JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1988, 21, 369–380 NUMBER 4 (WINTER 1988)

AN EVALUATION OF LABELING-THEN-DOING WITH MODERATELY HANDICAPPED PERSONS: ACQUISITION AND GENERALIZATION WITH COMPLEX TASKS

DAVID P. WACKER, WENDY K. BERG, COLLEEN MCMAHON, MARK TEMPLEMAN, JEFF MCKINNEY, VALERIE SWARTS, MELISSA VISSER, AND PAM MARQUARDT

THE UNIVERSITY OF IOWA

We conducted two experiments in which moderately mentally retarded persons were trained first to label and then to enter characters into a computer, calculator, or checkbook (label-then-do) within a multiple baseline design. In Experiment 1, 5 young adults were trained to enter statistical programs into computers in an office setting. Following training, all subjects' use of verbal labels and key-entry skills generalized across tasks (programs) and settings (offices and computer terminals). In Experiment 2, 3 junior high school students were trained with self-labeling procedures to complete a key-entry task and to balance a checkbook. The performance of all students generalized across tasks and settings, and the use of labels generalized for 2 of the students. Results are discussed relative to mediated generalization and to establishing verbal control over behavior.

DESCRIPTORS: verbal mediation, vocational tasks, mentally retarded adolescents, generalization

Training moderately mentally retarded individuals to self-instruct appropriate behavior may promote generalized performance across settings or tasks, because the instructions constitute a stimulus that controls subsequent responding (Gifford, Rusch, Martin, & White, 1984; Karlan & Rusch, 1982). With respect to vocational training of mentally retarded persons, self-instruction has been used to improve speed of completing tasks (Crouch, Rusch, & Karlan, 1984), to improve task performance in a cafeteria (Rusch, Morgan, Martin, Riva, & Agran, 1985), and to discriminate letters on a sorting task (Whitman, Spence, & Maxwell, 1987). For moderately and severely mentally retarded persons, Wacker and Berg (1986) recommended a slight variation of self-instructional procedures, which they described as self-labeling. In this procedure, the individual is taught to produce a verbal label (instead of an instruction) that guides performance by making relevant aspects of the task more salient, rather than by explicitly stating what the client should do. For example, Wacker, Carroll, and Moe (1980) taught 3 moderately handicapped elementary students to construct a three-piece object by first stating the color of each piece; Wacker and Greenebaum (1984) taught 7 adolescents with moderate and severe mental retardation to complete a sorting task by first training them to label the relevant dimension they were sorting. In both experiments, generalization of performance (across settings or tasks) occurred.

The present investigation was conducted to extend previous research with self-labels in several ways. First, Experiment 1 was conducted to determine whether the self-labeling procedure could be used on a more complex task (entering statistical programs into a computer) than those used in previous studies. This task is more complex because it requires the client to respond differentially and sequentially to multiple visual stimuli (numbers, characters, and spaces) that are simultaneously available on a data-entry sheet. In previous investigations, clients were required to label only three or fewer stimuli. In addition, generalization across settings, statistical programs, and computer terminals was evaluated to demonstrate further the generalized effects of training with verbal labels.

Experiment 2 was conducted to evaluate more

The authors thank Agnes DeRaad, Lori Popp, George Starr, Lynn Hansen, and the parents and students at Hoover School.

Requests for reprints should be sent to David P. Wacker, Division of Developmental Disabilities, The University of Iowa, Iowa City, Iowa 52242.

systematically the relationship between verbal labels and subsequent performance on complex tasks. The participants' use of labels was evaluated prior to, during, and following training to determine whether changes in correct responding corresponded to changes in their use of verbal labeling on two complex tasks. In previous investigations, labeling and correct responding were not scored separately, thus limiting our understanding of the role of verbal labels in guiding subsequent behavior.

EXPERIMENT 1

Method

Participants and Settings

Five young adults (18 to 25 years old) classified as moderately or severely mentally retarded participated in the experiment. The clients had IQ scores that ranged from 40 to 55 based on assessments with the Stanford-Binet, and all lived in group homes or county care facilities. The clients were independent in all self-help skills, had no significant motor or sensory impairments, and had few behavior problems. All could count to 10, generally knew the letters of the alphabet, and were familiar with a keyboard. None had operated a computer, but all indicated that they were familiar with a typewriter (but had not used a typewriter). All clients spoke in simple sentences but had articulation problems. Language assessments classified the clients as having communication disabilities that were commensurate with their intellectual levels of functioning.

The experiment was conducted in a university hospital for individuals with developmental disabilities. The clients were selected from those receiving a 3-week vocational evaluation in the inpatient unit. As part of their evaluation, the clients received brief training on various tasks, including clerical tasks. For the purposes of this experiment, data-entry tasks were selected and agreed upon by each client's program team in the hospital and by the client's residential staff because key-entry work and typing were considered to be potential work skills for each client.

