JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1994, 27, 483–491 NUMBER 3 (FALL 1994)

EFFECTS OF IMPOSED POSTFEEDBACK DELAYS IN PROGRAMMED INSTRUCTION

JOHN CROSBIE AND GLENN KELLY

DEAKIN UNIVERSITY

Imposed postfeedback delays promote discrimination training; the present experiments determined whether they also improve performance in programmed instruction. In two experiments, college students completed 45 sets of Holland and Skinner's (1961) programmed text on behavior analysis in a computerized format in a three-component multiple schedule. In Experiment 1, the conditions were (a) no delay between questions, (b) a 10-s delay after each question (noncontingent delay), and (c) a 10-s delay after each question answered incorrectly (contingent delay). Noncontingent delay produced better performance than no delay and contingent delay. To determine whether performance increased in the noncontingent delay condition because subjects studied the material during delay periods, Experiment 2 tested three conditions: (a) no delay between questions, (b) a 10-s delay after each question (noncontingent delay), and (c) a 10-s delay after each question with the screen blank during the delay period. Noncontingent delay produced better performance than no delay, but there was no difference in performance between no delay and noncontingent delay with blank screen. Hence, noncontingent delay improved performance because students used delay periods to study. Furthermore, subjects preferred noncontingent delay to the other conditions, and session time increased only slightly.

DESCRIPTORS: computer-based instruction, programmed instruction, multiple schedule, delay, college students

Programmed instruction is discrimination training: Learners examine a stimulus, emit a response, receive feedback, and consequently make appropriate responses in the presence of novel stimuli (Skinner, 1954, 1968). Several researchers have obtained superior discriminated performance with college students when delays were imposed after feedback on concept-identification tasks (e.g., Bourne, 1957; Bourne & Bunderson, 1963; Bourne, Guy, Dodd, & Justesen, 1965; White & Schmidt, 1972). In a typical study (Bourne, 1957), subjects were presented with geometric patterns that had seven dimensions, but only two were necessary for correct classification. Subjects reported which combination of dimensions was correct (e.g., large square, large triangle, small square, or small triangle). After all stimulus patterns had been presented, subjects pressed one of four unlabeled keys that corresponded to the four possible combinations of relevant dimensions. Following a delay between 0 s and 8 s, a light was illuminated above the correct key. It was found that discrimination performance increased as prefeedback delay decreased. Because prefeedback delay plus postfeedback delay always totaled 10 s (the constant intertrial interval), it was not clear whether improved performance was produced by decreasing prefeedback delay or increasing postfeedback delay. To answer this question, Bourne and Bunderson (1963) used a between-groups design with complexity (one and five irrelevant dimensions), prefeedback delay (0 s, 4 s, and 8 s), and postfeedback delay (1 s, 5 s, and 9 s) as independent factors. Prefeedback delay had no significant effect on performance, but postfeedback delay produced a linear decrease in error rate. Similar results were obtained with longer delays, and the consistent finding was that moderate postfeedback delays (i.e., 10 s to 15 s) produced better performance than longer and shorter delays (Bourne

The present experiments were supported by grants from Deakin University's Internal and Science Research Budgets. We are grateful to Linda Blackman, Suzanne Brown, Sezen Ildiri, and Jo Taraborrelli for collecting the data in Experiment 1; Andrew Eltringham for helping to prepare test items; and Phil Chase and Mike Perone for helpful comments on an earlier version of this manuscript.

Reprint requests should be sent to John Crosbie, who is now at the Department of Psychology, West Virginia University, Morgantown, West Virginia 26506-6040.

et al., 1965; White & Schmidt, 1972). Furthermore, if the stimulus was available during the delay interval, there was an additional increase in performance.

