The Psychological Record, 2006, 56, 205-218

ACQUISITION AND RETENTION OF ESPERANTO: THE CASE FOR ERROR CORRECTION AND IMMEDIATE FEEDBACK

GARY M. BROSVIC and MICHAEL L. EPSTEIN
Rider University

ROBERTA E. DIHOFF
Rowan University

MICHAEL J. COOK
Rider University

Participants completed 5 laboratory examinations during which the number of responses permitted (1 response, up to 4 responses) and the timing of feedback (no feedback control: Scantron form; delayed feedback: end-of-test, 24-hr delay; immediate feedback: assistant, response form) were manipulated. Participants completed a 100-item cumulative final examination which included 10 items from each laboratory examination, plus 50 entirely new items; this cumulative examination was presented again 3 and 6 months later to determine if the timing of feedback affected retention. Timing of feedback and opportunity to engage in iterative responding (IR) interacted significantly to enhance the cumulative final examination performance of participants provided with immediate feedback, with comparable performance demonstrated when immediate feedback was provided by either an experimental assistant or the Immediate Feedback Assessment Technique (IF AT) form. Together with prior reports from our laboratory, these results highlight the robustness of immediate feedback and IR to facilitate learning in a manner that is comparable between laboratory and classroom settings, although the size of the immediate feedback effect observed in the laboratory is less robust than that observed in the classroom.

Many institutions of higher learning, through their academic counseling services, offer tutorials that characterize both the distinctions in purpose (i.e., simple recognition to analysis and synthesis) and structure (i.e., grammar, inclusion of qualifiers or superlatives) of test materials presented in multiple-choice format. It has been generally assumed that adequately learned facts are not subject to interference from the confluence of correct and incorrect item choices, especially because misinformation serves as the context within which the correct response must be discriminated. This assumption is surprising in light of the repeated demonstration

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

that exposure to misinformation causes the learner to later perceive or remember misinformation as being correct (e.g., Remmers & Remmers, 1926). This phenomenon represents the memorial aftermath of exposure to misinformation, and more recently, it has been described as the Negative Suggestion Effect, or NSE (Brown, Schilling, & Hockensmith, 1999).

The NSE is observed when discriminations are presented in either a dichotomous (i.e., true-and-false) or multiple-choice format (e.g., Brown, 1986), and within the multiple-choice format, the NSE is evident even after the explicit labeling of distractors and misinformation as incorrect (e.g., Brown et al., 1999). The robustness of the NSE, especially for the multiple-choice format, has considerable implication for classroom testing in light of the modal use of Scantron forms for recording answers and predictable delays in the scoring and return of classroom examinations. Indeed, this combination, which characterizes the typical classroom, permits the learner to exit test questions without the affirmation of correct responding and the correction of inaccurate responding; thus, on subsequent tests such as cumulative final examinations and standardized assessments, learners are less likely to repeat correct responses and more likely to repeat incorrect responses. In general, the literature supports the efficacy of feedback for correcting errors, but there is considerable disagreement as to whether delayed or immediate feedback is more effective, and there has been a paucity of research on the effects of the affirmation of correct responding.

Much of the support for delayed feedback stems from the early work of Brackbill and her associates (e.g., Brackbill, Bravos, & Starr, 1962) in which brief delays in the presentation of feedback promoted the retention of meaningful material, an outcome replicated when the delay intervals were substantially lengthened (e.g., Kulhavy & Anderson, 1972; Surber & Anderson, 1975). The finding that delayed feedback enhances retention has become known as the delay-retention effect (DRE), and its proponents generally adhere to the Kulhavy and Anderson (1972) interference-perseveration hypothesis: Initial errors do not compete with to-be-learned correct responses if corrective information is delayed, because errors that are forgotten should not interfere with retention.

Presumably, the retention of correct responses is unaffected by delays in the provision of feedback although DRE proponents have been silent on this point and on how it is diametrically opposed to the NSE explanation. Proponents of immediate feedback theorize that the overall efficiency of retention mechanisms is enhanced by the immediate provision of corrective information. Acquisition of verbal stimuli (Ammons, 1956) and simple motor skills (Anderson, Magill, & Seklya, 2001), for example, is enhanced by immediate feedback, despite the contention that corrective information increases proactive interference. In fact, the results of a number of studies have suggested that the recall of incorrect responses can facilitate the acquisition of correct solutions (see Peeck, van Den Bosch, & Kreupeling, 1985). The contradictory results of studies on immediate and delayed feedback can, in part, be explained through noteworthy differences in procedures, test stimuli, and operationalizations

of 'immediate' and 'delayed' feedback. Despite these differences, it has generally been held that immediate feedback is more effective in the laboratory and that delayed feedback is more effective in the classroom.