The training setting was an administrative office located in the hospital. The computer consisted of a Hewlett-Packard terminal and keyboard that were connected to a mainframe located in another building. The computer was located in the back of an office. The generalization setting was the medical records office located on a different floor, and the computer was a Viewpoint terminal and keyboard, also connected to the mainframe. The Viewpoint computer was located at the rear of the office. During all conditions of the experiment, the clients worked in the context of normal office routines and frequently were seated next to clerical staff working at adjacent tables or computers.

Tasks and Behavior

The task required the clients to enter characters from a standard IBM computer sheet into a computer terminal. The characters comprised standard statistical programs from the Statistical Programs for Social Services (SPSS) manual. Two sets of programs were used: a set of three descriptive programs (Condescriptive) and a set of three correlational programs (Pearson Correlation). Each program consisted of 17 to 19 lines of characters, with each line containing up to 36 characters. Included within the characters were letters, spaces (up to seven spaces between characters), numbers, and other symbols (e.g., parentheses, commas, and brackets). The programs contained seven or eight lines of letters and formatting characters, followed by nine or 10 lines of numbers and run completion statements. In addition, the clients were required to push the return key after each line. Each character entered was displayed on the computer screen.

The computer sheets used by the clients were standard IBM program sheets with 32 lines and 80 columns. The programs were written in pencil and were typical of the programs generated by the research office of the hospital. The clients were required to enter between 237 and 244 characters (including pressing the return key) to complete the programs. Exact spacing was required.

Each client was trained to enter one of the two sets of programs (Condescriptive or Pearson Correlation), with the remaining program used as the

generalization task. The set of programs to be trained was counterbalanced across clients.

In addition to entering the program, the clients were instructed to verbalize (label) each character. In most cases, the clients accurately stated the number or letter after two training sessions (incorrect labeling was followed by a verbal demonstration), but they were permitted to use their own labels for characters other than numbers and letters. For example, Max labeled commas as "hooks" and parentheses as "things" (gesturing in the direction they pointed). These labels were acceptable as long as they were used consistently by the client.

Reliability

Each character or return key entered by a client was simultaneously but independently scored by two observers on observation sheets that matched the programs used by the clients. A circle was drawn around each character labeled correctly (posttraining only), and a slash was placed through each character (or return key) entered correctly. The observers sat behind or next to the client. An agreement occurred when the same verbal label and character entered within a space on the IBM sheet were scored by both observers as either correct or incorrect. Interobserver agreement was conducted by the first two authors or one author and a vocational aide.

Reliability checks were conducted across all conditions and participants on 32% of all sessions. Each probe involved a minimum of 196 entries. There were only three disagreements, all of which occurred during baseline.

Design and Procedures

A multiple baseline across subjects design was used to evaluate the results on the training task. A multiple probe design was used to evaluate generalization across tasks and settings.

Baseline. During baseline sessions across both settings and tasks, the client was seated at the computer terminal and was given a practice sheet that contained 10 numbers and letters on two rows, with spaces between some characters. The client was told that he or she was going to type the characters on the computer. The experimenter then demonstrated entering the characters and pressing the return key. In addition, the experimenter verbally stated each character before entering it and made one error to demonstrate how a correction could be made (backspace and retype). The client then practiced entering the same set of characters, receiving correction or praise following each entry. However, the use of verbal labels was not prompted or praised. Following the completion of the practice sheet, no further praise or correction was provided; the client was given the target IBM sheet and was told to type it.

A session consisted of one complete attempt to enter the entire program. However, in many cases, it was unclear what the client was typing or the client began to enter irrelevant information (his or her name over and over again). In these cases, the session was discontinued after 15 min of "wrong" responding. Any character entered in the correct space was scored as correct, even if it was entered as part of an apparently random or wrong string of characters. Finally, spacing errors between rows (return key not pressed or hit more than once) were scored as only one error. The client's self-labels were not scored.

The practice session preceding each baseline session and the liberal scoring during baseline were used to determine whether the client needed extra training to complete the task. If a client was able to complete the task with this minimal amount of training or if the client made only spacing errors, then the use of an extra treatment package such as self-labeling was not indicated. Baseline continued for a minimum of three sessions on the training task, on the generalization task (in the training setting), and in the generalization setting (with the training task).

Training. Training was provided only for the training programs (Condescriptive or Pearson Correlation) in the training setting (administrative office). Two steps were used to train the clients to label-then-do. During Training Step 1, the clients were instructed to label each character, space, or return key. The clients were praised after each correct label, and wrong labels were corrected. If the

