JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1991, 24, 361–368 NUMBER 2 (SUMMER 1991)

COMPUTER-PROGRAMMED INSTRUCTION: THE RELATION OF REQUIRED INTERACTION TO PRACTICAL APPLICATION

ROGER M. TUDOR AND DARREL E. BOSTOW

UNIVERSITY OF SOUTH FLORIDA

Computers were used to evaluate the effects of supplying answers to programmed instruction frames. A group experimental design compared passive reading, covert responding to frame blanks, and actively typing answers to blanks with and without immediate confirmation of correctness. Effects of a 315-frame program, teaching elements of programmed instruction design, were evaluated by analyzing answers to posttest generalization questions and an application test. Results strongly supported the effectiveness of requiring the student to supply fragments of a terminal repertoire while working through a program. Students who could either covertly respond to frame blanks or who were required to type frame answers performed significantly better on the frame generalization posttest and, more importantly, carefully followed program rules when preparing elements of a new instructional program.

DESCRIPTORS: academic behavior, programmed instruction, computer-based instruction, instructional design, generalization

Programmed instruction does not yet play a significant role in our public schools (Skinner, 1984). Skinner was referring to carefully composed, sequentially ordered, brief presentations of verbal material requiring constructed responses (Skinner, 1968). However, the advent of microcomputers and educational software has revived interest in automated, if not programmed, instruction. Unfortunately, operant learning principles, which are part of well-programmed instruction, are not easily identified in present-day software.

Until the 1980s, the "hardware problem" had not been solved, and most research with programmed instruction was accomplished via paper presentation. The advantages of strict stimulus structuring possible only with machine delivery were not part of most research conditions. It can be argued that a fair evaluation of programmed instruction, as initially described by Skinner, is not possible unless such precise structuring of contingencies is accomplished.

Reviews evaluating past research in programmed instruction do not support convincingly the necessity of overt-constructed responding (Anderson, 1967; Silverman, 1978; Tobias, 1973). There may be many reasons for this. Instructional programs differed from one experiment to another on many dimensions and were often apparently quite brief. Incomplete descriptions of programs prevent accurate comparison. Program accomplishments were usually evaluated using multiple choice or fill-in-the-blank test questions. When active responding produced greater learning, the difference between groups was usually quite small, and the criterion for significance was nearly always statistical rather than practical. The quality of programs used in research has been questioned (Holland & Kemp, 1965; Kemp & Holland, 1966; Vargas, 1986). Issues concerning prior familiarity with subject matter have been raised (Abramson & Kagen, 1975; Tobias, 1969) as has the relation of emotional conditions (Tobias & Abramson, 1971). It is clear that rules that might guide the designer of better instructional software cannot be easily extracted from past research.

The measurement index used to evaluate the effects of programmed instruction is an especially important issue. Is the difference between a pretest and a posttest performance in the form of answers to questions (i.e., the intraverbal, see Skinner, 1957)

This research was conducted in partial fulfillment of the requirements for the PhD degree by the first author, who is now affiliated with Georgia Southern University.

Requests for reprints should be sent to Roger M. Tudor, Department of Psychology, Landrum Box 8041, Georgia Southern University, Statesboro, Georgia 30460.

the appropriate measure? Although the written answer to a written question is the ubiquitous measure of "knowledge" in our institutions, this form of measurement may be a superficial and incomplete representation of what has been learned. It may be that important collateral and practical effects of interactive programmed instruction have been missed.

With the exception of the study by Avner, Moore, and Smith (1980) little research evaluating the effects of automated instruction has measured broader effects on a student's repertoire. Avner et al. measured whether chemistry students performed better in the laboratory after they had received computer-presented instruction. However, they gave few details about the specific measurement of application skills in their study.

The present experiment evaluated the importance of active student responding while using the microcomputer to deliver the contingencies of programmed instruction. The automated instructional program used as the foundation of the experiment was designed to teach a content specialist the steps necessary in the development of effective programmed instruction. The present research measured aspects of programmed instruction design and focused on student products.

METHOD

Subjects and Setting

Seventy-five undergraduates (18 males, 57 females) from an educational psychology course at the University of South Florida served as subjects. Most were juniors and seniors in degree programs preparing them to be teachers. Only 1 of those surveyed reported having previously taken a course in behavior analysis. All students received a grade of "pass" for participating in this study as part of the course requirement. Students did not receive letter grades in relation to experimental group differences. All students were randomly assigned to the experimental groups.