The provision of immediate feedback in the classroom typically requires the types of computerized testing employed in the laboratory, and as a practical matter, the availability of a sufficient number of computers to conduct classroom examinations under controlled and supervised conditions. We have reported extensively upon the development and validation of the Immediate Feedback Assessment Technique,[1] or IF AT (Epstein, Brosvic, Dihoff, Lazarus, & Costner, 2003; Epstein et al., 2002), and the results of our studies indicate that immediate feedback significantly enhances retention. It is of interest, therefore, to examine how these procedures might affect retention in the laboratory.

In the present study we report on how immediate and delayed feedback differentially affect the laboratory acquisition and retention of an artificial language, Esperanto, which was selected to represent the types of definitions and concepts that are presented to students during lectures and examinations in introductory courses. The length of each lab session was equivalent to those of traditional lectures and the number of sessions and examinations were comparable to those reported in prior studies (Epstein et al., 2002, 2003), thus equating the instructional and cognitive requirements between classroom and laboratory settings. Similarly, the procedures for recording responses and delivering feedback were identical to those in our classroom studies. The comparability of procedures permits the opportunity to determine if the affirmation of correct responding is as effective in the laboratory as it is the classroom and if the Negative Suggestion Effect is a phenomenon that is as operative in the laboratory as it is in the classroom.

Method

Participants

Twenty male and 80 female students enrolled in liberal arts and sciences undergraduate courses served as voluntary participants. The modal participant was a Caucasian female, approximately 19 years of age and majoring in either the liberal arts and sciences or elementary education.

Materials

Five laboratory examinations assessing definitions of simple Esperanto words were prepared from the Concise Esperanto and English Dictionary (Wells, 1992), each with 50 items, and each item with four response options (i.e., A, B, C, D). The cumulative final examination consisted of 100 items which included 10 randomly selected items from each laboratory examination, plus 50 entirely new items.

[1]Registered trademark of E3 Corporation

Design and Procedures

Participants completed five laboratory components, each of which consisted of seven 1-hr sessions. Sessions 1 and 2 consisted of formal lectures on vocabulary materials; Sessions 3 and 4 consisted of individual programmed learning exercises; Sessions 5 and 6 consisted of small-group learning activities followed by an overall review of materials; Session 7 consisted of the presentation of a 50-item laboratory examination. The sequence described above was employed to approximate the pedagogical procedures, time course, and classroom activities used in our prior studies of the effects of feedback on classroom learning (Dihoff, Brosvic, & Epstein, 2003, 2004).

Ten participants (2 male and 8 female) were randomly assigned to each of the 10 groups generated from the combination of the two response conditions (1 response permitted, up to 4 responses permitted) and the five feedback conditions (Scantron form, 24-hr delayed feedback, end-of-test feedback, immediate feedback/assistant, and immediate feedback/IF AT form). Participants limited to making only one response were not permitted to change their initial response; participants permitted to answer until correct were permitted to make up to four responses. The Scantron form groups served as controls, with all answers recorded with a pencil on a Scantron form. In the end-of-test feedback condition, 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 min. In the 24-hr 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 min. During this review process all participants not assigned to the end-of-test and 24-hr delayed feedback conditions were required to remain silent and to read noncourse materials, and were not permitted to share their materials nor to ask any questions of the session monitors. Participants in the immediate feedback/assistant condition sat in pairs, and an experimental assistant sat between the members of each pair. These participants recorded their answers on Scantron forms, and after making a response, the experimental assistant signaled that a correct response was made by the holding up of 3- x 5-inch index cards. In the event of an incorrect response and when iterative responding was available, additional cards identifying already-selected responses (e.g., A, B) were made visible to each participant, separately, who continued to select responses until the correct answer was discovered. In the immediate feedback/form condition, answers were recorded using the IF AT form which enabled participants to receive immediate affirming or corrective feedback; if appropriate to the experimental condition, the participant was permitted to continue selecting responses until the correct answer was discovered.

The cumulative final examination was administered 1 week after completion of the fifth laboratory examination, and at the time, all

participants used Scantron forms to record their answers. Once the cumulative final examination was completed, participants identified those items they believed had been repeated from one of the laboratory examinations, their initial response to those items, and whether their initial response had been correct. As a test of long-term retention, the same cumulative final examination was completed 3 and 6 months later using Scantron form.

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 and 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. Performance on the laboratory examinations served as the primary measure of acquisition and performance on laboratory examination items that were repeated on the cumulative final examinations (initial and 3- and 6-month followups) served as the primary measure of retention.