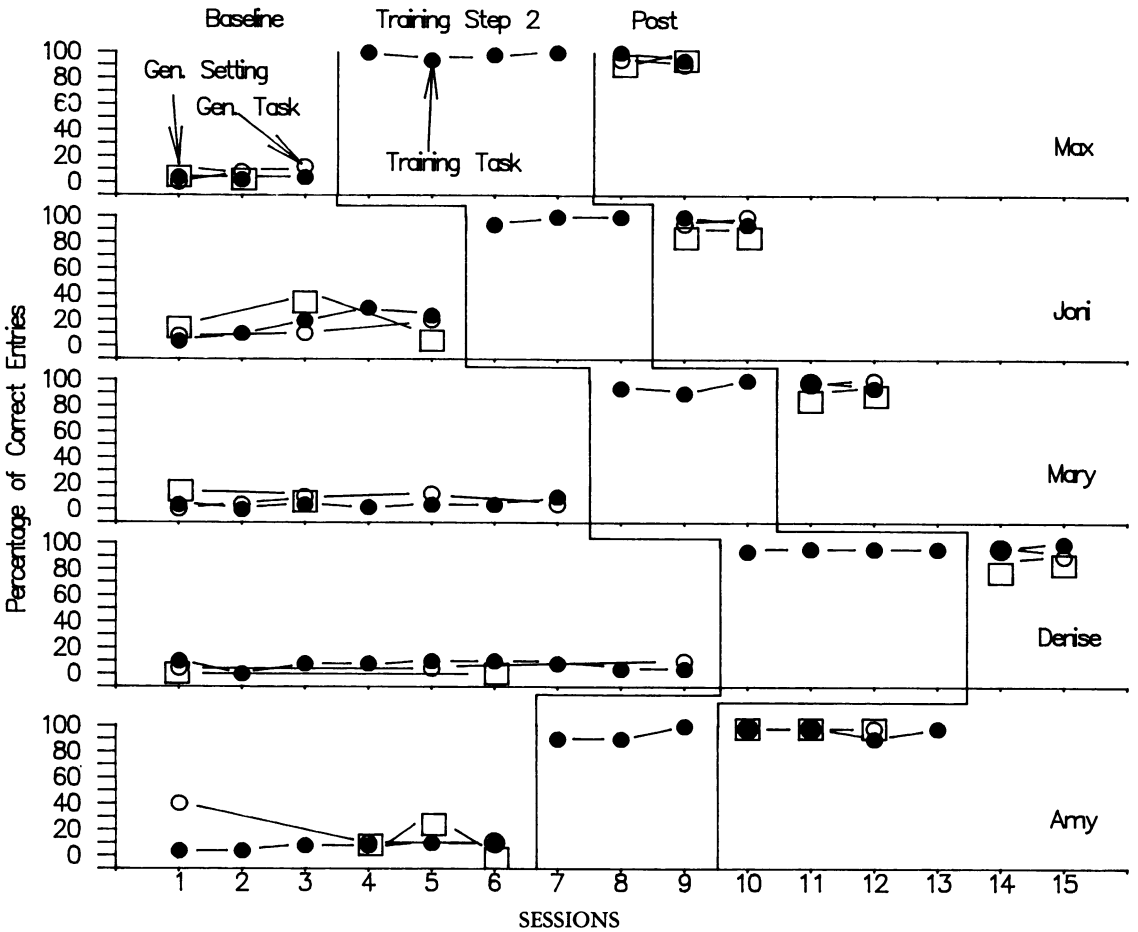

Figure 1. Percentage of entries (characters, spaces, and returns) performed correctly on the training task, generalization task, and in the generalization setting.

client did not know a label, he or she was asked to "guess." Training Step 1 continued until the clients labeled 80% of the characters correctly (or consistently) on two consecutive sessions.

During Training Step 2, the clients were trained to state (label)-then-do (type). At the beginning of each session, a client was told to say the character and then to type it into the computer. Correction was provided for errors, and praise was provided for correct performance. Following a labeling error, the client was asked to repeat the label. If the label was still incorrect, the experimenter said the correct label, and the client imitated the response. For entry errors, the experimenter said "stop," which signaled the client that an error had been made. If the client did not locate the error, the experimenter pointed it out, and the client corrected it. A correct response was scored only for a correct label and a correct entry emitted independently of the experimenter. Praise was delivered only after the label-then-type chain was completed correctly. Thus, the label-then-type sequence was considered one response during training. Training Step 2 continued until the client entered all characters correctly (with corresponding labels) with at least 95% accuracy for each sheet of data for the training program.

Posttraining. Posttraining assessments were conducted for two sessions on the training task, on the generalization task, and on the training task in the generalization setting. No correction or praise was provided. In addition, the clients were not provided with a demonstration or with practice

sheets. Instead, they were given the programs and told to type them.

A correct response during the posttraining assessments required that the entries match exactly with the IBM sheet. Therefore, spacing errors between rows were scored as errors for all characters in each subsequent row (in addition to an error for the return key). A correct label was not required for a correct response to be scored, and labeling was scored independently of key-entry responses.

Results and Discussion

The performance of the clients on the key-entry tasks is shown in Figure 1. During baseline, none of the clients entered more than 40% of the characters (numbers, letters, spaces, and symbols) and returns correctly, with Max, Mary, and Denise consistently entering only about 10% of the characters and returns correctly. However, the clients' performance was scored in a liberal fashion, with spacing and row (return key) errors accounting for only one error.

Training Step 1 was completed by the clients within two to four sessions, during which time they learned to label the characters consistently. Training Step 2 was also completed within four sessions, with all clients completing the training task with at least 95% accuracy.

The clients displayed similar levels of performance on the training task during the posttraining assessments, even though praise and correction were not provided. In addition, the clients performed the untrained key-entry task and the trained task in the novel setting with at least 85% accuracy, even though more stringent scoring criteria were used during generalization than for baseline.