The experimental room contained all computer equipment and software. Up to 7 students at a time worked through the instructional programs on separate computers. Members of the various experimental groups worked side by side but were nearly always working at different points in the program.

Apparatus, Programs, and Materials

The experiment used five identical content versions of a 315-frame instructional program. These frames covered operant conditioning and instructional design principles that related directly to the preparation of programmed instruction (Bostow, 1989). The objective of the program was to teach students how to construct effective programmed instruction. These objectives were not taught during regular class meetings.

Each IBM PC Junior® microcomputer station included a disk drive, keyboard, and color monitor. An authoring software program contained on each diskette (*Teacher Turned Author,* 1985) precisely controlled the sequence of frames and answer-contingent messages (if any). A series of three disks at each computer station contained the entire program.

Design and Procedures

The functional analysis in this study required group comparison, with its attendant difficulties in resolving power, for several reasons. Cumulative instructional effects ruled out a reversal design. A multiple baseline design could not be used because students could easily converse with each other and make comparisons, the effects of which could not be controlled or evaluated. Also, the content of the instructional program could not be broken into truly independent and equal segments enabling sequential treatment comparisons. Repetition was deliberately built into the entire program to sustain response strength. Students' entering skills would undoubtedly produce changing difficulty between program content frames; this could vary in unknown ways, making comparison between segments difficult. Experimental condition order effects might easily occur if the enduring results of less effective conditions generalized to performance on subsequent frames, confusing terminal test and task performances in unknown ways.

Subjects in Group 1 received computer-presented programmed frames without blanks, a condition similar to reading prose from a book. Students read each frame and simply tapped the ENTER key to produce the next video frame.

Subjects in Group 2 read through the same complete frames as those in Group 1 (i.e., no words were missing). As did Group 1 students, Group 2 students read each frame and selected ENTER to produce the next frame. However, to approximate more closely what might be a critical element in the remaining experimental conditions, the potential "answer word" (for fully programmed instruction) was presented at the top of each subsequent frame. In other words, each new frame began with the statement "The answer was {x}," which presented the word the student should have typed had this word been missing in the preceding frame. (This condition was included to determine whether simple reappearance of answer words might contribute to posttest response strength.) The students were instructed to read these words as they would any other in the program.

Subjects in Group 3 responded to frames with blanks. Students were instructed to read each frame and to "think" the correct answer covertly. Selecting ENTER produced the next frame with the correct answer to the previous frame at the top. This condition approximated paper-delivered programmed instruction, although computer delivery (in any of the experimental conditions) did not permit skipping backwards or ahead.

Subjects in Group 4 responded to exactly the same frames as did Group 3 subjects. However, these students were required to type their responses to the blanks. Selecting ENTER then produced the next frame with the correct answer presented at the top.

Subjects in Group 5 experienced frames identical to those of Groups 3 and 4. However, confirmation of correctness occurred immediately, contingent on entry of a student's answer. In other words, the word "correct" or "incorrect" appeared on the video screen for 2 s, depending on the entered response. The correct answer then appeared on the next frame. This final experimental condition represented what might be considered the full complement of variables commonly argued to represent orthodox programmed instruction.

Dependent variables. Each student supplied written answers to a 47-item fill-in-the-blank pretest-posttest. Each question contained a sentence with a blank to be filled in by the student. These blanks required terms that were the correct answers for program frames (for Groups 3, 4, and 5). The wording of these test questions was similar but not identical to program frames. Posttest reliability was .87 as measured by the KR-20 (Borg & Gall, 1983).

Students also applied program-presented concepts and principles to the solution of a novel instructional design problem. Each student was given a randomized list of the steps or rules necessary for washing clothing in an automatic washing machine and was asked to design and construct two programmed frames for computer presentation. Some of these steps included setting the wash cycle, sorting clothes by color, adding soap, closing or opening the lid, and starting the machine.

This applied task consisted of a *process* and a *product* component. The process component assessed several precurrent design and organizational behaviors considered necessary for writing effective programmed instruction. These were the terminal objectives of *Preparing Automated Instruction* (Bostow, 1989). Program objectives included identifying important concepts, formulating rules, ordering rules in a developmental sequence, constructing a rule matrix, selecting a deductive or inductive frame format, constructing rule and example sentences, and determining placement of review frames.