Postfeedback delays have also improved performance on demanding computer tasks. In one study (Boehm, Seven, & Watson, 1971), graduate computing students received a computerized map of an urban area plus information concerning the location and frequency of emergencies, and were given 2 hr to find the optimal location for emergency hospitals. For some subjects, the keyboard was locked for 5 min after each response (and there was no opportunity to correct responses), whereas others had free access to the computer. The 5-min lockout group outperformed the free-access group by finding the most effective solution with fewer computer interactions. In another study (Dannenbring, 1983), programmers debugged a program as fast and accurately as possible within a 30-min period. Corrections were made by retyping the line that contained an error. Postfeedback delays between 0 s and 14 s were imposed after each correction was entered. Delay did not affect performance, but longer delays produced fewer line deletions. Because subjects often worked during delay periods, it seems that delays provided an opportunity to study the problem and, therefore, make more efficient corrections (Salmoni, Schmidt, & Walter, 1984). Delays may also have been aversive. When punishment is employed in discrimination training, subjects respond more slowly and attend more carefully to stimulus features (Barlow, 1933; Muenzinger, 1934a, 1934b), and discriminated performance improves (Getsie, Langer, & Glass, 1985; Harris & Tramontana, 1973; F. Miller, Moffat, Cotter, & Ochocki, 1973). Consequently, performance may have improved because postfeedback delay functioned as time-out.

Postfeedback delay may also serve another function. During a course that used computerized testing (Crosbie & Kelly, 1993), students frequently made mistakes because they responded too quickly. To the extent that test completion and escape from class functioned as reinforcers, presentation of each question was a conditioned negative reinforcer on a ratio schedule. Faster responding produced the conditioned reinforcer more quickly, so rapid responding (racing) became the dominant pattern. One way to avoid racing may be to impose postfeedback delay. Racing would never be followed immediately by a conditioned reinforcer, and, therefore, would not increase because of adventitious reinforcement (Sidman, 1960). This use of postfeedback delay is directly analogous to the common technique of imposing a delay between components of a multiple schedule.

The aim of the present experiments was to assess the effectiveness, efficiency, and social validity (Wolf, 1978) of postfeedback delay presented both contingently and noncontingently on incorrect responses. Both types of delay were used to promote attention to the material and prevent racing. Response-contingent delay was also used to examine whether it functions as a punisher.

Apart from the experimental manipulations, the present experiments have important differences from typical studies of programmed instruction. Previous studies have often presented only a small number of frames (Holland, 1961; Tudor & Bostow, 1991), have used programs of dubious quality (Holland, 1967; Kemp & Holland, 1966; Tobias, 1973), have used a groups design to assess differences between conditions (Holland, 1961; Kulik, Cohen, & Ebeling, 1980), and have been unable to manipulate contingencies precisely because of the presentation medium (Tudor & Bostow, 1991). In contrast, the present experiments used 1,711 frames of a well-tested text (Holland & Skinner, 1961), a single-subject design to assess differences between conditions, and a computer to control presentation of material and timing. These are important modifications. If only a few frames are presented in one brief session, then it is unlikely that large differences between conditions will emerge. Furthermore, if group-based analyses are performed on these data, intersubject variability may further obscure differences between conditions. The present experiments avoided these problems by presenting all experimental conditions each session in a multiple schedule. A computer provided precise presentation of materials, control of contingencies, and recording. Furthermore, the present experiments assessed not only the percentage of frames answered correctly

and time required to complete each set (i.e., effectiveness and efficiency), but also subjects' satisfaction with the procedures (i.e., social validity).

EXPERIMENT 1

METHOD

Subjects and Setting

Four female college students (H21, H22, H23, and H24), aged 20 to 35 years, served as subjects as a course requirement of an undergraduate research project. All subjects had completed one course in applied behavior analysis (based on L. Miller, 1980; see Crosbie & Kelly, 1993, for details), but had not seen the material used in the present study (Holland & Skinner, 1961). In the applied behavior analysis course, H21 had received a grade of C, H22 and H23 had received a B, and H24 had received an A. All subjects were naive with respect to the aims and rationale of the present experiment and had not participated previously in similar studies.

