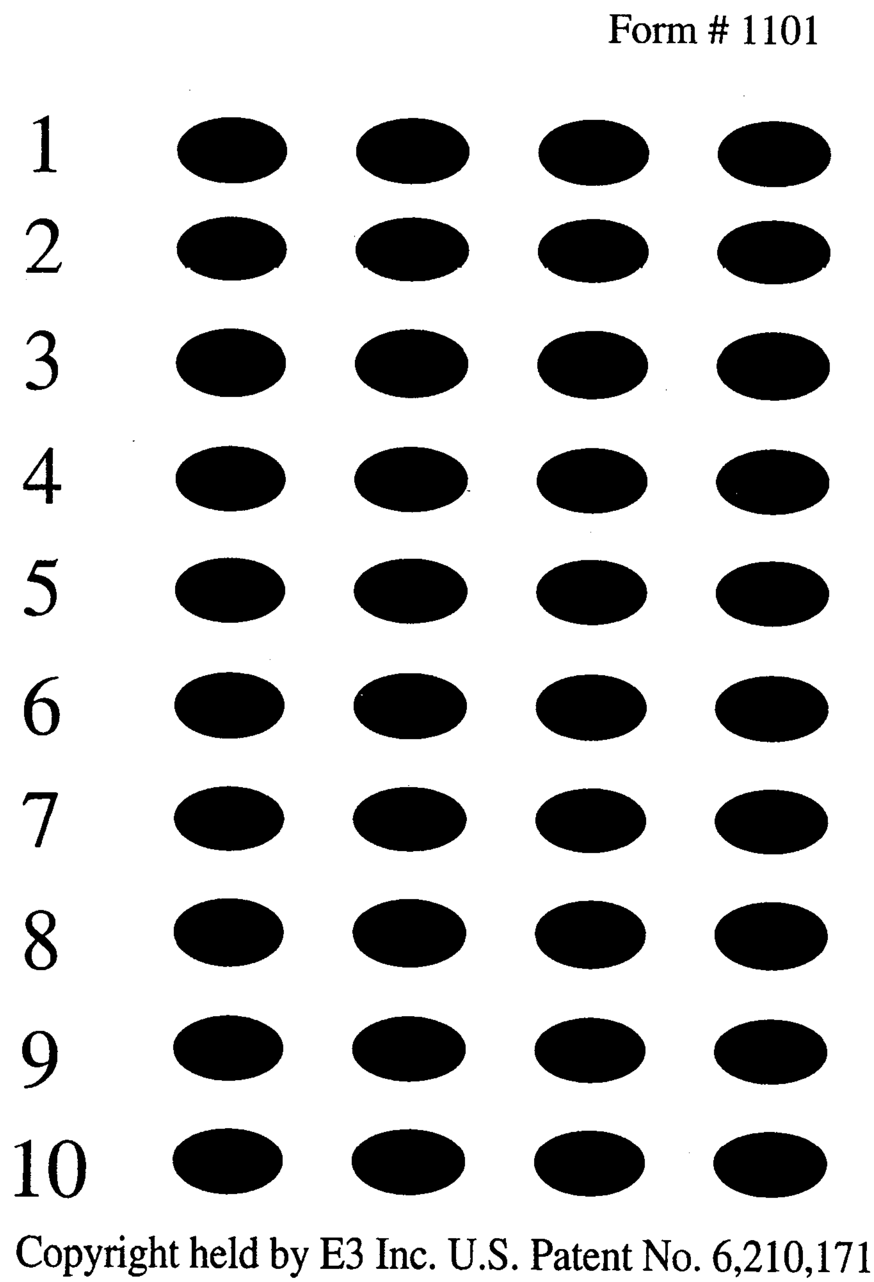

Figure 1. Sample portion of the Immediate Feedback Assessment Technique (IF AT) form. Patent is held by E3 Corporation.

[1]IF AT is a registered trademark of E3 Corporation.

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 the correct answer is discovered (a self-correction procedure), and thus exits each question knowing the correct answer.

Indeed, without corrective feedback, the learner likely exits an examination assuming that incorrect responses were actually correct; thus, an examination that does not employ feedback may promote misconceptions (e.g., Brown, Schilling, & Hockensmith, 1999). In comparison, the IF AT is an effective tool for promoting the correction of initially inaccurate response strategies and the retention of learning, and its use has been shown to facilitate the learning and retention of participants with developmental delays (Epstein, Brosvic, Dihoff, Lazarus, & Costner, 2003). In recent studies we found that the IF AT form could be easily used by preschool children with developmental delays studying academic readiness materials (provided sufficient motor skill was present), by junior high school students classified with mild mental retardation studying a life skills curriculum (Epstein et al., 2003), and by elementary school students classified with mild mental retardation acquiring mathematical fact series (Dihoff Brosvic, Epstein, & Cook, 2005), with more favorable outcomes observed when feedback was provided by an educator than by the IF AT form. The concurrent presentation of both sources of immediate feedback (educator and the IF AT) was more effective than the presentation of either source separately, and thus the inclusion of the IF AT facilitated the teaching-learning process to promote higher levels of independent learning and retention. These outcomes raise several interesting practical and theoretical issues for learning and retention that require further study and form the bases of the present study: the timing of feedback, the source of feedback, and the opportunity to engage in iterative responding (i.e., to answer until correct).

Method

Participants

Thirty-one male and 79 female students enrolled in a liberal arts and sciences undergraduate course served as voluntary participants. The modal participant was a Caucasian female, approximately 20 years of age and majoring in the liberal arts and sciences.

Materials

Five classroom examinations 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). The final examination consisted of 100 items which included 10 randomly selected items from each classroom examination, plus 50 entirely new items.

Design and Procedures

At least 11 participants (3 male and 8 female) were randomly assigned to each experimental group generated from the factorial combination of five feedback groups (Scantron, Delayed, End-of-Test, Immediate/Educator, and Immediate/Form) and two levels of iterative responding (a maximum of one response permitted, a maximum of four responses permitted). Prior to recording an answer for each test item, participants reviewed the four answers that were provided, and from those answers, recorded the number of answers that they determined could be correct; participants then selected the "best" answer and rated confidence in their selection on a scale ranging from 1 (no confidence) to 100 (complete confidence). This process was repeated, in the event of either an initial or subsequent incorrect response when iterative responding was permitted, for each test item.