The product component assessed the degree to which each student engaged in the repertoire developed by the automated instructional program. Technical correctness of frames composed by students was judged on the basis of the presence or absence of (program-defined) critical components of programmed instruction. A rater scored each student product using a checklist, tallying whether each critical component was present or absent (see Table 1). Individual percentage scores for the prod-

Table 1
Rated Items on the Product Component of the Applied Task

1.	Does the first frame contain at least two sentences?
2.	Does the first frame contain a rule sentence?
3.	Does the first frame contain an example sentence?
4.	Have the rule and/or example sentences been edited or changed in any way?
5.	Does the first frame adhere to the format selected (deductive or inductive)? If no format was selected, check "no" column.
6.	Does the first frame contain a blank? (programmed instruction type)
7.	Is the blank in the second half of the first frame?
8.	Is the required response identical or similar to the previously identified answer? (If no answer was identified, check the "no" column)
9.	Is there a prompt in the first frame?
10.	Is there a prime for the second frame response in the first frame?
11.	Does the second frame contain at least two sentences?
12.	Does the second frame contain a rule sentence?
13.	Does the second frame contain an example sentence?
14.	Have the rule and/or example sentences been edited or changed in any way?
15.	Does the second frame adhere to the format selected (deductive or inductive)? If no format was selected, check "no" column.
16.	Does the second frame contain a blank? (programmed instruction type)
17.	Is the blank in the second half of the second frame?
18.	Is the required response identical or similar to the previously identified answer? (If no answer was identified, check "no" column)
19.	Is there a prompt in the second frame?

uct were the sum of the number of components present divided by the total possible.

The authoring program used to create the *Preparing Automated Instruction* program automatically recorded whether a student correctly or incorrectly answered each frame of the program. Records were kept of this performance for members of Group 4 and 5 who typed their answers to frames.

Specific procedures. One week before the experiment, students met in their regularly scheduled classrooms to take the fill-in-the-blank pretest and sign up for computer lab time. At the computer lab, students listened to standardized instructions describing the program and the evaluation tasks that followed. Because students used the computer differently, each was asked to attend to their own video screen. Each student maintained a time log of the number of minutes needed to complete the program (excluding breaks, if any). The experimenter prompted and informally verified student accuracy in recording stop and start times. Once he or she had begun, further instructions indicated whether the student would be passively reading through the program, mentally constructing answers, or physically typing answers to program blanks. After completing the program, each student completed the posttest and the applied task.

RESULTS

Each student's performance on a 47-item fill-in-the-blank pretest-posttest is shown in Figure 1. All students performed very poorly on the pretest; 68% of them produced only one or two correct answers. None of the pretest group means differed, $F(4, 70) < 1$. For comparison, pretest scores were ordered from lowest to highest within groups (Figure 1) and revealed no consistent relation to posttest scores.

However, the type of response made during the program influenced the subsequent production of correct posttest responses, $F(4, 70) = 3.44$, $p < .05$. Posttest group means ranged from 35% correct in Group 1 to 49% in Group 5. Students who typed answers to blanks (Groups 4 and 5) achieved an average of 14% more correct than those who read frames without blanks (Groups 1 and 2), $LSD = 9.19$. Students who were instructed to "think" answers to program blanks (Group 3) performed significantly better than Group 1 readers and approximately equal to the overt-response group members (Group 4 and 5). The effect of interaction (i.e., the necessity of responding to a blank covertly or overtly) was evaluated by combining the non-interactive performance data (Groups 1 and 2) and the interactive group data (Groups 3, 4, and 5). This comparison revealed a 13% greater gain (*M*