The study was conducted in a section of a laboratory (180 cm by 200 cm) partitioned with two room dividers (each 123 cm by 180 cm). Subjects sat at a desk with a computer and monitor at the back, a keyboard in front, and exhibits to the right.

Apparatus

Computer equipment. A Samsung® S330 (IBM® PC-compatible) computer with 640K RAM, a 20-Mb hard disk, and an ECM 36-cm EGA color monitor were used. Before each instructional set was presented, all questions and correct answers for that set were read from the hard disk into memory. This ensured that there was no unprogrammed delay between questions while text was being read, and facilitated a quick return to questions that were answered incorrectly (see the Lesson Format section below for further discussion of this feature).

Characters were displayed by assembly language routines (Crosbie, 1990); consequently, text appeared on the screen instantaneously (i.e., <10 ms). Text was displayed in primary colors on a black background; questions were white, and prompts were yellow. Each question was presented in the center of the screen, the subject's response appeared below the question, and the correct response appeared below the subject's response. Instructions were presented at the bottom of the screen. When a delay was programmed, a bar was shown at the bottom of the screen with the instruction "Please wait until the bar is gone." This bar functioned as an analogue clock because, during the delay period, the bar became progressively shorter until it was gone. While any portion of the bar was on the screen, the program would not read input from the keyboard. At the end of the delay period, the keyboard buffer was cleared to remove characters typed during this period.

Lesson format. All lessons had a linear format (Skinner, 1954, 1968); that is, questions were presented in a predetermined sequence. Questions had one or more words missing, and subjects typed the missing words. This format was used because unfamiliar technical material is learned better with constructed responses (Tobias, 1973).

With Skinner's original procedures (1968), after the final question had been answered for a topic, the teaching machine returned to questions that were answered incorrectly. This review feature was incorporated in the present program because it has increased performance by 10% (Holland & Porter, 1961), and we wanted to assess how postfeedback delays affect performance on an optimal implementation of programmed instruction.

Exhibits. Each subject studied the first 45 sets of the Holland and Skinner (1961) text. Explanatory material (e.g., a graph or figure) is presented at the start of several sets in the textbook. This material was photocopied and compiled into a booklet that was placed on the table during sessions. Computer-presented sets were, therefore, identical to the original text in all ways except method of presentation and experimental contingencies.

Design and Procedure

Each subject received the following conditions in a multiple schedule (Hersen & Barlow, 1976): (a) no delay between questions (no delay), (b) a 10-s delay after each question (noncontingent delay), and (c) a 10-s delay after questions answered

incorrectly (contingent delay). Each condition had a discriminative label (e.g., "No delay between questions") displayed in the top left corner of the screen throughout the set. Sets varied in length between 25 and 75 frames, so time required to complete a set was variable. Three sets were completed each session, and sets took a mean of 20 min to complete. Hence, sessions lasted approximately 1 hr, and 45 sets were completed in 15 sessions. For the first five sessions, 15 sets were presented with no delay to establish a baseline against which to compare the effects of the two delay conditions and to adapt subjects to the equipment and procedures. Following baseline, the multiple schedule operated for the final 10 sessions. During this phase each set in a session was presented with a different condition, and condition sequence varied randomly across sessions.

Instructions. Before the first session each subject read the following instructions:

> In this experiment you will use computer-assisted instruction to learn about an important area of psychology: the experimental analysis of behavior and its application. You will cover three topics per session, and there will be one session every day (Monday to Friday) for 15 sessions; each session will last approximately 1 hour.
>
> Each topic has between 25 and 75 questions and, for each, you will supply one or two missing words. This study has three experimental conditions: (1) no delay between questions, (2) a 10-second delay between all questions, and (3) a 10-second delay only after questions answered incorrectly. During a lesson, only one of these conditions will apply (shown in yellow at the top of the screen). For the first five sessions you will receive only Condition 1, but for all other sessions you will receive all three conditions. You must answer all questions correctly before you finish a topic, so when you have answered the final question, the program will take you back to the questions you answered incorrectly, and will continue this process until all questions have been answered correctly. At the end of each lesson, you will be asked a question concerning the lesson, and you will type a number between 1 and 9 to show your answer.
>
> Between sessions do not study the material or talk to other people about the topics; this might ruin the experiment, and make the results uninterpretable.