In the no-feedback control group, answers were recorded with a pencil on a Scantron form. In the end-of-test-feedback group, answers were recorded with a pencil on a Scantron form, and upon the 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 groups were requested to remain seated and to read noncourse materials until the end of the test period under the supervision of test monitors. As with the aforementioned condition during the review process, all participants were required to remain silent and were permitted neither to share their materials nor to ask any questions of the test monitors. In the delayed-feedback group, 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 groups were requested to remain seated and to read noncourse materials until the end of the test period. During this review process all participants were required to remain silent and were permitted neither to share their materials nor to ask any questions of the test monitors. Participants in the immediate/educator group were seated in pairs, and an experimental assistant sat between the members of each pair. After the participants recorded their answers on Scantron forms, the experimental assistant indicated by holding up 3- x 5-inch index cards if the response was correct; if an incorrect response was made and iterative responding was available, additional cards identifying already selected responses (e.g., A, B) were made visible to each participant who continued to select responses until the correct answer was discovered. In the immediate/form group, answers were recorded using the IF AT form (E3 Corporation) which enabled participants to receive immediate affirming or corrective feedback; if appropriate to the experimental group the participant was permitted to continue selecting responses until the correct answer was discovered.

After completing the fifth classroom examination, each participant then completed a 15-item questionnaire assessing ease of understanding and ease of completing response requirements, perceived fairness of and preference for an answer-until-correct procedure, and involvement in the test-taking process, as described previously by Epstein and Brosvic (2002). The final examination was administered 1 week after completion of the fifth classroom examination, and at the time, all participants used Scantron forms to record their answers. Once the final examination was completed, participants reviewed each examination item and identified those items they believed were repeated from one of the classroom examinations, and their initial responses to those items, both correct and incorrect, and then rated confidence in the accuracy of their identifications on a scale ranging from 1 (no confidence) to 100 (complete confidence). Performance on the items carried over from classroom examinations to the final examination served as the primary measure of retention. 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 correct responding on the second, third, or fourth attempt may 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.

Results

There were no differences in SAT scores, current semester classroom performance, overall GPA, or any other dependent measure described below as a function of sex of participant, feedback group, the opportunity to engage in iterative responding (IR), or their interaction, all $F < 1$, all $p > .5$. There were no differences in any dependent measure between the delayed-feedback and the end-of-test-feedback groups, all $F < 1$, all $p > .5$; responses were aggregated and hereafter referred to as delayed feedback. A similar lack of differences between the immediate/educator and the immediate/form groups was observed, all $F < 1$, all $p > .5$; responses were aggregated and hereafter referred to as immediate feedback.

Scores on classroom examinations. Potential differences in mean scores (see Figure 2) were examined using a 3 (feedback group) x 2 (IR) x 5 (classroom examination) ANOVA, with significance observed for neither the main effects nor their interactions, all $F < 1$, all $p > .5$.

Final examination scores. Potential differences in mean scores (see Figure 3) were examined using a 3 (feedback group) x 2 (IR) ANOVA, with significance observed for each main and their interaction, all $F > 17.92$, all $p < .001$. Scheffe comparisons indicated that scores were significantly (a) higher for the immediate-feedback than for the delayed-feedback and control groups, (b) higher for the delayed-feedback than for the control group, and (c) highest when immediate feedback was combined with IR, all $p < .005$.

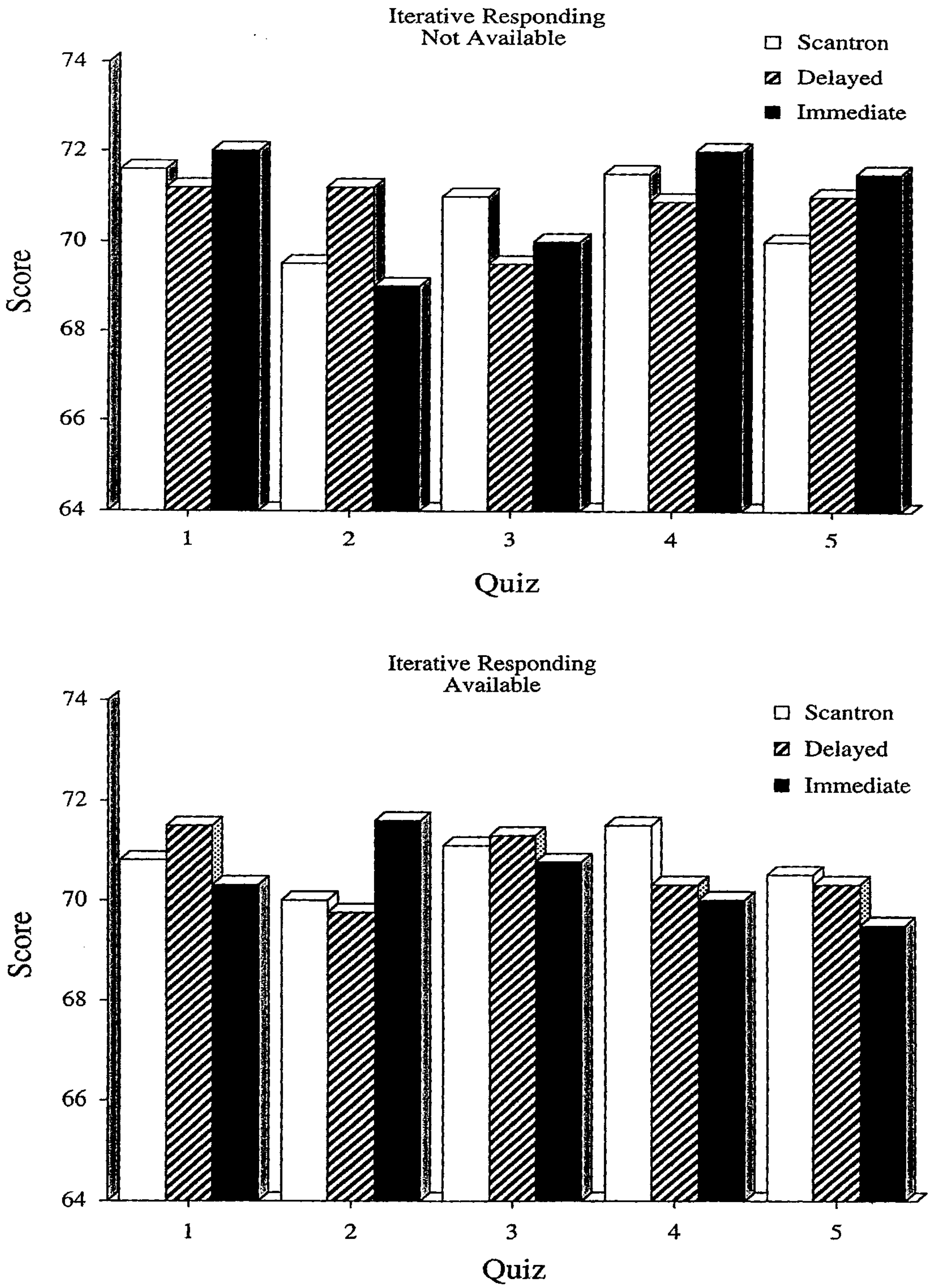

Figure 2. Scores on the five classroom examinations for the control and feedback groups when IR was (lower panel) or was not (upper panel) available.