To assess further the clients' continued use of self-labeling, their performance on the key-entry tasks during posttraining assessments (training task, generalization task, and generalization setting) was evaluated both with and without verbal labels. These data are presented in Table 1. As shown in the table, only Mary did not continue to use the label-then-do sequence for every character entered during posttraining. Although Mary was accurate even when she did not label, she continued to use the labels most of the time. These data indicate that the clients' use of both verbal labels and key-entry skills generalized across programs and settings.

Table 1
Continued Use of Verbal Labels and Correctness of Performance During Posttraining

		Verbal label		No verbal label	
Client	Task*	Correct	Incorrect	Correct	Incorrect
Max	TT	472	6	—	—
	GT	444	9	—	—
	GS	424	8	—	—
Joni	TT	439	14	—	—
	GT	465	13	—	—
	GS	440	39	—	—
Mary	TT	388	8	28	0
	GT	430	2	—	—
	GS	387	7	83	2
Denise	TT	437	1	—	—
	GT	422	2	—	—
	GS	417	21	—	—
Amy	TT	864	5	—	—
	GT	670	0	—	—
	GS	682	0	—	—

* TT = training task; GT = generalization task; GS = generalization setting.

The data from Experiment 1 replicate previous findings (Wacker et al., 1980; Wacker & Greenebaum, 1984) by demonstrating that the use of a self-labeling procedure can produce rapid acquisition and generalization across tasks and settings. The results extend previous findings in several ways: The clients learned to complete a complex task that required considerable precision and accuracy, the self-labeling procedure resulted in improved performance with an untrained program and in a novel setting with a different computer terminal, and the procedures were used to train a vocational skill that had been selected as having job potential for the clients. In fact, 2 clients applied this skill to employment as typists after they returned to their residential environments.

The accuracy criterion for the computer tasks was valid based on the required accuracy of performance established by the university key-entry service. Most key-entry operators (who enter data

at much faster rates than these clients) have minimum performance expectations of 80% accuracy. For the clients' skills to be more functional in view of their slower rate, we required 95% accuracy. Finally, coworkers in the training and generalization settings were asked if the self-labeling used by the clients was disruptive. In most cases the coworkers were unaware that the clients were self-labeling, and only 1 client (Max) was considered to be too loud while working in the generalization setting.

EXPERIMENT 2

One of the most striking findings from Experiment 1 was the fact that 4 of 5 clients labeled every character during all posttraining and generalization conditions and that they labeled almost all characters correctly. This is a positive finding, but it does not permit an analysis of whether or not the verbal labels controlled behavior. Instead, we can conclude only that these clients continued to emit both sets of behavior.

A related concern is that we did not score verbal labels independently of correct responses during baseline and training. It is possible that the clients already emitted verbal labels prior to training, but that the verbal labels did not control responding during baseline. Similarly, we cannot conclude that corresponding increases in verbal labeling and correct responding occurred with training, only that both occurred with high accuracy during the posttraining conditions.

Experiment 2 was conducted to evaluate more specifically the role of verbal labels in guiding behavior on equally complex tasks. Younger and less verbal participants were selected to increase the likelihood that they were not familiar with the training tasks, and more important, that they did not use verbal labels to guide their responding. Their use of verbal labels was scored independently of their responding on two distinct tasks during all conditions to determine whether their use of verbal labels increased correspondingly with correct responding, whether correct responding occurred more frequently following a correct label, and whether training resulted in generalized use of the procedure across settings and tasks. Finally, the use of self-labels required the client to emit two responses (state-then-type) that might actually impede rate of responding. Although the training procedures used in Experiment 1 seemed to be efficient given that the effects generalized, it may be that they were actually inefficient relative to the speed with which the clients completed the task. For this reason, in Experiment 2, we evaluated the clients' speed in completing a key-entry task.

Method

Participants

The 3 participants were between the ages of 12 and 13 years and attended a junior high school program for students with moderate mental retardation. The students were classified in school records as moderately mentally retarded based on measures of their intelligence (Stanford-Binet, IQ below 50) and adaptive behavior. All were verbal but had articulation problems making them very difficult to understand. The students used simple sentences, but only Mitch was understandable out of context. The students could count to 10 and recognized most letters of the alphabet; all were classified as having moderate communication disabilities commensurate with their intellectual levels. Carl was also labeled as hyperactive and received Ritalin. None had sensory or motor handicaps. The students had not received previous training on the target tasks, had never typed or used a computer, and had not been taught to self-label.

Settings and Tasks

The students were trained in their school library. Regular school activities were conducted in the library during the experiment, resulting in the students' frequently working near peers. The generalization setting was an office located at the university.