After they had read the instructions, subjects interacted with a short demonstration lesson that showed the following aspects of the procedure: (a) the types of question that can be asked (i.e., when one word "_____," two words "_____ _____," or several words "***" are missing, or when a technical term is required "TT"; see Holland & Skinner, 1961, for further details of these conventions); (b) how to type an answer and erase typing errors with the BACKSPACE key; (c) how to score the answer by typing C if it is correct or I if it is incorrect; (d) how the program returns to questions that were answered incorrectly; and (e) how to score satisfaction with the condition. At the end of each set, the program asked subjects how satisfied they were with the experimental condition in operation during the set. Subjects typed a number between 1 and 9, where 1 was *not at all satisfied* and 9 was *extremely satisfied.*

Typical session. At the start of a session, the second author inserted the subject's key disk (which contained her code name and session number) and ran the experimental program. If there was an exhibit for the current set, the subject read it and referred to it later as required. When the subject was ready to start the set, she pressed any key. Question 1 appeared immediately in the center of the screen. When she was ready, she typed her response followed by the ENTER key. Before she pressed ENTER, she could correct her response with the BACKSPACE key; after ENTER was pressed, however, there was no opportunity to change the response. Immediately after she pressed ENTER, the correct response was displayed below her response. It was clear which was her response and which was the correct response because they were

labeled and shown in different colors. She was then instructed to "Type C if your answer is correct or I if it is incorrect." If she typed C, the number of correct responses (continuously displayed in the score box in the top right corner of the screen) was incremented by one. If she typed I, the number of incorrect responses (also displayed in the score box) was incremented by one, and that frame was tagged for subsequent review. The program did not continue until the subject typed C or I. The subject's responses, correct responses, and scoring (i.e., correct or incorrect) were transferred to a disk file for verification by an experimenter; all scoring was accurate.

During no delay, immediately after the response had been scored, the subject was instructed to "Type any key to continue." At this stage, the question, response, and correct response were all displayed and remained on the screen until a key was pressed. When a key was pressed, the screen was cleared immediately, and the next question was presented. During noncontingent delay, immediately after the response had been scored (and regardless of how it was scored), the delay bar was displayed for 10 s (during which time typing had no programmed consequences); then the subject could type any key to continue. Observation and debriefing reports suggest that subjects typically spent the delay period reading the question, response, and correct response. During contingent delay, immediately after a response had been scored as incorrect, the delay bar was displayed for 10 s; then the subject could type any key to continue. If a response was scored as correct, however, the delay bar was not displayed, and the subject could type any key to progress to the next frame. When all questions in the set had been answered, the program returned to frames that had been scored as incorrect. This review continued until all frames were scored as correct. The subject then answered the question "How satisfied were you with the experimental condition, 1 = not at all satisfied and 9 = extremely satisfied," by typing a number between 1 and 9. If she had not completed three sets during the session, she started the next set; otherwise, she left the laboratory. Apart from the 10-s delay period, no time constraints were imposed. At each stage she could proceed as quickly or slowly as she wished.