Role of prior experience with test items. The percentage of correct responding on the 50 test items previously included on classroom examinations is presented in Figure 4. Potential differences in percentages of correct responding were examined using a 3 (feedback group) x 2 (IR) ANOVA, with significance observed for each main and their interaction, all *F*

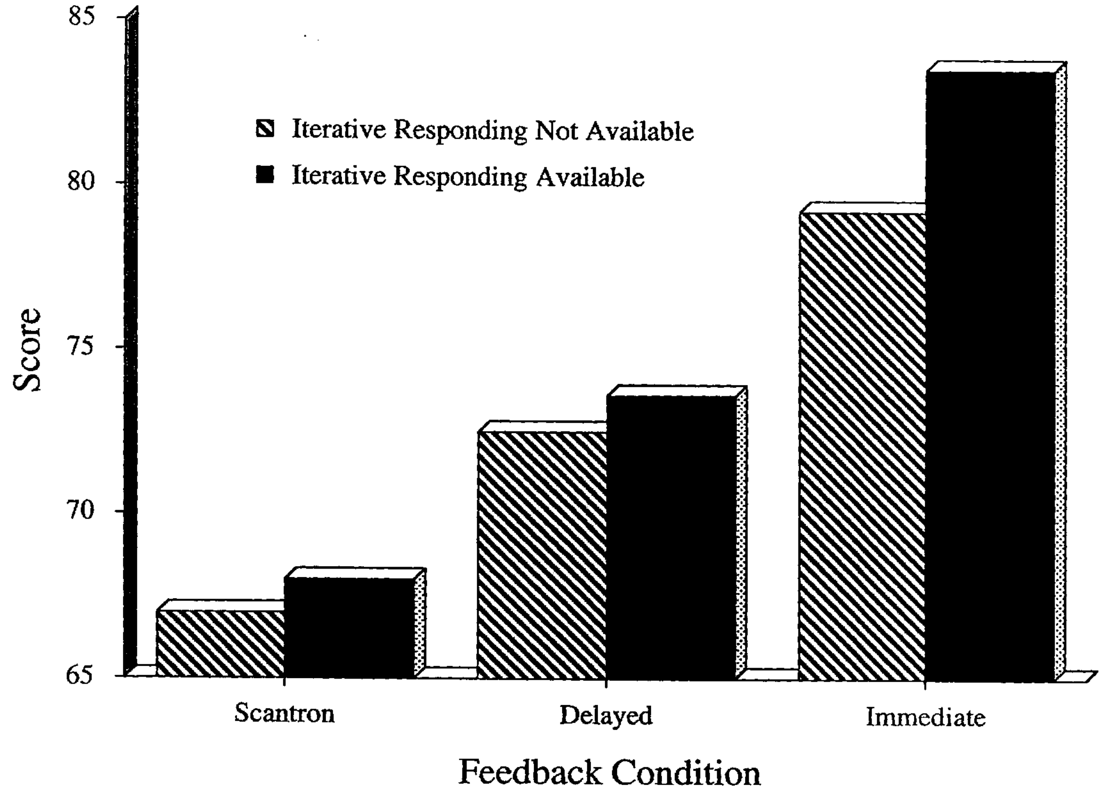

Figure 3. Final examination scores for the control and feedback groups when IR was available.

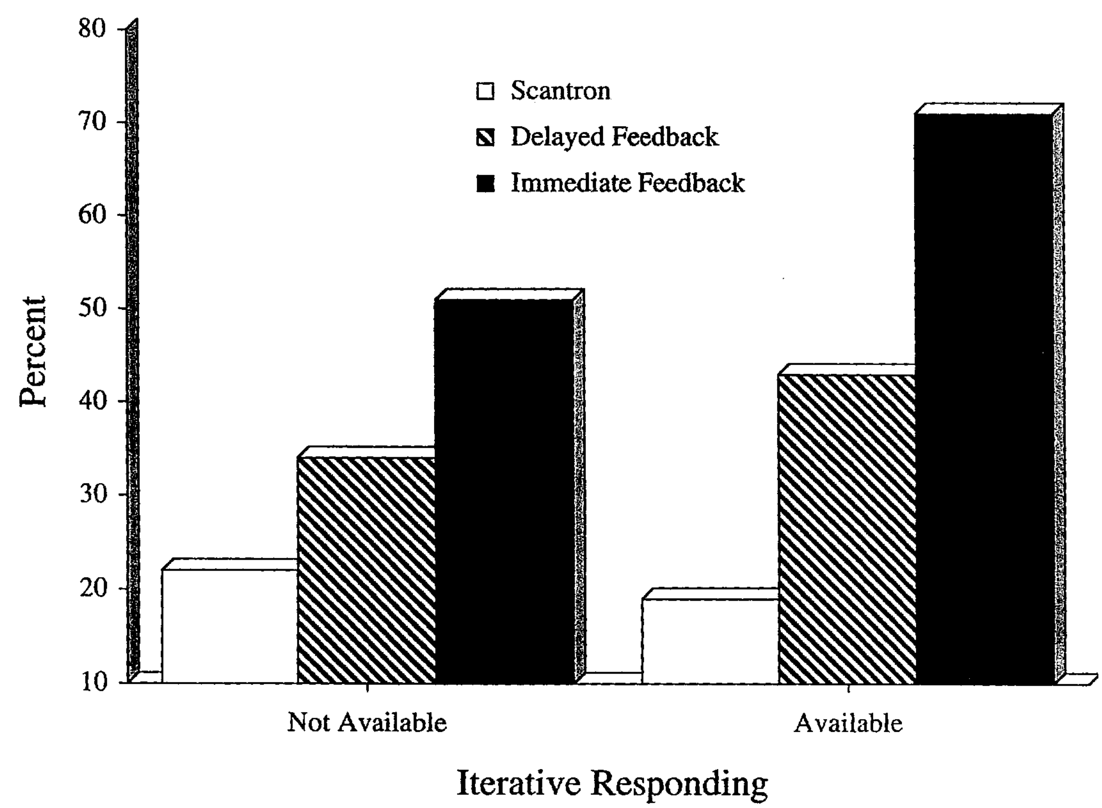

Figure 4. Percentages of correct responding on final examination items that were previously included on one of the classroom examinations.