Results

There were no differences in SAT scores, current semester classroom performance, overall GPA, foreign language placement test scores, and past and current semester classroom performance in the foreign languages as a function of sex of participant, feedback condition, 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 24-hr delayed feedback and the end-of-test feedback conditions, all $F < 1$, all $p > .5$, and thus responses for these two conditions were aggregated and are referred to hereafter as delayed feedback. No differences were observed in any dependent measure between participants assigned to the immediate feedback/assistant and the immediate feedback/IF AT form conditions, all $F < 1$, all $p > .5$, and thus responses for these two conditions were aggregated and are referred to hereafter as immediate feedback. No differences in any dependent measure described below was observed as a function of sex of participant, all $F < 1$, all $p > .5$, and thus all dependent measures were collapsed over sex of participant.

Scores on laboratory examinations. Potential differences in mean scores on the laboratory examinations were examined using an analysis of variance with feedback condition and IR as between-subject factors and laboratory examination as the within-subjects factor. Significance was observed for neither the main effects nor their interactions, all $F < 1$, all $p > .5$.

Cumulative final examination scores. Scores on the cumulative examination are presented in Figure 1. Potential differences in scores on the cumulative examination were assessed using an analysis of variance with feedback condition and IR as between-subject factors. Significance was observed for both main effects and their two-way

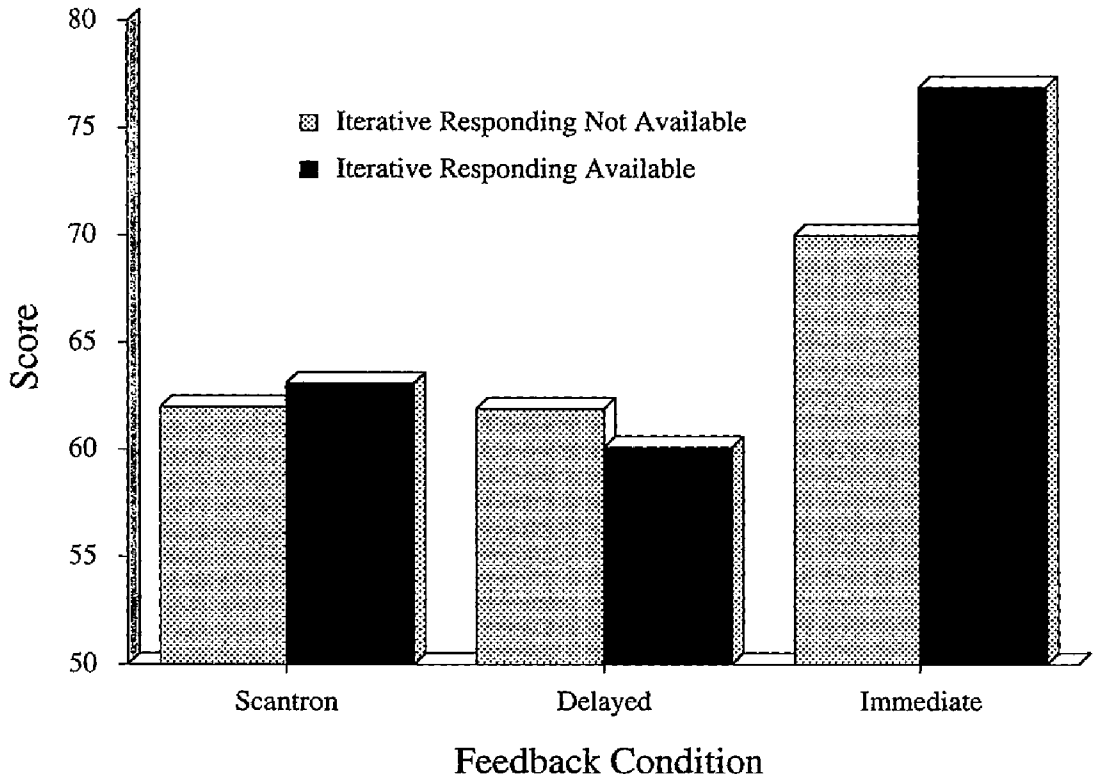

Figure 1. Cumulative final examination scores for the control (Scantron form) and feedback conditions as a function of iterative responding.

interaction, all $F > 17.92$, all $p < .001$. Scheffe comparisons indicated that scores on the cumulative final examination (a) did not differ between the delayed feedback and Scantron conditions, all $p > .5$, (b) were higher for the immediate feedback than for the delayed feedback and Scantron conditions, all $p < .005$, and (c) were highest when immediate feedback was combined with IR, all $p < .005$.