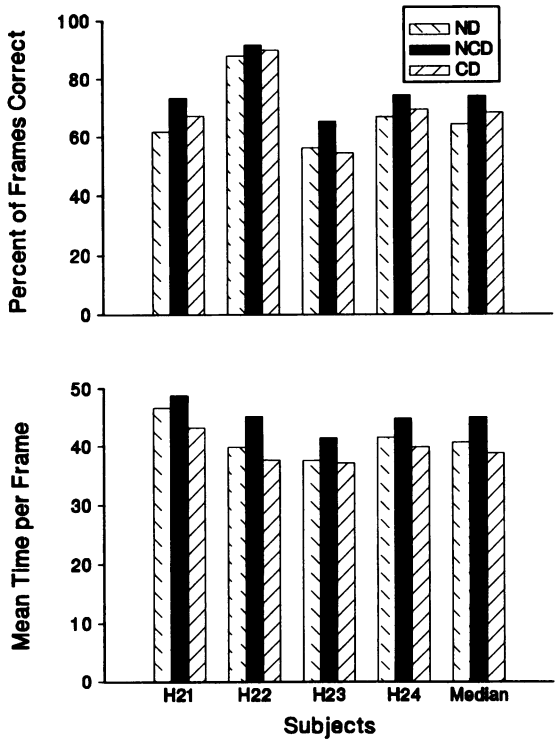

Figure 1. Experiment 1: Percentage of frames answered correctly and mean time per frame for each condition (ND = no delay, NCD = noncontingent delay, and CD = contingent delay; see text for details) and subject. Median results across subjects are also shown.

Results and Discussion

Because sets differed in terms of number of frames and difficulty, percentage of frames answered correctly per set differed markedly within experimental conditions. We therefore discuss consistency of differences among conditions, and show the total percentage of frames answered correctly for each condition.

Across sessions, all subjects correctly answered a greater percentage of frames during noncontingent delay (NCD) than during no delay (ND) and contingent delay (CD), which did not differ significantly. Of the 40 multiple-schedule sessions (10 per subject), NCD had a greater percentage correct

than ND in 29 sessions, and a greater percentage correct than CD in 28 sessions. Both differences are statistically significant ($p < .01$, sign test; Siegel, 1956). CD had a greater percentage correct than ND in 19 of the 40 sessions (not statistically significant). Overall, there was no consistent difference between CD and ND; therefore, CD was not a punisher.

For the 1,112 frames presented during the multiple-schedule phase, Figure 1 shows, for each subject and condition, the percentage of frames answered correctly and mean time per frame (in seconds). NCD produced better performance than the other conditions for all subjects. Median results across subjects show that NCD produced 10% more correct responses than ND and 6% more correct responses than CD. When performance during ND was high, NCD improved performance only slightly (see H22), presumably because there was little room for improvement. When ND performance was lower, however, NCD produced a much larger improvement (see H23 and H21). The importance of delay is highlighted by the finding that median performance was best for the condition in which delay was most frequent (NCD), worst when delay was not employed (ND), and intermediate when delay was intermittent (CD).

Figure 1 also shows that NCD had a longer mean time per frame than the other conditions for all subjects. Although 10 s was added to each frame in NCD, median results show that NCD took only 4 s per frame longer than ND and 6 s per frame longer than CD. Either subjects responded more slowly during ND than they did during baseline, or they responded more quickly during NCD. Mean time in ND did not change from baseline, so subjects must have worked more quickly during NCD than during ND.

Three subjects (H21, H23, and H24) consistently reported more satisfaction with NCD than the other conditions. The 4th subject (H22) had an NCD satisfaction rating similar to those of the other subjects, but gave ND and CD an atypically high rating. Median results across subjects show that all conditions had a moderate satisfaction rating (approximately 5 on a 9-point scale), and NCD was rated 0.5 points more highly than ND and 0.2 points more highly than CD. These results show that subjects were not distressed when delays were imposed.

EXPERIMENT 2

Experiment 1 found that NCD produced superior performance. NCD has two aspects that could have achieved this: an additional opportunity to inspect material, and a delay that might have reduced racing. The present experiment, which provided a systematic replication of Experiment 1, assessed both possibilities.

Subjects received the following conditions in a multiple schedule: (a) no delay, (b) noncontingent delay, and (c) noncontingent delay with the screen blank during the delay period (NCB). If extra inspection time, instead of delay, improved performance, then NCD would be superior to both NCB and ND, which would produce similar performances. Alternatively, if delay was the important factor, then NCB and NCD would have similar results, and both would be superior to ND.