> 7.44, all $p < .005$. Scheffe comparisons indicated that the percentage of correct responding was (a) higher for the immediate-feedback and delayed feedback groups than for the control group, (b) higher for the immediate-feedback than for delayed-feedback group, and (c) highest for the immediate-feedback group when IR was permitted, all $p < .001$.

Retention as a function of the time interval between examinations. The percentage of correct responding on the set of final examination

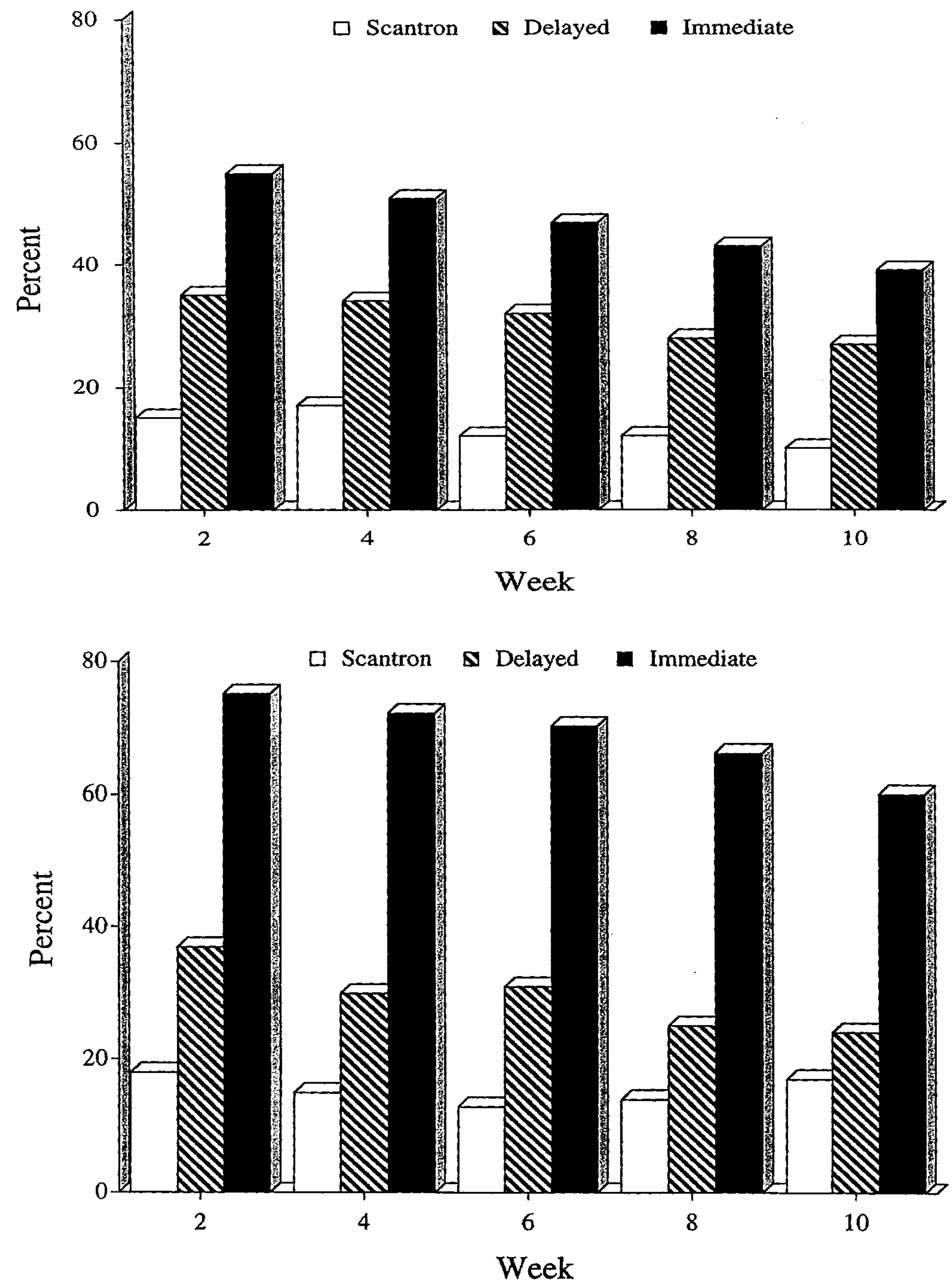

Figure 5. Percentages of correct responding on final examination items as a function of the number of weeks between the classroom examination and the final examination when IR was (lower panel) or was not (upper panel) available.