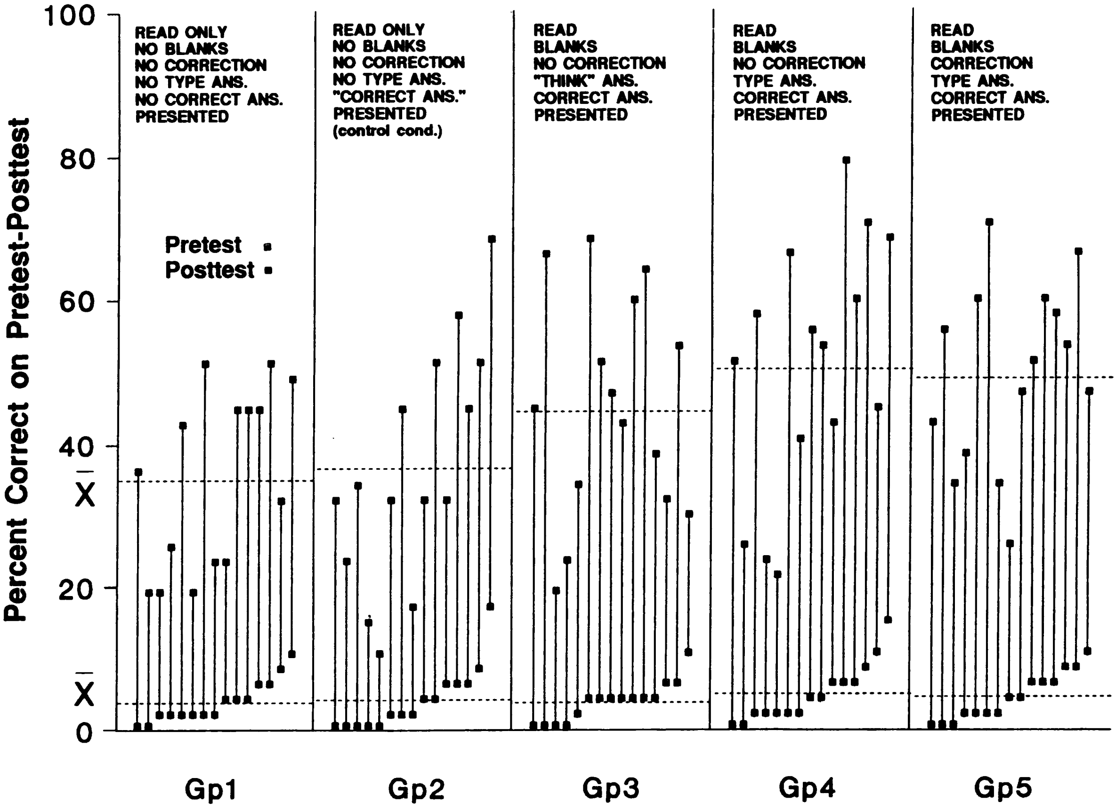

Figure 1. The percentage of correct responses made on the fill-in-the-blank pretest and posttest. Dotted lines represent group means.

= 48.2%) for active responders than for passive responders (M = 35.7%), $F(1, 70) = 12.5, p < .001$.

Variations among the instructional programs produced differences in the time required to complete the program, $F(4, 70) = 13.6, p < .001$. The means ranged from 151 min in Group 1 to 196 min in Group 5. Students who typed their answers (Groups 4 and 5) required an average of 45 min longer than those who simply read frames, LSD = 15.08 (Groups 1 and 2). Students who presumably "thought" responses to program blanks spent an average of 15 min longer than passive readers but 30 min less than students who typed their answers.

Each student's product, in the form of sample frames constructed following the posttest, was an important dependent variable in this study. Initially, the entire task was evaluated by scoring the evidence of both the steps leading to the construction of frames (process), as well as the two required frames (product). Differences resulting from the combined analysis of process and product components were not statistically significant, $F(4, 70) = 2.06$.

An evaluation of written frames alone (the product component only) showed significant group differences in ratings, $F(4, 70) = 3.17, p < .05$. Figure 2 presents individual data and group means of this applied task. Ratings of these frames differed widely, ranging from an average of 39% correct in Group 1 to 72% correct in Group 5. Students who responded overtly to program blanks produced a higher percentage of frames adhering to program-taught rules than did students reading frames without blanks, LSD = 17.36. Groups 4 and 5 differed

