JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2007, **40,** 179–184 NUMBER 1 (SPRING 2007)

# *STRENGTHENING SCIENTIFIC VERBAL BEHAVIOR: AN EXPERIMENTAL COMPARISON OF PROGRESSIVELY PROMPTED AND UNPROMPTED PROGRAMMED INSTRUCTION AND PROSE TUTORIALS*

DARREL R. DAVIS, DARREL E. BOSTOW, AND GUDMUNDUR T. HEIMISSON

UNIVERSITY OF SOUTH FLORIDA

Web-based software was used to deliver and record the effects of programmed instruction that progressively added formal prompts until attempts were successful, programmed instruction with one attempt, and prose tutorials. Error-contingent progressive prompting took significantly longer than programmed instruction and prose. Both forms of programmed instruction substantially increased the appropriate use of behavioral vocabulary during subsequent interpretive essays. These behavioral gains extended to a different setting, suggesting that more was being learned than simply how to answer programmed tutorial frames correctly.

DESCRIPTORS: programmed instruction, Web interaction, constructed response, practical application, behavioral vocabulary, prompting, computer instruction

Programmed instruction is the sequential arrangement of reinforcement contingencies that cumulatively establish terminal repertoires as well as their stimulus control. This term is most frequently associated with teaching verbal behavior (Skinner, 1957) but may broadly refer to all forms of instruction that are response contingent and arranged in a cumulative sequence. Programmed instruction involves carefully crafted response–consequence interactions of a complexity that requires an extensive understanding of operant reinforcement contingencies, which may partially explain the absence of its wide application in schools, universities, and commerce.

The results of any form of instruction may generalize to noninstructional situations and may interlock with practical performances. Kritch and Bostow (1998) explored this kind of generalization with respect to programmed instruction of verbal behavior and discovered that functional relations between repertoires may automatically arise even though they have played no role in verbal instructional contingencies. Research that measures such generalization is warranted, because its results may justify the extensive effort in (a) integrating instructional learning principles, (b) acquiring a thorough understanding of the content, and (c) software programming necessary to deliver programmed instruction.

A recent renewal of interest has grown with computer technology (e.g., Kritch & Bostow, 1998; Miller & Malott, 1997). Kritch and Bostow confirmed that increasing the density of required overt interaction improved the ability of learners to apply rules taught by programmed instruction. The present research revisited the relation between programmed instruction and the generalization of tutorial performance to subsequent practical application, in this case, the nature of interpretive behavioral essays.

## METHOD

### *Participants and Class Context*

One hundred eighteen undergraduate students participated in the fall semester experiment, and 107 students participated in a replication in the following spring semester. These students came from two courses with substantially overlapping

We thank John Ferron for his helpful advice with respect to the statistical analyses in this manuscript.

Reprints may be obtained from Darrel E Bostow, Department of Psychological and Social Foundations, College of Education, University of South Florida, Tampa, Florida 33620 (e-mail: bostow@coedu.usf.edu).

doi: 10.1901/jaba.2007.93-05

behavioral content. They were randomly distributed into the experimental groups in the first experiment and were stratified by class and then randomly distributed into experimental groups in the replication.

Prior to the commencement of the first experiment, students read and took quizzes on Skinner's *Walden Two* (1974), and either Skinner's *Beyond Freedom and Dignity* (1971) or Alberto and Troutman's *Applied Behavior Analysis for Teachers* (2003) depending on the class from which they were drawn. In the experimental procedures, all learners studied tutorials explaining the role of feelings in behavior analysis. These experiments were an additional 1-week assignment in the first semester's experiment and were the initial assignment in the next semester's replication.

### *Setting, Apparatus, Instructional Materials, and Procedure*

Data collection occurred in a laboratory containing 20 computers connected to the Web. The first author supervised the laboratory and used a standard set of vocal instructions to maintain procedural integrity.

The 6,300-word eight-set tutorial about feelings was formatted into programmed instruction (frames) and also into prose in a scrolling Web page format. It covered contrasts between respondent and operant conditioning, setting conditions, public and private behavior, and the difference between feelings and practical behavior. It explained emotions as by-products of contingencies rather than causes, and how feelings can be modified by context management.

Priming and thematic formal prompting (Skinner, 1957) were used during programmed instruction. Generalization was induced with examples. No more than two or three sentences appeared in one frame. Typically, frames contained one blank to which the user responded by typing a word or phrase. The required answers did not appear elsewhere in the frame, and the blanks tended to be near the end of the frame. Many frames contained an abstract statement of a behavioral relation (i.e., a rule) and then an example. Redundancy maintained high correct answer strength as the program introduced new examples. Frame step size was refined through reiterative field testing, such that the probability of answering correctly was maintained at better than .70 per frame and generalization responses were intermittently required. In summary, the tutorial was direct and devoid of content not directly related to emission of correct answers, as suggested by Holland's (1964, 1967) early work. These specifications define *programmed instruction* as used in the present research.