Role of prior experience with test items. Percentage of correct responding on the cumulative final examination was dichotomized into separate scores for the 50 items repeated from the laboratory examinations and for the 50 new items. Potential differences in percentages of correct responding were examined using an analysis of variance with feedback condition, IR, and item novelty (new, repeated from a laboratory examination) as between-subject factors. Significance was observed for each main effect and their interaction, all $F > 20.22$, all $p < .0001$. Scheffe comparisons indicated that scores on the novel items did not differ between the feedback and IR conditions, all $p > .5$. Scheffe comparisons for the repeated items (see Figure 2) indicated that scores were (a) comparable within the delayed feedback and Scantron conditions and did not differ as a function of IR, all $p > .5$, (b) higher for the immediate feedback than for the delayed feedback and Scantron conditions in the absence of IR, $p < .005$, and (c) highest in the immediate feedback condition when IR was permitted, $p < .001$.

Identification of original examination items, responses, and errors.

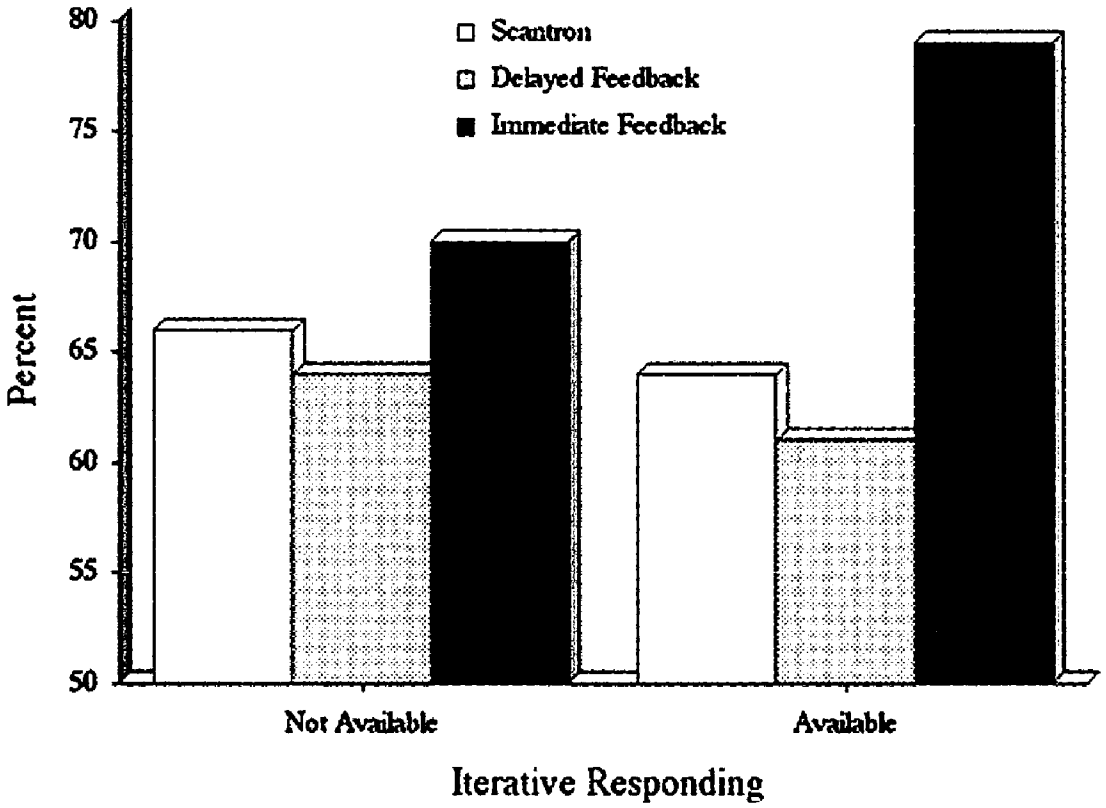

Figure 2. Percentages of correct responding on cumulative final examination items that were previously included on a laboratory examination as a function of the opportunity to engage in iterative responding.

Upon the completion of the cumulative final examination, participants reviewed each item with the task of identifying items repeated from laboratory examinations, their initial response to those items identified as repeated, and if their initial response had been correct (see Figure 3). Potential differences in the percentage of correctly identified repeated items (upper panel), correctly recalled initial responses (middle panel), and accuracy at recalling if initial responses were correct (lower panel) were examined using separate analyses of variance with feedback condition and IR as between-subjects factors and laboratory examination as the within-subjects factor. Significant main effects and interactions were observed, in each analysis, for feedback condition and IR, all $F > 24.52$, all $p < .0001$. Scheffe comparisons indicated that accuracy at identifying repeated test items, recalling initial responses, and recalling original responses were (a) higher for the immediate feedback than for the delayed feedback and Scantron conditions, and (b) highest for the Immediate feedback condition when IR was permitted, all $p < .001$.