In the present experiment, 4 undergraduates (1 male and 3 female), with no background in behavior analysis, were paid $6 for each session completed. Thus, the generality of the results obtained in Experiment 1 was assessed with subjects who differed in terms of background knowledge and motivation.

Results and Discussion

NCD produced a greater percentage correct than ND for 30 of the 38 sessions in which the different conditions produced different results ($p < .01$, sign test; 2 of the 40 comparisons were discarded because the sign test cannot accommodate tied scores); all subjects showed this pattern. This result replicated Experiment 1. NCD also produced a greater percentage correct than NCB for 27 of the 38 sessions, and each subject showed this pattern. There was no difference, however, between NCB and ND.

Figure 2 shows that NCD produced better performance than the other conditions for all subjects.

Median results across subjects show that NCD produced 9% more correct responses than ND and 8% more correct responses than NCB. The difference between NCD and ND is very similar to the result obtained in Experiment 1 (see Figure 1). ND and NCB produced virtually identical results. These results support the extra inspection time proposition and cast doubt on the racing hypothesis.

Figure 2 also shows that NCB had a longer mean time per frame than the other conditions for all subjects, but NCD and ND did not differ consistently. Median results show that NCB took 11 s longer than both NCD and ND. Hence, NCD did not increase session time for all subjects.

NCD had a moderate to high satisfaction rating for all subjects (median of 6; range, 5 to 9 on a 9-point scale). Median results show that NCD produced 0.6 points greater satisfaction than ND and 1.0 points greater satisfaction than NCB.

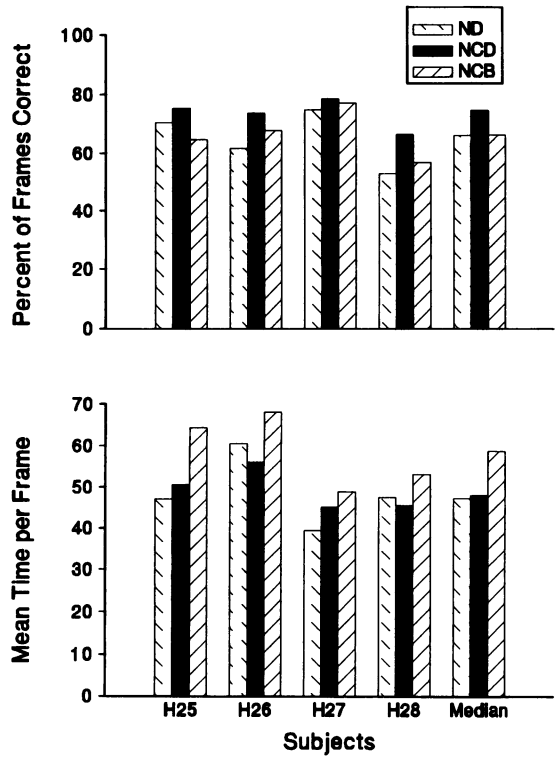

Figure 2. Experiment 2: Percentage of frames answered correctly and mean time per frame for each condition (ND = no delay, NCD = noncontingent delay, and NCB = noncontingent delay blank screen; see text for details) and subject. Median results across subjects are also shown.

GENERAL DISCUSSION

The present experiments assessed whether postfeedback delays improved performance because they provided additional study time, reduced racing, or punished incorrect responses. Additional study time was found to be important, but racing and punishment were not. NCD seems to improve performance because it forces students to spend greater periods of time on task, which frequently leads to higher achievement (Tobias, 1973). CD did not function as a punisher; perhaps the 10-s delay was too short. Future studies should try CD with longer delays or response cost (i.e., point or money loss contingent on incorrect responses). Response cost may be preferable because it would not increase session time and would permit a larger range of potential punishers. Although the racing hypothesis has a reasonable theoretical basis, it was not supported by the present data.