### *Independent Variable (Treatment) Conditions*

*Progressive prompting programmed instruction.* The instructional content was broken into eight sets (lessons), averaging 31 frames per set. In this condition, each frame was a screen presentation, and content initially appeared without prompt letters (i.e., portions of the answer word) surrounding the blanks. If the user's answer was correct, the program advanced to the next frame. Answering incorrectly resulted in seeing ''incorrect'' and ''try again'' below the content, and then one letter of the answer appeared next to the blank in the re-presented frame. If the next attempt was correct, the program advanced the next frame. If not, an additional prompt letter appeared (alternately and cumulatively) before or after the blank. This progressive prompting continued until a final incorrect response resulted in the correct answer being displayed. An unseen software clock timed each student interaction in all conditions. A constantly present ''progress box'' displayed the total number of frames in the set, stated the current frame's serial number, and showed the current frame's ''contribution remaining.'' The latter datum began at 100% with the first attempt for a given frame. An incorrect response reduced the frame's contribution to the terminal percentage correct for that tutorial. The reduction was based on the

percentage of the total letters necessary for that frame's correct answer, and fell to zero if the answer was eventually revealed.

*Traditional programmed instruction with one attempt.* This condition was identical to the progressive prompting condition, with the exception that (a) the learner viewed a frame without letter prompts and (b) the learner was given only one try before the answer was revealed. After the correct answer was supplied by the learner or viewed as a correction, the program advanced to the next frame. Also, this frame's "contribution remaining" box was absent from the tutorial screen because a given frame's contribution was either 100% or 0% depending on the one possible attempt.

*Scrolling prose condition.* The same frame content appeared as paragraphs on one scrolling Web page without blanks. Software timing began with opening the Web page, and clicking on an "I'm done reading" button terminated the timer.

### *Dependent Variables*

*Tutorial percentage correct scores.* In the progressive prompting and traditional conditions, individual frame performance contributed to an accumulating percentage correct score that became the final performance score at tutorial termination. In the progressive prompting condition, individual frame contributions were based on the remaining contribution at the point at which the student achieved frame advancement.

*Posttest.* A 30-item posttest was delivered in the same format as the traditional programmed instruction tutorials. Answers for blanks in the tutorials were also answers in this posttest. However, text surrounding the posttest blanks was terse and devoid of thematic or formal prompting to assess more sensitively the standing strength of intraverbals (Skinner, 1957) and to yield scores distributed at the middle of a 0% to 100% range. Weakened intraverbal supplementation of this nature (Skinner) was assumed likely to have reduced the possibility of a ceiling that could artificially restrict the upper range of possible score variation and with it the possible detection of differences between experimental effects.

The same answer words were occasionally required in different items on the posttest. Analysis of this 30-item posttest, using the results of the 118 participants from the first experiment, yielded a Kuder-Richardson internal consistency score of .84. This score described the degree to which individual items correlated with the remaining items and suggested that separate item performances were under the control of similar variables.

*Essays.* Prior to the commencement of the various treatment conditions and then again immediately following completion of the computer posttest, the lab manager guided students to a Web site on which they wrote brief essays in response to the following statements: (a) "Explain the relationship of your feelings to what you do." (b) "Describe why you feel the way you do." (c) "Say what you would do to create a world in which you will feel better." The output from responses to these essays was used to calculate the proportion of the total number of learner-produced words that matched author-selected words. This measurement yielded scores that were defined by the authors as the degree of behavioral interpretation present within student essays. Vocabulary gain scores were calculated by subtracting preexperimental essay proportions from postexperimental essay proportions.

## RESULTS AND DISCUSSION

Figure 1 presents fall semester tutorial time score distributions, which indicate how long students spent completing tutorial instruction. A statistically significant difference among groups was found, $F(2, 113) = 158.4$, $p < .001$. The traditional PI (one-try) group spent three times as much time as did prose readers (while achieving an average within-tutorial score of 60%). The progressive prompting group