Each student completed a key-entry task and a checkbook task. The key-entry task was very similar to the tasks in Experiment 1 and consisted of the same SPSS Pearson Correlation and Condescriptive programs. The Pearson Correlation program required entry of 202 characters, and the Condescrip-

tive program involved 246 characters. Each student was trained to enter one program, with the other serving as a generalization program. Lisa and Mitch were trained on the Pearson Correlation program, and Carl was trained on the Condescriptive program. An Apple microcomputer was used during training, and either an Apple or an IBM microcomputer was used during the assessment of generalization across settings. A key-entry task was selected for training because the students used microcomputers for other academic tasks (e.g., spelling). Therefore, key entry was considered by the students' teachers and parents to be a useful skill and was included in the students' individual educational programs (IEP).

The checkbook task consisted of the use of a calculator to determine the balance of a checking account after several checks had been written. Real checkbooks from a local bank were used, with each student having his or her own checkbook. The task consisted of subtracting three- to four-digit numbers from four- to five-digit numbers, keeping the decimal point in the correct position. The students subtracted three checks before achieving a balance and then entered the balance into the checkbook. Two different sets of numbers were used by the students, with one set serving as the training task and the second set serving as the generalization task. Mitch and Lisa were trained on Task 1, and Carl was trained on Task 2. The checkbook task was selected based on the IEP goals of the students.

Target Behavior

For both tasks, the target behavior was to (a) state (label) the correct character being entered into the computer, calculator, or checkbook, and (b) enter the character correctly. For the computer task, the characters consisted of letters, numbers, spaces, return key, and miscellaneous characters (i.e., parentheses, semicolons, etc.). For the calculator task, the characters consisted of numbers, minus sign, and equal sign.

For the key-entry task, the percentage of the data sheet (program) completed within 30 min was also recorded. Although the students eventually entered an entire program during each session (or indicated they had completed the program), it was considered reasonable that they complete the entire program within 30 min. The percentage of program completion within 30 min was recorded for all training, posttraining, and generalization sessions.

Reliability

Interobserver agreement was conducted in the same manner as Experiment 1, except that verbal labels were scored independently of responses during all conditions. Reliability checks were conducted across all conditions and participants on 52% of the key-entry sessions and 21% of the checkbook sessions. Reliability ranged from 98% to 100% across all conditions, students, and tasks for both verbal labels and responses.

Design

Data were collected within a multiple baseline design across students for each training task. For Carl and Lisa, a multiple baseline across tasks was also conducted. The specific task trained was counterbalanced across both target tasks, and generalization probes (across tasks for the checkbook task and across tasks and settings for the key-entry task) were conducted for every student.

Checkbook Procedures

Baseline. One demonstration was provided before each session, in which the experimenters correctly labeled all characters, entered them into the calculator, and wrote the balance in the checkbook. One error was always made, and the process was repeated to show the students how to correct errors (start over). Following the demonstration, a student was instructed to determine how much money he or she had in the bank. The student then received the checkbook, with the checks entered in pencil on the balance sheet, and was instructed to use the calculator and to write the balance. Once a balance was entered, the session was completed. No correction or praise was provided, but a student was prompted verbally to write a balance.

Training. Two steps were used to teach the students the training task. During Step 1, the students received praise after the label and after the

response. For Step 2, praise and correction were delivered only after the label-then-do sequence. The criterion for moving from Step 1 to Step 2, and from Step 2 to posttraining, was two sessions with at least 95% accuracy for entering the characters. No criterion was established for labels.

Posttraining assessments and generalization probes. Posttraining assessments and posttraining generalization probes were identical to baseline except that the demonstration was not provided prior to each session. Pretraining generalization probes were identical to baseline.

For Lisa, one change in procedures was implemented at the completion of training after she broke her glasses. After a 1-week delay to obtain new glasses, probes were conducted to determine whether her new glasses affected her performance prior to conducting posttraining assessments.

Key-Entry Procedures

Baseline. Baseline was similar to Experiment 1, with a demonstration trial (with one error) again preceding each session. The students received a program written on an IBM sheet and were told to enter the program. In most cases, a session was completed within 30 min. If a student appeared fatigued or frustrated, two 15-min sessions were used. Only Mitch required 15-min sessions. No prompts, correction, or praise were provided during baseline. For scoring purposes, exact placement of characters was required for a correct response.

Training, posttraining assessments, and generalization probes. Training, posttraining assessments, and generalization sessions were identical to those described for the checkbook task. The criterion for each step was 95% or greater accuracy in entering the characters.

Results and Discussion

The students' performance on the checkbook training task is presented in Figure 2. During baseline, the students completed the task with 10% or less accuracy, and no improvement occurred in performance across sessions. During Training Step 1, improvement occurred immediately for all students with respect to both their labeling and their correct responses. Mitch and Lisa each completed training within 15 sessions. For Mitch, an initial relationship between labeling and correct performance occurred during Training Step 1, but he subsequently began to emit labels independently less frequently. During Step 2, he continued to complete the task errorlessly, but again showed variability in his use of labels. For Lisa, a more consistent relationship occurred, and she needed the least amount of training. Carl took much longer to complete training, requiring 41 sessions. Carl's initial performance during Training Step 1 showed little consistency between labeling and performance. After the seventh training session a more consistent relationship occurred, followed by an inconsistent pattern at the completion of Training Step 1 that continued for Training Step 2.