In ND, after a subject had scored an answer (i.e., typed C or I), the question, answer, and correct answer were all displayed until a key was pressed, then the next question was displayed. Consequently, after feedback was provided, subjects in any condition could read the question and correct answer for 10 s, as they did in NCD, and obtain the same benefits. In other words, in ND subjects were not obliged to proceed to the next frame immediately, and could study material as much as they did in NCD. The important result is that in ND subjects *did not* spend additional time studying materials. Debriefing reports showed that subjects were motivated by negative rather than positive reinforcement. Participating in the experiment was not intolerably aversive, but neither was it as positively reinforcing as things subjects could do outside the laboratory. Programmed instruction is normally considered to be based on positive reinforcement with no aversive consequences. In most classroom and experimental situations, however, performance is based substantially on escape and avoidance. If substantial positive reinforcement were available, perhaps the nature of the task would

change, but this possibility seems unlikely for most schools and laboratories in the near future.

Given that 10% is the target error rate for the Holland and Skinner material (Holland, 1960; Holland & Doran, 1973), the present subjects' error rates may seem too high. There are, however, five points that should be considered when making such an assessment. First, the materials were developed and tested with outstanding students (at Harvard and Radcliffe circa 1960) who were very competitive and were assigned grades according to their performance on the material (J. G. Holland, personal communication, April 28, 1992). Present subjects were less able academically and less motivated. Second, the goal of 10% error is for a frame, not a complete set. That is, instructional designers will modify a frame if more than 10% of students answer it incorrectly (and Holland did that with the present material). Whether all students will correctly answer 90% of frames in a set is a different question, one that was rarely addressed during program development. Third, for experimental purposes sessions were often long (i.e., the multiple schedule required that three sets were completed per session); fatigue may have impaired performance. Fourth, no external reinforcement was provided. In our recent experiments in which subject payment was contingent on performance, NCD was superior to ND, and the percentage correct increased greatly. Finally, in the most recent study of computerized programmed instruction (Tudor & Bostow, 1991), subjects in the condition most similar to the present procedures (Group 5) correctly answered between 30% and 70% of frames, with a median of 52%.

Although the present results are significant, it is important that they be kept in perspective. The research did not include a complete technology, but rather was an experimental analysis of the effects of postfeedback delays within a rigorous laboratory context. Delays of 10 s were selected arbitrarily (although there is support for the efficacy of such durations in the concept-identification literature discussed above) and may not provide optimal conditions. Indeed, given that subjects in Experiment 1 worked 6 s per frame more quickly during NCD (i.e., although 10 s was added to each frame, they took only 4 s longer per frame), it would be instructive to use a 6-s noncontingent delay to determine whether performance would be improved with no additional time required. Similarly, three sets were presented each session for experimental purposes, and such long sessions may not be optimal. The present results probably underestimate what can be achieved with noncontingent postfeedback delays, and future research should try to refine the procedure to determine its maximum potential. The current research showed that a small environmental adjustment can significantly affect academic performance, with few, if any, undesirable side effects; behavior analysts have an important role to play in assessing how to improve effectiveness, efficiency, and consumer satisfaction with teaching.

REFERENCES

Barlow, M. C. (1933). The influence of electric shock in mirror tracing. *American Journal of Psychology, 45,* 478–487.

Boehm, B. W., Seven, M. J., & Watson, R. A. (1971). Interactive problem-solving—An experimental study of "lockout" effects. *AFIPS Conference Proceedings, 38,* 205–210.

Bourne, L. E., Jr. (1957). Effects of delay of information feedback and task complexity of the identification of concepts. *Journal of Experimental Psychology, 54,* 201–207.

Bourne, L. E., Jr., & Bunderson, C. V. (1963). Effects of delay of informative feedback and length of postfeedback interval on concept identification. *Journal of Experimental Psychology, 65,* 1–5.