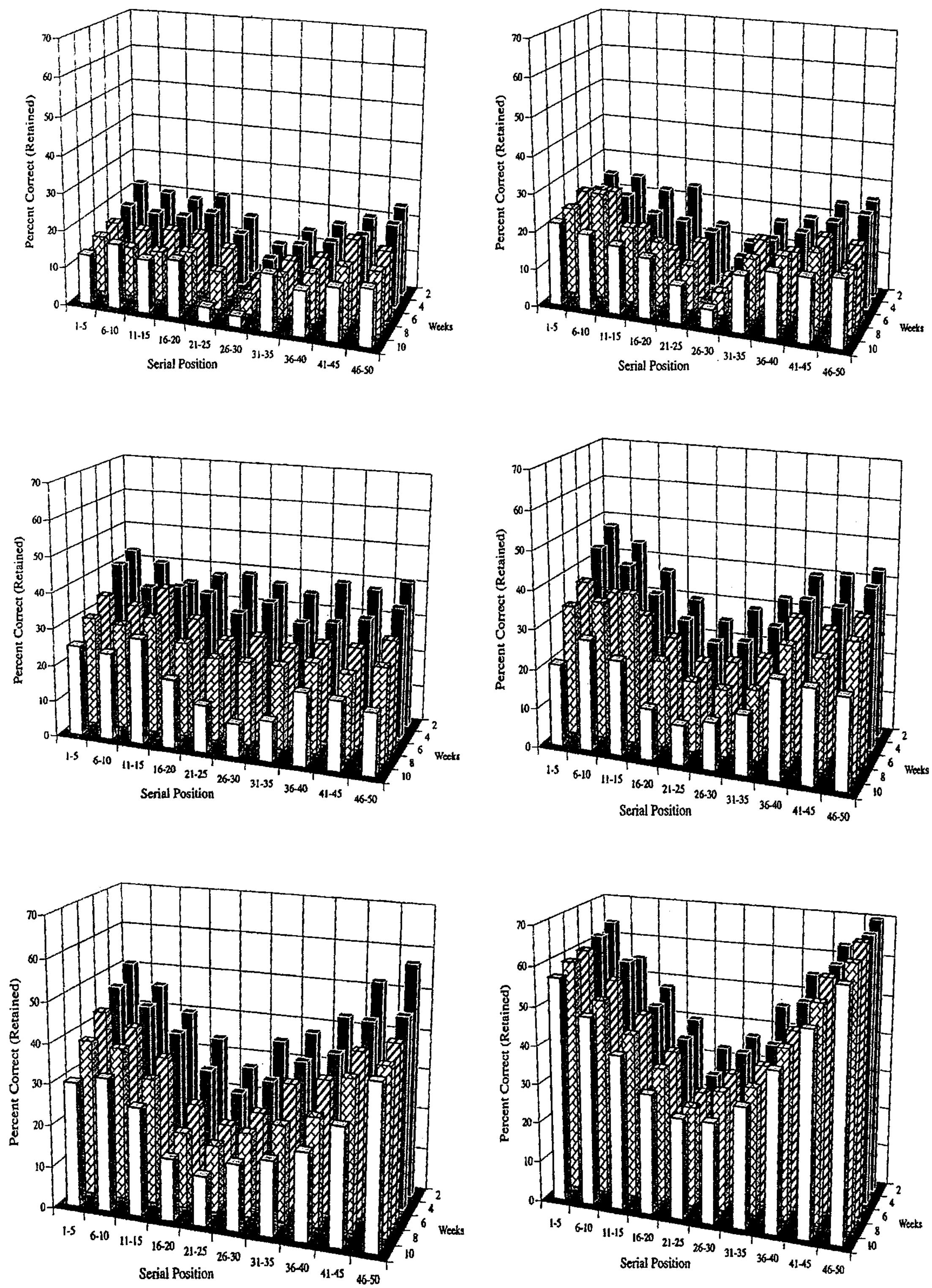

Figure 6. Percentages of correct responding on final examination items that had been repeated from classroom examination as a function of the serial position of those items on the classroom examinations and the number of weeks between the classroom examination and the final examination for the control group (upper panel), the delayed-feedback group (middle panel), and the immediate-feedback group (lower panel). Figures on the right represent values when IR was available; figures on the left represent when IR was not available.

items that had been included on the classroom examinations served as the primary measure of retention. Potential differences in correct responding (see Figure 5) were examined using a 3 (feedback group) x 2 (IR) x 5 (weeks between the initial and subsequent presentation of test items) ANOVA, with significance observed for each main effect and their interactions, all $F > 9.92$, all $p < .001$. Scheffe comparisons indicated that the percentage of retention was (a) highest for the immediate-feedback group at each interval, (b) higher when IR was available, (c) higher for both feedback groups at each interval than for the control group, (d) higher at each interval for the immediate-feedback group when IR was available, and (e) higher at Weeks 2 and 4 than at Weeks 8 and 10 for both feedback groups, Scheffe comparisons, all $p < .005$.

Serial position effect. The percentage of correct responding on final examination items repeated from the classroom examinations is presented in Figure 6. A visual inspection of each panel demonstrates that the distribution of correct responding conforms closely to the classic serial position curve. Potential differences in percentages of correct responding were examined using a 3 (feedback group) x 2 (IR) x 5 (weeks between the initial and subsequent presentation of test items) ANOVA, with significance observed for each main effect and their interactions, all $F > 6.83$, all $p < .001$. Scheffe comparisons indicated that mean overall percentage of correct responding was (a) higher for both feedback groups than for the control group, (b) higher when IR was available, (c) higher for the immediate-feedback than for the delayed-feedback group, and (d) highest for the immediate-feedback group when IR was available, all $p < .001$. Scheffe comparisons within the immediate-feedback group indicated (a) higher values when IR was available, and (b) significant declines in percentages of correct responding across the time intervals for the items at the middle, all $p < .05$, but not at the ends, all $p > .5$, of the serial position curve. Scheffe comparisons within the delayed-feedback group indicated significant declines in the percentage of correct responding across the time intervals for the items at each serial position, with the largest declines observed at the beginning and end of the serial position curve, all $p < .05$.

Identification of items repeated on the final examination, initial responses, and initial errors. Potential differences in the percentage of correctly identified repeated items (Figure 7, upper panel), initial responses (Figure 7, middle panel), and initial errors (Figure 7, lower panel) were examined using separate 3 (feedback group) x 2 (IR) x 5 (classroom examination) ANOVAs, with significance observed for the main and interactive effects of feedback group and IR, all $F > 21.28$, all $p < .0001$. Scheffe comparisons indicated that the percentage of identifying repeated items, initial responses, and initial errors was (a) higher for both feedback groups than for the control group, (b) higher for the immediate-feedback than for the delayed-feedback group, and (c) highest for the immediate-feedback group when IR was permitted, all $p < .001$.