During the posttraining assessments, all students continued to display nearly errorless performance. Mitch decreased his use of labels, but this did not affect his performance. Both Carl and Lisa continued to emit the labels and performed at a high level of accuracy.

The students' performance on the generalization task is also presented in Figure 2. Their performance prior to training was comparable to baseline on the training task in that no characters were entered correctly in any session. Only one error was made (by Carl) on the generalization task. Mitch did not continue to use verbal labels consistently, whereas both Carl's and Lisa's use of labels generalized across tasks.

The performance of the students on the key-entry training task is displayed in Figure 3. During baseline, students made very few correct responses. Only Lisa made any response other than a random response, and she entered a maximum of 5% of the characters correctly. In addition, during only one session (for Mitch) was a correct label emitted.

Correspondence between the use of labels and correct responding occurred for all students during Training Step 1. The most obvious pattern occurred for Carl, who demonstrated consistent improvement in both labeling and entering characters across sessions. Mitch again demonstrated the weakest pattern and emitted the labels with a declining frequency through Training Step 2. Mitch was observed to self-label following an error, but after

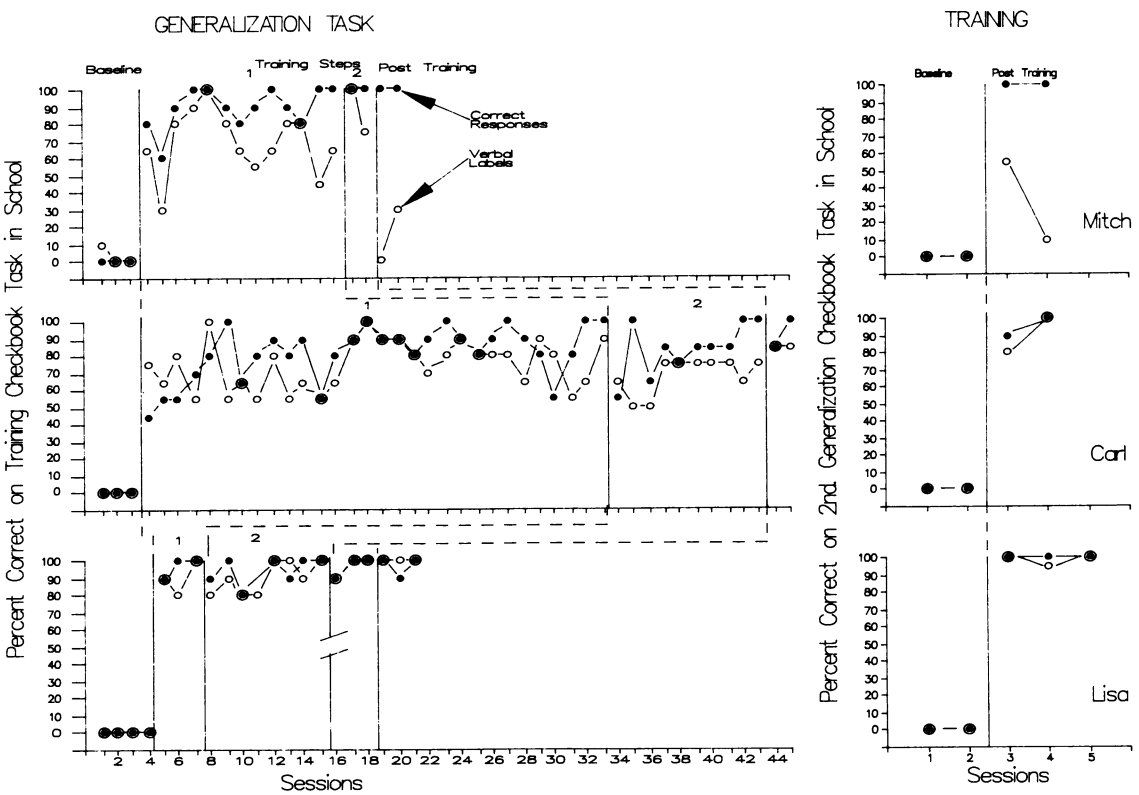

Figure 2. Percentage of correct verbal labels and responses for both checkbook tasks.

labeling several subsequent characters, he would again discontinue his use of the labels. The students also displayed increased speed in entering the data, as shown by the percentage of the data sheet completed within 30 min. By the end of training, which was completed in eight sessions, all students were entering the entire program.

During posttraining assessments, both Mitch and Carl continued to perform at criterion. However, as was the case for the checkbook task, Mitch overtly labeled the characters only intermittently. Of interest is that the accuracy of performance for both Lisa and Carl varied as a function of their use of labels.

Similar results were observed with the generalization tasks and settings, as shown in Figure 3; generalization across both tasks and settings occurred for Lisa and Carl, and for Lisa, errors were not preceded by labels. Mitch continued to perform at criterion, but his speed of performance decreased substantially on the generalization task. In addition, Mitch was less consistent in his use of verbal labels in both generalization conditions, although he used the labels with a much higher frequency than he did prior to training. It is possible that Mitch covertly used the labels, but this possibility cannot be directly evaluated. His use of the labels following errors suggests this possibility, but it is not conclusive.