Bourne, L. E., Jr., Guy, D. E., Dodd, D. H., & Justesen, D. R. (1965). Concept identification: The effects of varying length and informational components of the intertrial interval. *Journal of Experimental Psychology, 69,* 624–629.

Crosbie, J. (1990). The Microsoft mouse as a multipurpose response device for the IBM PC/XT/AT. *Behavior Research Methods, Instruments, & Computers, 22,* 305–316.

Crosbie, J., & Kelly, G. (1993). A computer-based personalized system of instruction course in applied behavior analysis. *Behavior Research Methods, Instruments, & Computers, 25,* 366–370.

Dannenbring, G. L. (1983). The effect of computer response time on user performance and satisfaction: A preliminary investigation. *Behavior Research Methods & Instrumentation, 15,* 213–216.

Getsie, R. L., Langer, P. D., & Glass, G. V. (1985). Meta-analysis of the effects of type and combination of feedback on children's discrimination learning. *Review of Educational Research, 55,* 9–22.

Harris, L. M., & Tramontana, J. (1973). Discrimination learning of retarded children as a function of positive reinforcement and response cost. *American Journal of Mental Deficiency, 78,* 216–219.

Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs: Strategies for studying behavior change in the individual.* New York: Pergamon.

Holland, J. G. (1960). Teaching machines: An application of principles from the laboratory. *Journal of the Experimental Analysis of Behavior, 3,* 275–287.

Holland, J. G. (1961). Evaluating teaching machines and programs. *Teachers College Record, 63,* 56–65.

Holland, J. G. (1967). A quantitative measure for programmed instruction. *American Educational Research Journal, 4,* 87–101.

Holland, J. G., & Doran, J. (1973). Instrumentation of research in teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 286–317). Chicago: Rand McNally.

Holland, J. G., & Porter, D. (1961). The influence of repetition of incorrectly answered items in a teaching machine program. *Journal of the Experimental Analysis of Behavior, 4,* 305–307.

Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior.* New York: McGraw-Hill.

Kemp, F. D., & Holland, J. G. (1966). Blackout ratio and overt responses in programmed instruction: Resolution of disparate results. *Journal of Educational Psychology, 57,* 109–114.

Kulik, J. A., Cohen, P. A., & Ebeling, B. J. (1980). Effectiveness of programmed instruction in higher education: A meta-analysis of findings. *Educational Evaluation and Policy Analysis, 2,* 51–64.

Miller, F. D., Moffat, G. H., Cotter, P. D., & Ochocki, T. E. (1973). Verbal reinforcement combinations, task complexity, age and sex in children's discrimination learning. *Perceptual and Motor Skills, 37,* 531–534.

Miller, L. K. (1980). *Principles of everyday behavior analysis* (2nd ed.). Monterey, CA: Brooks-Cole.

Muenzinger, K. F. (1934a). Motivation in learning: I. Electric shock for correct response in the visual discrimination habit. *Journal of Comparative Psychology, 17,* 267–277.

Muenzinger, K. F. (1934b). Motivation in learning: II. The function of electric shock for right and wrong responses in human subjects. *Journal of Experimental Psychology, 17,* 439–448.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95,* 355–386.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* Tokyo: McGraw-Hill Kogakusha.

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24,* 86–97.

Skinner, B. F. (1968). *The technology of teaching.* New York: Appleton-Century-Crofts.

Tobias, S. (1973). Review of the response mode issue. *Review of Educational Research, 43,* 193–204.

Tudor, R. M., & Bostow, D. E. (1991). Computer-programmed instruction: The relation of required interaction to practical application. *Journal of Applied Behavior Analysis, 24,* 361–368.

White, R. M., & Schmidt, S. W. (1972). Preresponse intervals versus postinformative feedback intervals in concept identification. *Journal of Experimental Psychology, 94,* 350–352.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203–214.

Received June 1, 1993
Initial editorial decision August 2, 1993
Revisions received November 2, 1993; November 23, 1993
Final acceptance December 22, 1993
Action Editor, Anthony J. Cuvo