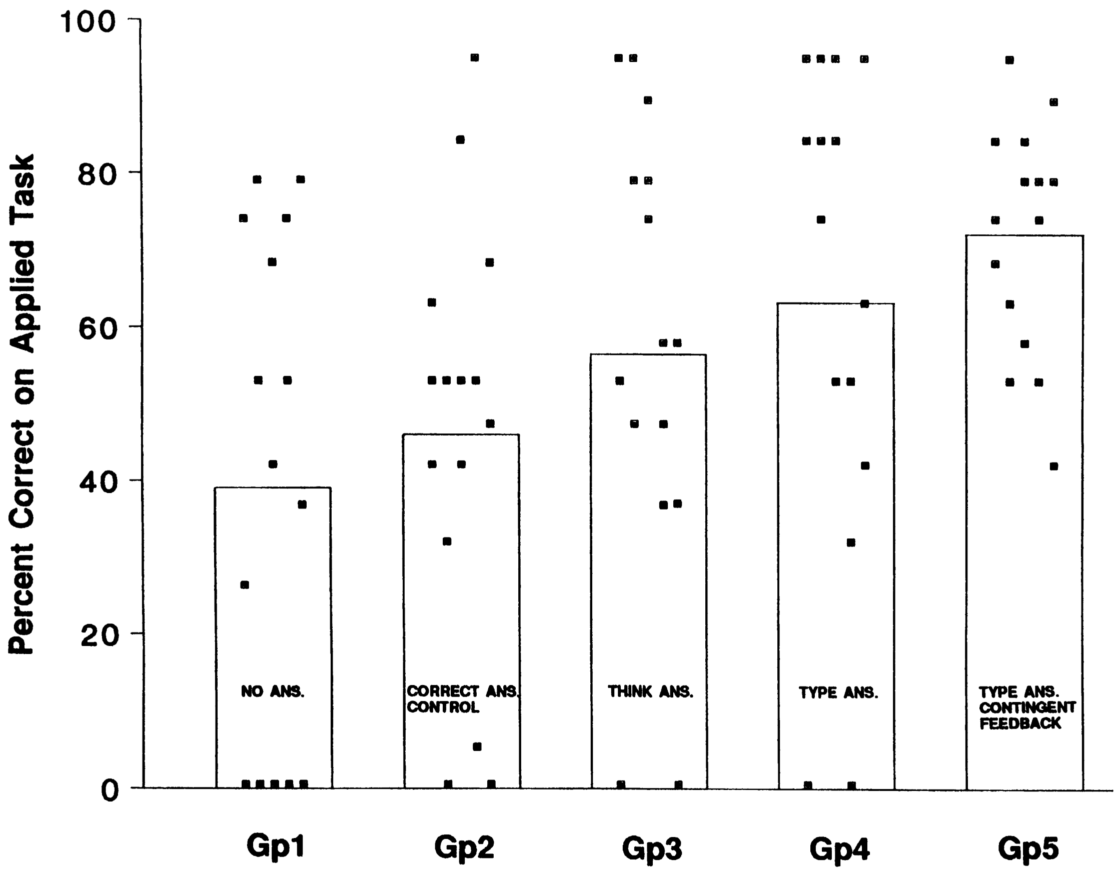

Figure 2. The percentage of program rules followed on the product component of the applied frame construction task.

significantly from Group 1, and Group 2 from Group 5. One third of Group 1 student products were scored 0% correct. Some of these students left entire pages of the product component blank. Students who responded covertly to blanks (Group 3) performed better than readers (Group 1) but statistically equal (although separated by 15%) to students who typed their responses (Groups 4 and 5). When the performance of students who simply read completed frames was combined (M = 43% correct) and compared with the three remaining groups who interactively constructed answers either covertly or overtly (M = 64% correct), the difference was substantial (21%), $F(1, 70) = 10.1$, $p < .001$. To evaluate the reliability of frame scoring, an independent rater scored a randomly selected 10% of the student products. An item-by-item comparison with the ratings of the primary observer produced 98% agreement.

The authoring program automatically recorded whether students in Groups 4 and 5 answered each frame correctly as they worked through the instructional program. Adding the contingent statement of "correct" or "incorrect" (the experimental difference between Groups 4 and 5) did not affect the nature of within-program responding, $t(28) = 1.58$. Group 4 students had an average of 74% correct responses compared to 69% for Group 5. These results do not support the need for immediate confirmation of correctness.

DISCUSSION

The present research isolated variables that contribute to the effectiveness of programmed instruction. Student interaction in the form of overt or covert answer construction resulted in a 13% better performance on a fill-in-the-blank posttest and a

21% greater adherence to rules when constructing programmed frames for a computer. These differences are comparatively larger than reported in previous programmed-instruction experiments and appear to have educational importance.

This research extends the results of Avner et al. (1980) and concludes that the most appealing aspects of well-programmed instruction may be those that generalize to practical application. Students who responded interactively not only gave more correct test responses but also could actually write better frames. Differences in the quality of student-constructed frames were larger than those measured on the fill-in-the-blank posttest. It is not known whether this was directly related to the terminal objectives of *Preparing Automated Instruction* (i.e., teaching a content specialist how to construct technically correct programmed frames) or to the imprecision of using intraverbals to predict practical application. In this study the correlation between posttest and frame construction scores was only .55, suggesting that answers to questions may not accurately reflect what is learned in interactive programmed instruction.