The results of Experiment 2 replicate Experiment 1 by demonstrating that the self-labeling procedure resulted in relatively quick acquisition of complex tasks and in substantial generalization across both tasks and settings.

GENERAL DISCUSSION

Both experiments demonstrated, as have previous investigations (Wacker et al., 1980; Wacker & Greenebaum, 1984), that the label-then-do pro-

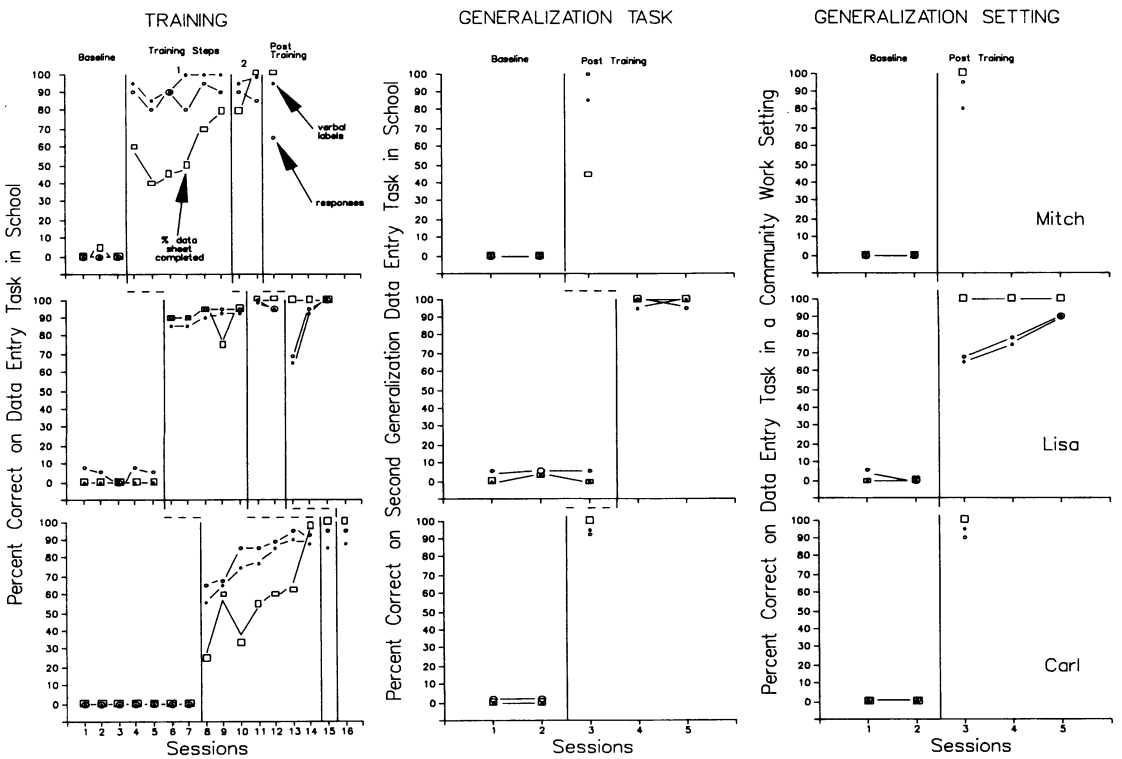

Figure 3. Percentage of correct labels, responses, and data sheets completed on the data-entry tasks during training and generalization.

cedure is effective. In all investigations, the participants rapidly acquired the target behavior, continued to perform at criterion in the absence of experimenter feedback, and generalized their performance across tasks and settings. In the present investigation, participants consistently used self-labels during the generalization conditions in all but one case (Mitch, Experiment 2).

At issue is whether or not the participants' verbalizations (labels) controlled subsequent behavior. Several types of data lend support to this supposition. In Experiment 2, a relationship between the labels and responding occurred for Carl and Lisa, suggesting that the labels controlled responding. This relationship was particularly striking on the key-entry task for Carl on Training Step 1 and for Lisa during posttraining assessments and generalization across settings. When Lisa made errors during these conditions, she also usually failed to emit the label. However, as Stokes, Osnes, and Guevremont (1987) have pointed out, this relationship, no matter how strong, is still correlational and is not conclusive of functional verbal control. During the baseline conditions of both tasks in Experiment 2, the students neither labeled nor entered the characters correctly; only when trained to emit verbal labels did a relationship emerge for Carl and Lisa. Again, however, we cannot conclude that the label-then-do paradigm was necessary. Instead, the most substantial improvement occurred after the students were taught to label (Training Step 1).

In both experiments the participants received no instructions from the experimenters to continue labeling in the posttraining or generalization conditions. Stokes et al. (1987) have suggested that experimenter instructions may control the demonstration of both verbal and nonverbal performance, rather than the verbalization controlling responding. Because the participants (with the exception of Mitch) continued to label consistently in

the absence of experimenter instructions or demonstrations, with corresponding increase in performance during the generalization conditions, it appears that the labels may have controlled responding.