End-of-term retention. Percentages of correct responding on items repeated from the laboratory examinations (see Figure 4) served as the primary measure of end-of-term retention. Potential differences in scores were examined using an analysis of variance with feedback condition and IR as between-subject factors and number of weeks between the initial and subsequent presentation of test items as the within-subjects

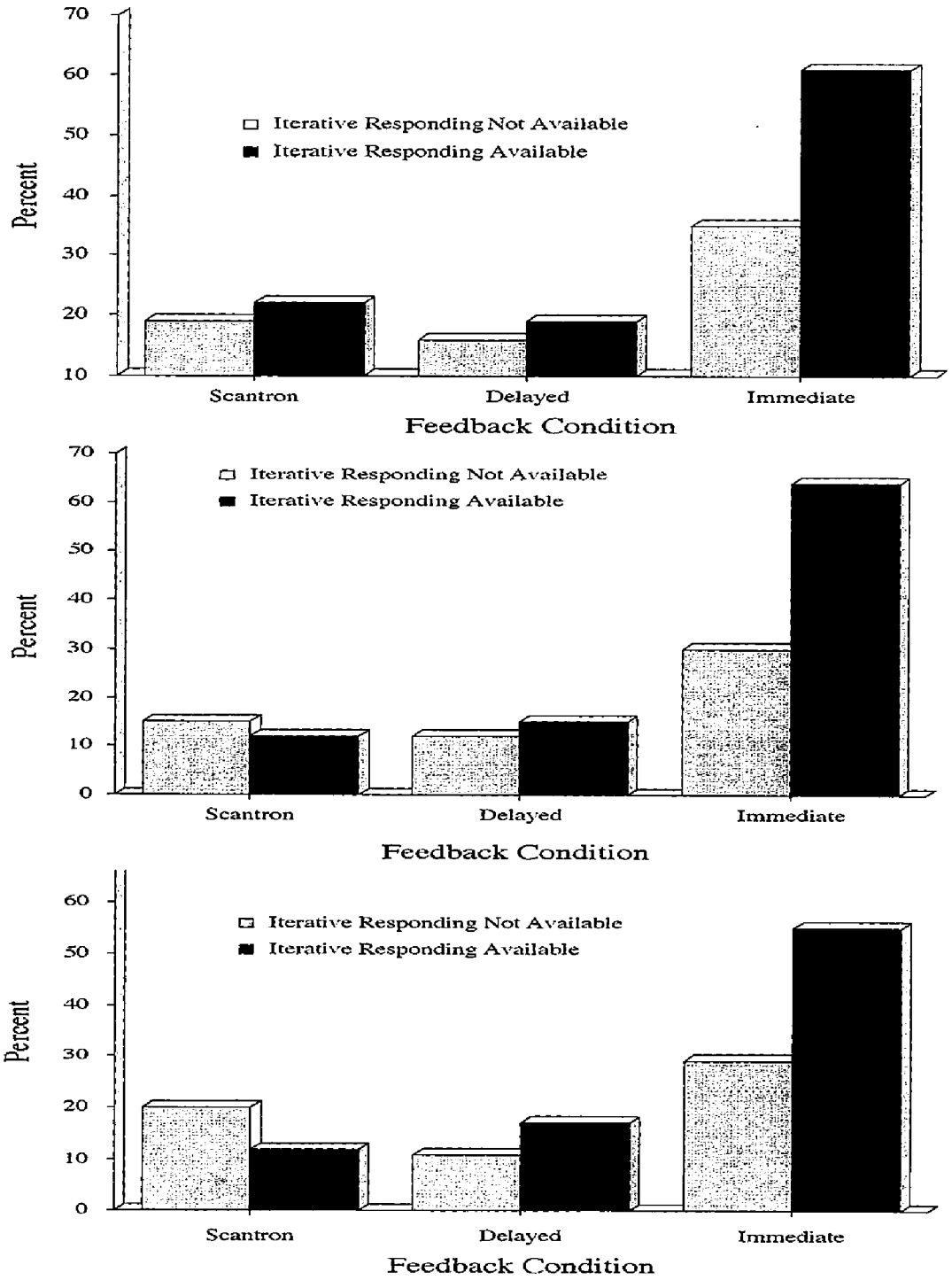

Figure 3. Percentage of correctly identified repeated items (upper panel), percentage of correctly recalled initial responses (middle panel), and accuracy at recalling if initial responses were correct (lower panel) on laboratory examinations items repeated on the cumulative final examination as a function of feedback condition and the opportunity to engage in iterative responding.

factor. Significance was observed for the main effects and interactions of feedback, IR, and the number of weeks, all $F > 13.88$, all $p < .001$. No differences in the scores of Scantron participants were observed