Repetition of inaccurate perseverative responding. Responses on the

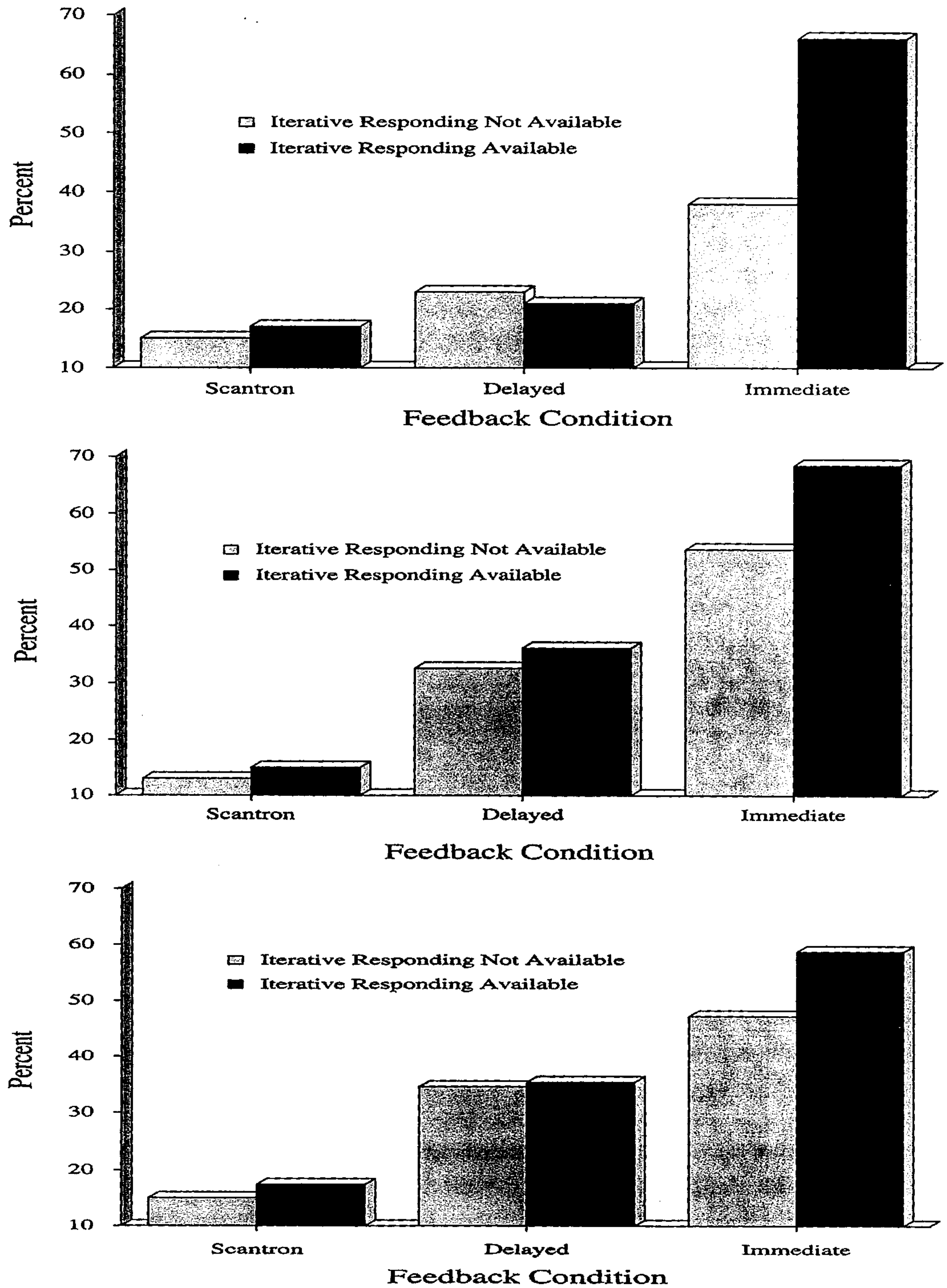

Figure 7. Percentages of accuracy at identifying initial responses (upper panel), initial errors (middle panel), and repeating the same incorrect response (lower panel) on final examination items that were presented on one of the classroom examinations.

second administration of the items presented on a classroom examination and on the final examination were dichotomized into the categories of repeating the same incorrect answer or making a different but also incorrect answer. Potential differences in the repetition of the same incorrect answer (see Figure 8) were examined using a 3 (feedback

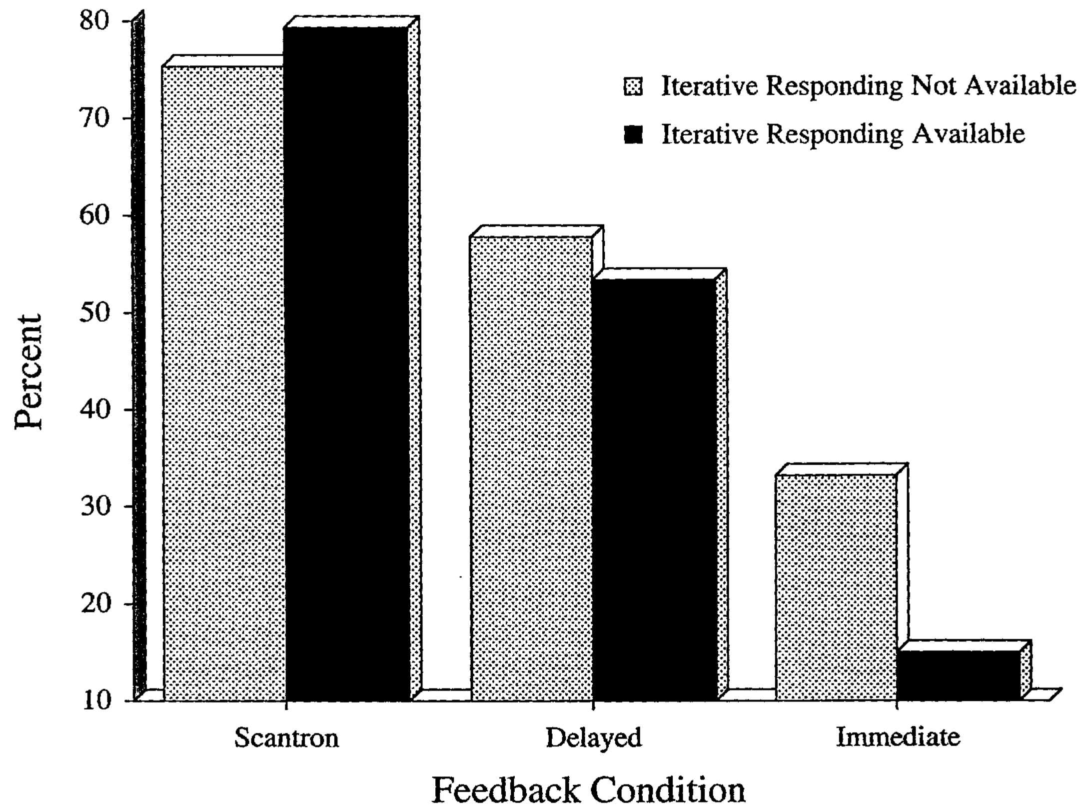

Figure 8. Percentage of respondents repeating the same incorrect response to test items presented on the final examination and one of the classroom examinations.

group) x 2 (IR) x 5 (classroom examination) ANOVA, with significance observed for the main and interactive effects of feedback group and IR, all $F > 8.64$, all $p < .0001$. Scheffe comparisons indicated that the percentage of participants repeating the same incorrect response was (a) lower for both feedback groups than for the control group, (b) lower for the immediate-feedback than for the delayed-feedback group, and (c) lowest for the immediate-feedback group when IR was permitted, all $p < .001$.