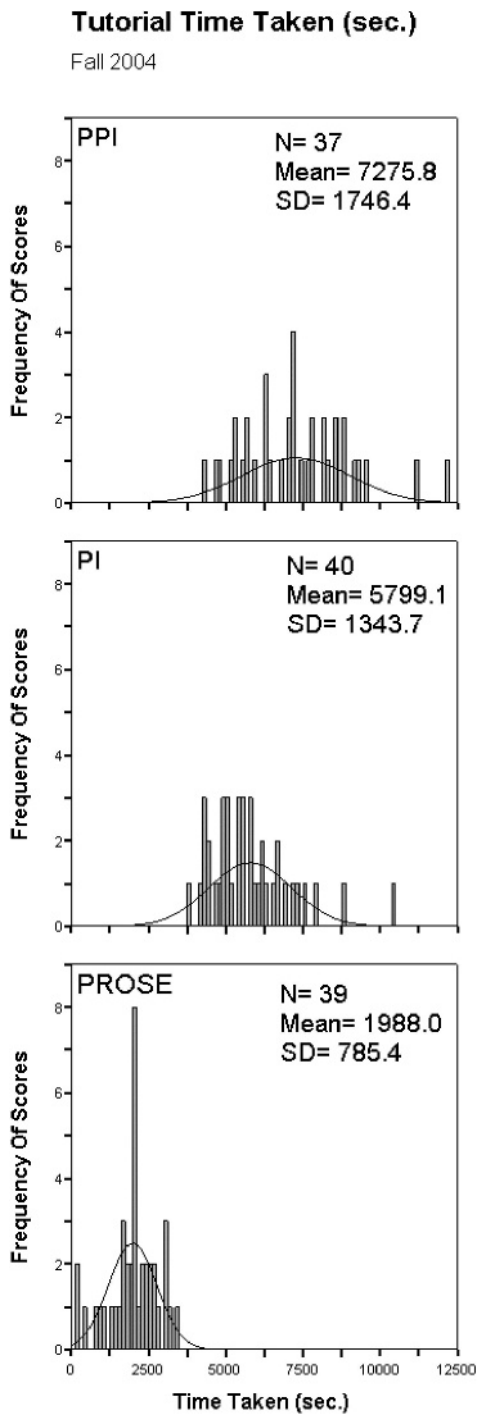


Figure 1. Frequency distributions of the time it took students to complete the tutorial task in the fall semester. Distribution bell curves were calculated using the mean, standard deviation, and range of each distribution.

devoted nearly four times as much time as the prose group did while achieving an 84% within-tutorial average during their tutorial sets.

Fall semester posttest results are presented in Figure 2. Although little difference between the two programmed instruction groups was observed (*M*s = 54.3% and 53.9%, respectively), both groups scored significantly ($p < .01$) higher than prose readers ($M = 32\%$).

Figure 3 presents data on essay scores before and after instruction. Again, a Tukey post hoc analysis indicated that both programmed instruction groups differed significantly ($p < .01$) from the prose group but not from each other (progressive prompting $M = 11.4\%$, $SD = 9.6\%$; traditional $M = 8.5\%$, $SD = 6.3\%$; prose $M = 4.2\%$, $SD = 6.3\%$).

The experiment was replicated in the following spring with the same courses and similar students; however, this time it occurred before any course content had been studied. Group assignment was random as in the fall, but also stratified; students were distributed proportionately by course. The resulting frequency distributions and group differences (not shown) obtained in this replication were similar to the fall term.

Pearson *r* correlations for both fall and spring terms revealed that posttest and essay gain scores were positively correlated with tutorial time taken ($r = .37$, $p < .01$).

As in previous studies, programmed instruction produced greater learning gains in comparison with prose reading but required significantly more time. The higher within-tutorial percentage correct rates produced by the progressive prompting procedure may maintain students' interest and help to maintain their study behavior; however, further investigation into the relation between within-tutorial scores and the motivation to continue is warranted.

In summary, the present research confirmed the relation between the density of overt interactions and both posttest and applied performance that was previously observed by Kritch and