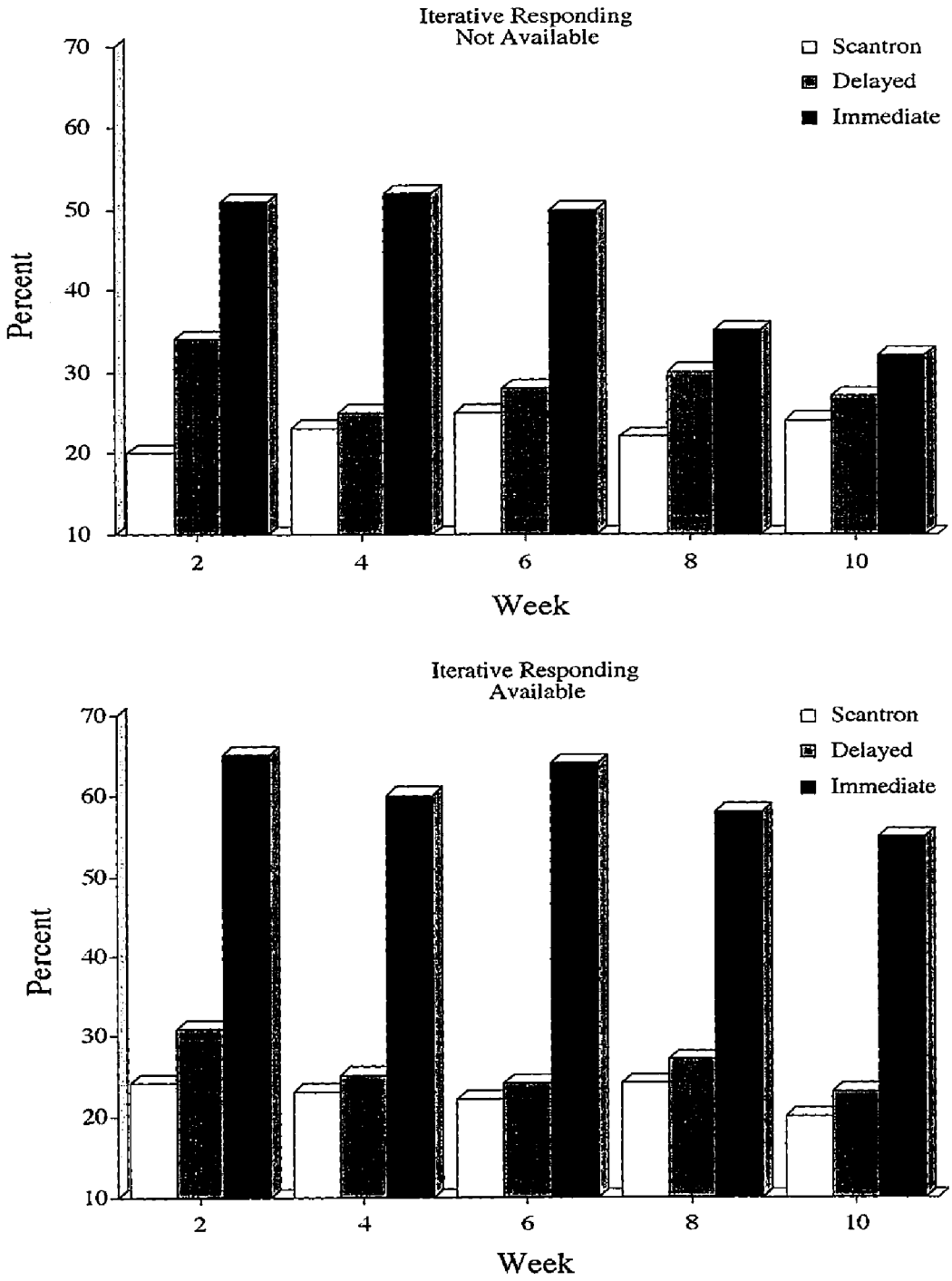

Figure 4. Percentages of correct responding on cumulative final examination items as a function of the number of weeks between the laboratory examination and the cumulative final examination when the opportunity to engage in iterative responding was (lower panel) or was not (upper panel) available.

as a function of IR, the number of weeks, or their interaction, Scheffe comparisons, all $p > .5$. Scores for the delayed feedback condition were higher than those of the Scantron condition at Week 2, Scheffe comparisons, all $p < .015$. During Weeks 2 through 10, scores for

the immediate feedback condition were higher than the those of the Scantron and delayed feedback conditions, and were highest when IR was permitted, all $p < .005$. Scheffe comparisons within the immediate feedback condition indicated that scores, when IR was permitted (Figure 4, bottom panel), were unaffected by the number of weeks between assessments, all $p > .5$, and that, when IR was not permitted (Figure 4, top panel), were higher during Weeks 2 through 6 than during Weeks 8 through 10, all $p < .005$.