Students who mentally supplied answers did essentially as well as those typing their answers. It was, of course, impossible to verify that students in Group 3 did, in fact, mentally construct answers to blanks. Other researchers have noted that subjects under similar conditions will often not respond and wait to see the answer or even "peek ahead" when printed materials permit this (Kulhavey, 1977; Silverman, 1978). However, in this study all students knew they were to receive a posttest covering the program. It is quite possible that students who could simply mentally construct answers were more likely to do so because they believed their performance was being graded. Constructing answers mentally may be equally effective when subsequent verification of learning accompanies the use of programmed instruction. It is safe to assume, however, that when an answer is not always required, it will not always be constructed. One distinct advantage of making progress through a program explicitly contingent on constructed responses is that subsequent testing is usually unnecessary. (This alone might justify the expense of a microcomputer.)

The length of time taken by each student was recorded because it was easy to measure. This deceptively simple measure included many variables. Time taken included computer mechanical operation (a rather lengthy 7-s loading time per frame) as well as the time it took to physically tap a key (Groups 1, 2, and 3) or type an answer to each frame (Groups 4 and 5). The average time spent by those requested to construct answers mentally was slightly more than simply reading frames and significantly less than actually typing answers on the keyboard. It is tempting to generalize that instruction requiring the investment of more time produces greater learning. However, time spent is not a cause of learning, but is rather a byproduct. Both time taken and skill acquired increased with the addition of more components of programmed instruction.

It is unlikely that changes in student behavior can be fairly represented by complex dimensions such as test or rating scores and time spent. As is true of any research based on test or rating scores, posttest and frame construction task scores in the present study probably did not correlate directly with any specific operant in a student's repertoire. The effectiveness and efficiency of instruction cannot be comprehensively evaluated until rates of operants in practical situations are measured. Future research might assess the functional significance of those behaviors a program is designed to produce. For example, can teachers design frames that actually change behavior? (In other words, can students use a washing machine correctly after completing a program?)

Group experimental comparisons similar to those found in this study are a crude functional analysis. Future application of within-subject repeated-conditions designs should bring differences found in the present study more sharply into focus. Individual-subject designs will permit a precise evaluation of the relation of student characteristics to the magnitude of behavior change produced by interactive instruction. Such an analysis, however, will require experimental conditions with independent segments of subject-matter content to prevent overlap and sequence effects (counterbalanced conditions also appear to be necessary). The group comparisons employed in the present study could be carried out without creating such an unusual program but, unfortunately, brought with them all the problems of combining individual performances into group means (Sidman, 1960). This area of research will remain a challenge for the behavior analyst.

REFERENCES

Abramson, T., & Kagen, E. (1975). Familiarization of content and different response modes in programmed instruction. *Journal of Educational Psychology*, **67**, 83–88.

Anderson, R. C. (1967). Educational psychology. *Journal of Educational Psychology*, **18**, 103–164.

Avner, A., Moore, C., & Smith, S. (1980). Active external control: A basis for superiority of CBI. *Journal of Computer-Based Instruction*, **6**, 115–118.

Borg, W. R., & Gall, M. D. (1983). *Educational research*. New York: Longman Inc.

Bostow, D. E. (1989). *Preparing automated instruction*. Tampa, FL: Custom Systems International, Inc.

Holland, J. G., & Kemp, F. D. (1965). A measure of programming in teaching-machine material. *Journal of Educational Psychology*, **56**, 264–269.

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

Kulhavey, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, **47**, 211–232.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Silverman, R. E. (1978). Programmed instruction. In A. C. Catania & T. A. Brigham (Eds.), *Handbook of behavior analysis: Social and instructional processes* (pp. 454–481). New York: Irvington.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

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

Skinner, B. F. (1984). The shame of American education. *American Psychologist*, **39**, 947–954.

Teacher turned author. (1985). Longwood, FL: Raster Sciences Inc.

Tobias, S. (1969). Effect of creativity, response mode, and subject matter familiarity on achievement from programmed instruction. *Journal of Educational Psychology*, **60**, 453–460.

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

Tobias, S., & Abramson, T. (1971). Interaction among anxiety, stress, response mode, and familiarity of subject matter on achievement from programmed instruction. *Journal of Educational Psychology*, **62**, 357–364.

Vargas, J. S. (1986). Instructional design flaws in computer-assisted instruction. *Phi Delta Kappan*, **64**, 738.

Received June 15, 1989
Initial editorial decision December 14, 1989
Revision received February 15, 1990
Final acceptance January 21, 1991
Action Editor, Terry J. Page