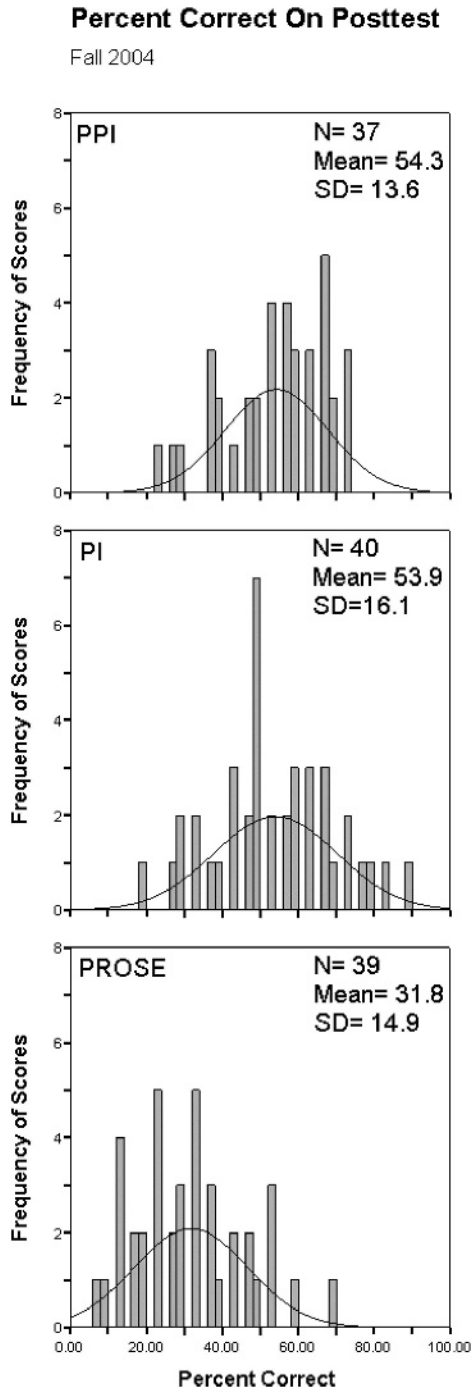


Figure 2. Frequency distributions of the scores earned on the computer posttest that immediately followed tutorial completion.

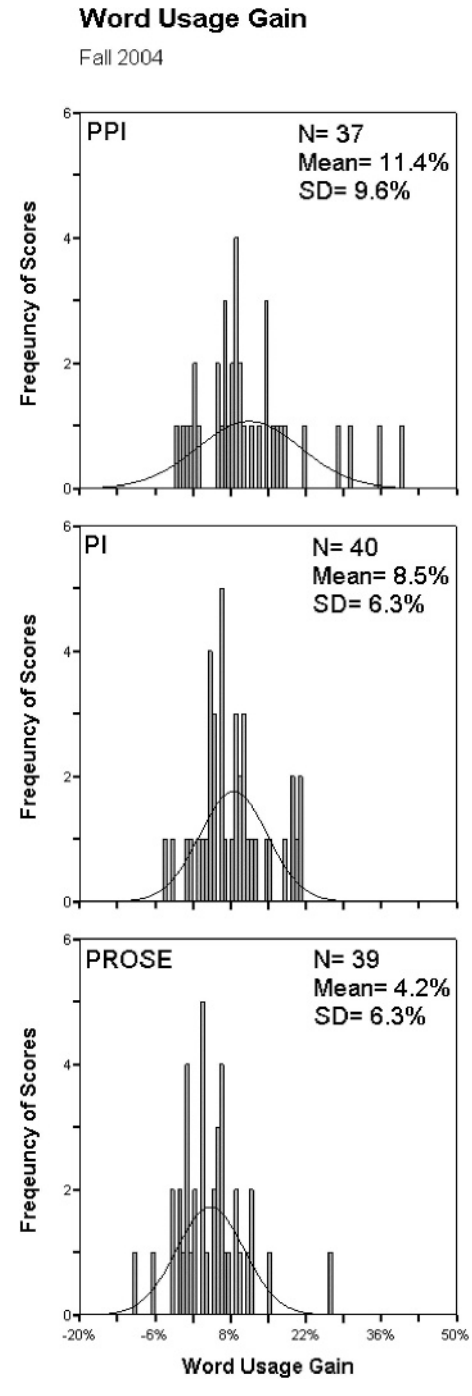


Figure 3. Frequency distributions of the gain scores calculated from subtracting initial essay scores from posttutorial essay scores.

Bostow (1998). In other words, the more frequent the response-contingent program requirement, the better the resulting performance. Furthermore, the current results extended the generality of tutorial-induced changes in behavior to scientific concept usage (essays).

Response-contingent programmed instruction programs that use rigid frame stimuli often result in substantially different specific frame performances, terminal performance scores, and inclination to continue working. They are rarely a perfect fit for specific learners. In the experience of the present investigators, within-tutorial scores correlate with subsequent performance and, incidentally, student complaints. Further investigation of programs that adjust to learners' performances should be a fruitful area of research because adjusting feedback could accommodate students with a broader range of beginning skills, allowing them to achieve higher within-tutorial scores and potentially profit from greater motivation resulting from higher performance scores. The added prompting built into each frame before a correct answer is either produced or shown may relate to terminal performance as well as to student satisfaction; both are promising areas for further research.

## REFERENCES

Alberto, P. A., & Troutman, A. C. (2003). *Applied behavior analysis for teachers*. Englewood Cliffs, NJ: Prentice Hall.

Holland, J. G. (1964). Response contingencies in teaching machine programs. *Journal of Programmed Instruction, 3*, 1–8.

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

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

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

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

Skinner, B. F. (1971). *Beyond freedom and dignity*. Indianapolis: Hackett.

Skinner, B. F. (1974). *Walden two*. New York: Macmillan.

*Received July 15, 2005*
*Final acceptance September 6, 2006*
*Action Editor, Richard G. Smith*