Postterm retention. Percentages of correct responding on items repeated from the laboratory examinations (see Figure 5) served as the primary measure of postterm retention on the 3- and 6-month followup tests. Potential differences in scores on the 3- and 6-month followup tests were examined using a repeated measures analysis of variance with feedback condition and IR as between-subject factors and time of testing (cumulative final examination, 3 months, 6 months) as the within-subjects factor. Significance was observed for the main effects and interactions of feedback, IR, and time of testing, all $F > 33.96$, all $p < .0001$. Scheffe comparisons indicated that scores were (a) comparable within the Scantron and delayed feedback conditions and did not differ as a function of IR, all $p > .5$, (b) higher at each assessment for the immediate feedback condition, all $p < .005$, and (c) highest within the immediate feedback when IR was permitted, all $p < .005$.

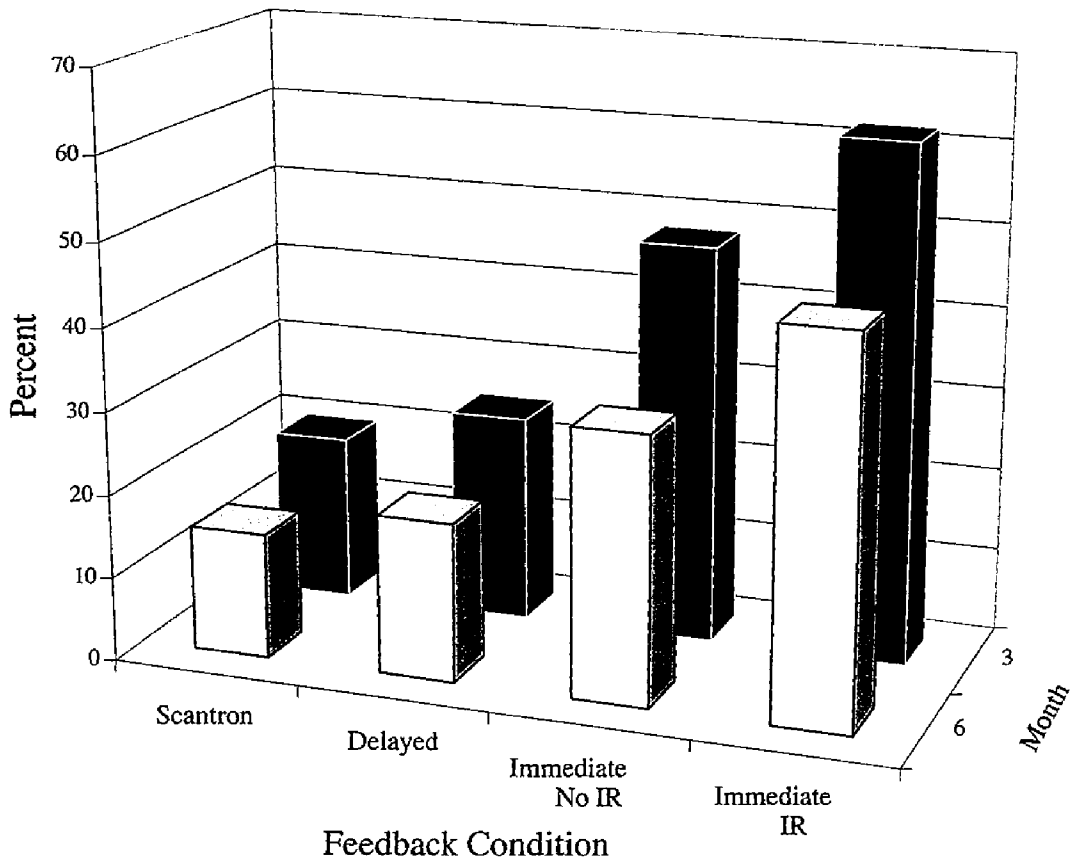

Figure 5. Percentages of correct responding on the cumulative final examination when it was administered 3 and 6 months after the initial presentation.

Conditional probabilities. Reductions in inaccurate perseverative responding were further evaluated for items administered on both a laboratory examination and the cumulative final examination by determining the conditional probabilities of correct and incorrect responding on the second (cumulative final examination) and the initial (laboratory examination) administration of each item. A review of the conditional probabilities values presented in Table 1 indicates that the hypothesis that inaccurate perseverative responding is reduced to a greater extent by the provision of delayed, rather than immediate feedback, cannot be supported. Potential differences in conditional probability values were examined using separate analyses of variance with feedback condition and IR as the between-subject factors and laboratory examination as the within-subjects factor.