Conditional probabilities. Reductions in inaccurate perseverative responding were further evaluated for the item set administered on a classroom examination and on the final examination by determining the conditional probabilities of correct and incorrect responding on the second (final examination) and the first (classroom examination) administration of each item. Potential differences in conditional probabilities were examined using separate 3 (feedback group) x 2 (IR) ANOVAs.

The likelihood of responding correctly on the second administration of an item after having responded either correctly (C2 / C 1) or incorrectly (C2 / I1) on its initial administration differed only as a function of feedback group, IR, and their interaction, all $F > 7.01$, all $p < .001$. Scheffe comparisons indicated that these values were (a) highest for the immediate-feedback group when IR was available, (b) higher for the immediate-feedback than for the delayed-feedback and control groups, and (c) higher for the delayed-feedback than for the control group, all $p < .001$. These outcomes highlight the centrality of affirming correct

Table 1

Conditional Probability (in percentages) of Final Examination Outcomes Given Initial Test Outcomes By Feedback Group

		Response Group		
	Iterative Responding	Scantron Form	Delayed Feedback	Immediate Feedback
C2/C1	No	31.01	48.04	61.08
	Yes	35.85	50.31	63.25
C2/I1	No	24.58	33.42	45.61
	Yes	22.81	29.07	56.67
I2/C1	No	68.99	51.96	38.92
	Yes	64.15	49.69	36.75
I2/I1	No	75.42	66.58	54.39
	Yes	77.19	70.93	43.33

responding (C2 / C1) and the immediate correction of initially incorrect responses (C2 / I1). The likelihood of responding incorrectly on the second administration (I2 / C1) of an item after having responded correctly on its initial examination administration differed as a function of feedback group, $F = 14.32$, $p < .0001$. Scheffe comparisons indicated that this conditional probability was (a) lower for both feedback groups than for the control group, (b) lower for the immediate-feedback than for the delayed-feedback group, and (c) lowest for the immediate-feedback group, all $p < .001$. The likelihood of responding incorrectly on the second administration (I2 / I1) of an item after having responded incorrectly on its initial examination administration differed as a function of feedback group, IR, and their interaction, all $F > 23.19$, all $p < .0001$. Scheffe comparisons indicated that this conditional probability was (a) lowest for the immediate-feedback group when IR was available, (b) lower for the immediate-feedback than for the delayed-feedback group, and (c) lower for the delayed-feedback than for the control group, all $p < .001$.

Satisfaction measures. Participants' evaluations of their experimental group were expressed on a brief questionnaire upon the conclusion of the fifth classroom examination. The six scales described by Epstein and Brosvic (2002) were verified through factor analysis, with potential differences in scale scores examined using analyses of variance with feedback group and IR as the between-subject factors. Mean responses on the six scales (see Table 2) measuring test anxiety and time requirements did not differ as a function of the main or interactive effects of feedback group and IR, all $F < 1$, all $p > .5$. Mean responses on the scales measuring satisfaction with response format, clarity of response requirements, the desirability of the response form, and the benefits of testing differed significantly as a function of the main and interaction effects of feedback and IR, all $F > 6.98$, all $p < .005$. Scheffe comparisons indicated that mean scores on these latter four scales were (a) higher for both feedback groups than for the control group, (b) higher for the immediate-feedback than for the delayed-feedback group, and (c) highest when immediate-feedback was combined with IR, all $p < .005$.

Table 2

Posttest Measures Assessing Perceptions as a Function of Feedback Group

		Scantron	Delayed Feedback	Immediate Feedback IR Not Available	Immediate Feedback IR Available
Test	*M*	3.12	3.05	3.17	2.98
Anxiety	*SD*	1.02	1.45	1.29	1.09
Time		2.77	2.82	2.87	3.01
Requirements		1.15	1.34	1.05	1.32
Satisfaction With		2.45	3.54	4.15	4.37
Response Format		0.83	1.56	1.63	1.78
Clarity of Response		3.02	3.24	3.87	4.29
Requirements		1.41	1.18	1.62	1.34
Benefits of		2.67	3.87	4.22	4.57
Testing		1.08	1.29	1.79	1.62
Desirability of		2.58	3.66	4.26	4.65
Response Format		1.38	1.54	1.22	1.06

Discussion

The present study was undertaken to examine the contributions of timing of feedback and iterative responding during classroom examinations on final examination performance. The rates at which answers were changed in the absence of iterative responding were comparable to previously published figures (Ramsey, Ramsey, & Barnes, 1987). Robust effects were observed for feedback and iterative responding, but not for other variables that have received positive reports in prior studies. Sex of participant did not differentially affect any dependent measure, supporting the absence of such differences in our prior studies, but not reports that gender influences the likelihood of changing answers and interacts with item difficulty (see Ramsey et al., 1987) and self-reported achievement level (Hanna, 1976). Feedback did not differentially affect test-taking anxiety (see Clark, Fox, & Schneider, 1998), and neither inhibited the types of mindful behavior that permit deeper levels of understanding nor influenced learners to discontinue further interest in test items (see Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Similarly, no dependent measure was differentially affected by the source of feedback (educator, IF AT), an outcome that does not support prior reports that more learning is observed when feedback is provided by an educator (see Dihoff et al., 2005; Sassenrath & Garverick, 1965). This latter outcome suggests that the differences observed in our prior study (Dihoff et al., 2005) are specific to participants' individual syndromes and the learning difficulties attributable to and originating from within them. We are currently examining the effects of the source of feedback on learning and retention using diverse samples of children with and without known developmental delays.

It should be noted that much of the literature on delayed feedback has focused on the acquisition of simple motor skills, and thus the outcomes described above may be specific to motor skills, the learning of which requires the repetition of a single movement or task, and may not apply to tests employing discrete and unrelated items presented within a multiple-choice format. In contrast, the present results indicate that immediate, rather than delayed feedback, enhances acquisition and retention; however, these outcomes may be specific to testing situations in which responses are selected from among options presented by the experimenter, as opposed to the groups in effect in studies of motor behavior in which responses are typically provided by participants. These caveats suggest that the processes underlying the provision of affirming and corrective information interact differentially with task requirements, and it should be expected that knowledge of results may or may not be beneficial across a diversity of experimental settings. For example, in studies of eyewitness recall ability, confirming feedback has increased distortions of recall (e.g., Wells, Olson, & Charman, 2003) and limited the ability to apply selected programming statements (Lee & Dwyer, 1994) whereas disconfirming feedback has increased accuracy (e.g., Wells et al., 2003) and enhanced performance on sentence completion tasks (Försterling & Morgensten, 2002), the acquisition of foreign language skills (Brandl, 1995), and the performance of low-ability students on social studies tests (Clariana, 1990).