Self-labeling, like self-instruction, can be considered as a form of mediated generalization (Stokes & Baer, 1977). With both procedures, the client emits a verbal response that serves to guide subsequent responding in both training and generalization conditions; or, as Baer (1988) has recently discussed, the verbal response permits a different level of control over responding than do the normally available stimuli. Although infrequently evaluated, both self-instruction (Rusch et al., 1985; Whitman et al., 1987) and self-labels (Wacker & Greenebaum, 1984) have produced generalization in mentally retarded persons. In addition, in two studies that directly compared verbal mediation with an external instructional program (Wacker & Greenebaum, 1984; Whitman et al., 1987), the verbal mediation procedures were more successful in producing generalization.

Self-instruction and self-labels are similar to the extent that both guide behavior through verbal control over responding. However, the two procedures are distinct in that with self-instruction, the client states the behavior to be performed, whereas with self-labels, the client states the antecedent stimulus that in turn guides behavior. By stating the antecedent stimulus, the client is increasing the saliency of that stimulus and limiting the range of antecedent stimuli that might guide responding.

Self-labels might be best conceptualized as self-echoic prompts (Skinner, 1957), verbal statements that are repeatedly stated by the subject and that serve as a verbal stimulus. The participants' repetition of the names of the characters may have guided behavior by continually prompting the participant as to which key was correct, and functioned to make a very complex task that involved multiple antecedent stimuli a rather simple task for the participant to perform correctly; control over behavior changed from a complex array of stimuli to a more simple stimulus–response relationship (Baer, 1988).

Future investigators may wish to evaluate four aspects of the label-then-do procedure. First, as discussed for Experiment 2, it is not clear that the two-step training sequence, used in this investigation and in previous investigations, is needed. Perhaps clients only need to be taught to correctly label the stimuli for improvement in behavior to occur. If labels function as self-echoic prompts, teaching clients to verbally emit the relevant antecedent stimuli may be sufficient.

Second, the efficiency of self-labeling needs further evaluation. It remains to be determined whether the clients' increased speed on the key-entry task was sufficient for most work situations, or whether the clients would have typed at a faster rate if another procedure had been used.

Third, the long-term effects of self-labeling need evaluation. Several studies have demonstrated that self-labels produce generalization across settings and tasks, but it is not known whether self-labeling facilitates long-term maintenance or whether clients continue to overtly emit verbal labels over extended periods.

Finally, generalization across distinct tasks has not been systematically evaluated. In the present experiments and in previous investigations, the generalization tasks have been very similar to the training tasks. Thus, whereas generalization across distinct settings has occurred, it is unclear whether clients who are trained to emit labels for one task will independently emit verbal labels on a second, distinct task.

REFERENCES

Baer, D. (1988, May). Self-instruction by children. In J. Parrish (Chair), *Complex cases of behavior: How best to understand and influence them.* Symposium presented at the annual convention of the Association for Behavior Analysis, Philadelphia.

Crouch, K., Rusch, F., & Karlan, G. (1984). Competitive employment: Utilizing the correspondence training paradigm to enhance productivity. *Education and Training of Mentally Retarded,* **19,** 268–275.

Gifford, J., Rusch, F., Martin, J., & White, D. (1984). Autonomy and adaptability: A proposed technology for maintaining work behavior. In N. Ellis & N. Bray (Eds.), *International review of research on mental retardation* (Vol. 12, pp. 284–314). New York: Academic Press.

Karlan, G., & Rusch, F. (1982). Correspondence between saying and doing: Some thoughts on defining correspon-

dence and future directions for application. *Journal of Applied Behavior Analysis,* **15,** 151–162.

Rusch, F., Morgan, T., Martin, J., Riva, M., & Agran, M. (1985). Competitive employment: Teaching mentally retarded employees self-instructional strategies. *Applied Research in Mental Retardation,* **6,** 384–407.

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

Stokes, T., & Baer, D. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis,* **10,** 349–367.

Stokes, T., Osnes, P., & Guevremont, D. (1987). Saying and doing: A commentary on a contingency-space analysis. *Journal of Applied Behavior Analysis,* **20,** 161–164.

Wacker, D., & Berg, W. (1986). Generalizing and maintaining work behavior. In F. Rusch (Ed.), *Competitive employment issues and strategies* (pp. 129–140). Baltimore: Paul H. Brookes.

Wacker, D., Carroll, J., & Moe, G. (1980). Acquisition, generalization, and maintenance of an assembly task by mentally retarded children. *American Journal of Mental Deficiency,* **85,** 286–290.

Wacker, D., & Greenebaum, F. (1984). Efficacy of a verbal training sequence on the sorting performance of moderately and severely retarded adolescents. *American Journal of Mental Deficiency,* **88,** 653–660.

Whitman, T., Spence, B., & Maxwell, S. (1987). A comparison of external and self-instructional teaching formats with mentally retarded adults in a vocational training setting. *Research in Developmental Disabilities,* **8,** 371–388.

Received December 28, 1987
Initial editorial decision March 24, 1988
Revisions received June 24, 1988; July 25, 1988
Final acceptance August 2, 1988
Action Editor, Nancy A. Neef