Table 1

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

		Response Condition		
	Iterative Responding	Scantron Form	Delayed Feedback	Immediate Feedback
C2/C1	No	25.09	31.11	57.36
	Yes	30.44	37.65	60.02
C2/I1	No	28.88	36.28	41.38
	Yes	26.64	34.98	50.91
I2/C1	No	74.91	68.89	42.64
	Yes	69.56	62.35	39.98
I2/I1	No	71.12	63.72	58.62
	Yes	73.36	65.02	49.09

The likelihood of (a) responding correctly on the second administration of an item after having responded correctly on its initial administration, and (b) responding correctly on the second administration of an item after having responded incorrectly on its initial administration differed as a function of the main effects and interaction of feedback and IR, all $F > 5.16$, all $p < .001$. Scheffe comparisons indicated that these probability values were (a) highest for the immediate feedback condition when IR was available, and (b) higher for the immediate feedback than for the delayed feedback and Scantron conditions, all $p < .001$. These outcomes highlight the centrality of affirming correct responding, and that affirmation thus functions as a form of reinforcement.

The likelihood of responding (a) incorrectly on the second administration of an item after having responded correctly on its initial examination administration, and (b) incorrectly on the second administration of an item after having responded incorrectly on its initial examination administration

differed as a function of the main effects of interaction of feedback condition and IR, all $F > 12.73$, all $p < .0001$. Scheffe comparisons indicated that these probability values were (a) lower for the immediate feedback than for the delayed feedback and Scantron conditions, and (b) lowest for the immediate feedback condition when IR was permitted, all $p < .001$.

Participant motivation. The contribution of salience was estimated by comparing the magnitude of the conditional probabilities observed in the present laboratory study with those determined in a prior classroom study in which identical instructional, experimental and control, and assessment procedures were used (Brosvic, Epstein, Cook, & Dihoff, 2005). Potential differences were examined using separate analyses of variance with test setting (laboratory, classroom), feedback condition (Scantron, delayed feedback, immediate feedback), and iterative responding as between-subjects factors. Significance was observed for the main effects of feedback condition and iterative responding and for the three-way interaction of test setting, feedback condition, and iterative responding, all $F > 16.84$, all $p < .0004$. Scheffe comparisons within the Scantron condition (Scantron form) indicated no differences in the magnitude of probability values between the two test settings, all $p > .5$. Scheffe comparisons within the delayed feedback condition indicated that C2/C1 values in the classroom setting were higher than those observed in the laboratory, and that I2/C1 values in the classroom were lower than those observed in the laboratory, all $p < .001$, with these outcomes likely reflecting the lower salience of affirming correct responding in a laboratory setting. Scheffe comparisons within the immediate feedback condition indicated no differences in probability values between the two test settings, all $p > .5$, and that C2/I1 and I2/I1 values in both test settings were higher when iterative responding was permitted, all $p < .005$.

Discussion

There were no differences in any dependent measure as a function of sex of participant, with a similar lack of differences between the control (Scantron form) and experimental groups on SAT scores, foreign language placement test scores, current and past performance in foreign language courses, GPA, and current classroom performance, and thus the outcomes described cannot be attributed to differences in either general intellectual or course-specific skills. The control and experimental procedures used in the present study under laboratory conditions were identical to the distribution of lecturing and assessment under classroom conditions in our prior studies (Dihoff et al., 2003, 2004), and the results of the participation motivation comparisons suggest that participants were sensitive to the operative reward systems and that immediate feedback coupled with iterative responding is more effective than delayed feedback for the correction of errors.

The larger probability values for C2/I1, together with greater accuracy of correctly identifying test items that had been repeated and their initial

errors on these items, suggests that the self-corrective information provided by immediate feedback and the opportunity to engage in iterative responding was incorporated by participants into their self-corrective mechanisms. Analogous outcomes were not observed for participants provided with delayed feedback or control (Scantron form) procedures as their initial errors were more likely to be repeated (I2/I1) whereas their initially correct (C2/C1) responses were less likely to be repeated. The former outcome, the failure to recall and to correct initial errors, supports some aspects of the delay-retention effect; however, the forgetting of initially correct responses would negate any potentially beneficial effects of delayed feedback. It is unfortunate that when feedback is provided in the classroom, it is typically delayed, and that when Scantron forms are used, it is altogether absent. The present results add to a growing body of literature indicating that immediate feedback is more effective than delayed feedback and directly question the common belief that immediate feedback is more effective in the classroom whereas delayed feedback is more effective in the laboratory.

References

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

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.

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

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.

BROSVIC, G. M., EPSTEIN, M. L., COOK, M. J., & DIHOFF, R. E. (2005). Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: Learning in the classroom. *The Psychological Record*, 55, 401-418.

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.

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., 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.

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

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.

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

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

WELLS, J. C. (1992). *Concise Esperanto and English Dictionary*. New York: McGraw-Hill.