Test developers typically assume that adequately learned facts are not subject to interference from the confluence of correct and incorrect statements; indeed, misinformation serves as the context within which correct responses must be discriminated when test materials are presented within a multiple-choice context. This assumption is surprising in light of the repeated demonstration that exposure to misinformation within test items causes participants later to perceive or to remember misinformation as being correct (e.g., Remmers & Remmers, 1926). This phenomenon represents the memorial effects of exposure to misinformation and has been described (see Brown et al., 1999) as the Negative Suggestion Effect (NSE). The robustness of the NSE has been demonstrated during multiple-choice tests (e.g., Rees, 1986), true-and-false tests (e.g., Toppino & Brochin, 1989), and spelling tests (e.g., Brown, 1986; Jacoby & Hollingshead, 1990). The interference of NSE with retrieval (cued recall) and discrimination (e.g., multiple-choice) can be demonstrated even when the misinformation is specifically identified as incorrect (e.g., Bartlett, 1932; Brown et al., 1999; Kay, 1955).

The robust effects of feedback, both in the present and in prior studies, cannot be demonstrated unless learners are presented with the same test items on a second occasion, such as a cumulative final examination. The final examination in the present study included 10 randomly selected items from each classroom examination, plus 50 entirely new items. No differences in performance on the new items were observed between any experimental or control group, an expected outcome, as the final examination represented the initial presentation of the new items. Performance on the items presented for a second time for participants who were provided with delayed feedback that had been coupled with iterative responding was superior to that of controls. Performance on the items presented for a second time for participants provided with immediate feedback coupled with iterative responding was superior to that of participants provided with delayed feedback and iterative responding, again replicating and extending the results of our prior studies.

References

BANGERT-DROWNS, R. L., KULIK, C. L., KULIK, J. A., & MORGAN, M. T. (1991). The instructional effects of feedback in test-like events. *Review of Educational Research*, 61, 179-212.

BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

BRANDL, K. K. (1995). Strong and weak students' preferences for error feedback options and responses. *The Modern Language Journa*l, 79, 194-211.

BROWN, A. S. (1986). Experiencing misspellings and spelling performance: Why wrong isn't right. *Journal of Educational Psychology*, 80, 488-494.

BROWN, A. S., SCHILLING, H., & HOCKENSMITH, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology*, 91, 756-764.

BRUNING, R., SCHRAW, G., & RONNING, R. (1999). *Cognitive psychology and instruction*. Columbus, OH: Merrill Prentice Hall.

CLARIANA, R. B. (1990). A comparison of answer until correct feedback and knowledge of correct response feedback under two groups of contextualization. *Journal of Computer-Based Instruction*, 17, 125-129.

CLARK, J. W., FOX, P. A., & SCHNEIDER, H. G. (1998). Feedback, test anxiety and performance in a college course. *Psychological Reports*, 82, 203-208.

DIHOFF, R. E., BROSVIC, G. M., & EPSTEIN, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record*, 53, 533-548.

DIHOFF, R. E., BROSVIC, G. M., & EPSTEIN, M. L. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, 54, 207-231.

DIHOFF, R. E., BROSVIC, G. M., EPSTEIN, M. L., & COOK, M. J. (2005). Adjunctive role for immediate feedback in the acquisition and retention of mathematical fact series by elementary school students classified with mild mental retardation. *The Psychological Record*, 55, 39-66.

EPSTEIN, M. L., & BROSVIC, G. M. (2002). Students prefer the immediate feedback assessment technique. *Perceptual and Motor Skills*, 90, 1136-1138.

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.

FÖRSTERLING, F., & MORGENSTEN, M. (2002). Accuracy of self-assessment and task performance: Does it pay to know to know the truth? *Journal of Educational Psychology*, 94, 576-585.

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

JACOBY, L. L., & HOLLINGSHEAD, A. (1990). Reading student essays may be hazardous to your spelling: Effects of reading incorrectly and correctly spelled words. *Canadian Journal of Psychology*, 44, 345-358.

KAY, H. (1955). Learning and retaining verbal material. *British Journal of Psychology*, 46, 81-100.

KLUGER, A. N., & DENISI, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-282.

KULHAVY, R. W., & STOCK, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279-308.

LEE, D., & DWYER, F. (1994). The effect of varied feedback strategies on students' cognitive and attitude development. *International Journal of Instructional Media*, 21, 321-330.

PRESSEY, S. L. (1950). Development and appraisal of devices providing immediate automatic sorting of objective tests and concomitant self-instruction. *The Journal of Psychology*, 29, 417-447.

RAMSEY, P. H., RAMSEY, P. P., & BARNES, M. J. (1987). Effects of student confidence and item difficulty on test score gains due to answer changing. *Teaching of Psychology*, 14, 206-210.

REES, P. J. (1986). Do medical participants learn from multiple choice examinations? *Medical Education*, 20, 123-125.

REMMERS, H. H., & REMMERS, E. M. (1926). The negative suggestion effect of true-false examination questions. *Journal of Educational Psychology*, 17, 52-56.

ROBIN, A. L. (1978). The timing of feedback in personalized instruction. *Journal of Personalized Instruction*, 3, 81-87.

SASSENRATH, J. M., & GARVERICK, C. M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology*, 56, 259-263.

SKINNER, B. F. (1968). *The technology of teaching.* New York: Meredith Corporation.

THORNDIKE, E. L. (1913). *Educational psychology, volume 1: The original nature of man.* New York: Columbia University.

THORNDIKE, E. L. (1927). The law of effect. *American Journal of Psychology*, 39, 212-222.

TOPPINO, T. C., & BROCHIN, H. A. (1989). Learning from tests: The case of true-false examinations. *Journal of Educational Research*, 83, 119-124.

WELLS, G. L., OLSON, E. A., & CHARMAN, S. D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. *Journal of Experimental Psychology: Applied*, 9, 42-52.