TRAINING TWO SEVERELY RETARDED ADOLESCENTS TO ASK QUESTIONS¹

SANDRA TWARDOSZ AND DONALD M. BAER

UNIVERSITY OF KANSAS

Two retarded boys were taught to discriminate items they knew how to label (training items) from items they did not know (probe items), to respond appropriately by naming any training items, and to ask a question about any probe items. The boys did not learn to question when appropriate questioning was modelled by the experimenter; however, when they were prompted and rewarded for asking questions about some training items, they then began asking questions about probe items. Both the modelling- and prompting-reinforcement procedures were introduced in an across-subject, multiple baseline design.

Imitation and reinforcement techniques have been applied to the problem of teaching retarded children to speak. Several studies have shown that when a child is taught to use specific examples of a grammatical rule, he also produces novel examples of this rule that have never been trained, *i.e.*, the rule is generative (Guess, Sailor, Rutherford, and Baer, 1968; Schumaker and Sherman, 1970; Wheeler and Sulzer, 1970; Baer and Guess, 1971). Ideally, a relatively small amount of language training would result in an unlimited amount of language.

The present study attempted to teach the generative use of questions. Asking questions is a highly useful skill, as it can lead to the

acquisition of a large amount of information, and can affect the manner in which the child is treated by adults (Graubard, Rosenberg, and Miller, 1971). Asking a question requires two skills: the ability to discriminate situations where information is required, and the ability to utter the appropriate words.

METHOD

Brief Summary

The subjects were first taught to name several letters, numbers, and colors (training items). Other items of the same kind (probe items) that they were not taught to name were interspersed with the training items. A series of training procedures was then instituted to develop question-asking in response to probe items that were always presented without training to see if question-asking would generalize to them.

To ensure that question-asking was due to the experimental procedures, a multiple baseline design across subjects was used (*cf.* Baer, Wolf, and Risley, 1968, for a general discussion; and Panyan, Boozer, and Morris, 1970, for an across-subjects example). This design requires that each subject experience a different number of sessions in the baseline condition, so that an effect cannot be attributed to the passage of time or the number of experimental trials.

¹This study is based on a thesis submitted by the first author to the Department of Human Development, University of Kansas, in partial fulfillment of the requirements for the degree of Master of Arts, and was partially supported by a Graduate School Honors Fellowship from the University of Kansas. The authors express appreciation to the people at Kansas Neurological Institute, particularly Dr. Doug Guess and Mrs. Minda Garcia, for their help while the study was in progress. Special thanks also go to Drs. James Sherman and Thomas Sajwaj for their critical evaluations and suggestions in preparing the manuscript, and to Drs. Todd Risley and Montrose Wolf for their comments on our reversal procedure. Reprints may be obtained from Sandra Twardosz, Dept. of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, Miss. 39216.

Subjects

Jack and Terry, 16-yr-old residents of the Kansas Neurological Institute, Topeka, participated in the study. Both were severely retarded, and their speech limited to a few short phrases. Ward personnel remarked that they had not heard either boy ask a question. Each could imitate at least a syllable at a time, neither was echolalic or seriously disruptive, and neither could identify the alphabet letters, numbers, or colors to be used in the study. Selection procedures aimed at screening subjects for these characteristics included consultation with ward personnel, observation of the children on their wards, and appropriate pre-tests in an experimental room.

Materials and Experimental Sessions

The items used were the 26 letters of the alphabet, seven colors, and seven numbers. The numbers were one through seven, and the colors were purple, blue, green, yellow, orange, red, and brown. All stimuli were displayed on 3 by 3 in. (7.5 by 7.5 cm) cards.

Each subject was seen separately in 30-min sessions, four or five days per week. Sessions were conducted in a room furnished with a table, two chairs, and a tape recorder. Tokens were given for correct responses, and were traded for candy at the end of the session.

Procedure and Experimental Conditions

Throughout all conditions, five letters and two numbers (Jack) or five letters and two colors (Terry) were the training items, and the remaining 21 letters and five members or five colors were probes. The subjects were taught to name the training items, while the probes were inserted to see if they would ask questions about them. For Jack, the training items were A, K, O, R, U, 2, and 7; for Terry, they were B, O, R, U, X, green, and red.

During each session, the training and probe items were presented to the subject in a sequence in which six training-item trials were followed

by a probe item, followed by another six training-item trials and another probe, *etc.*, until the 26 probes (21 letters and five colors or numbers) were used once each. A cycle thus consisted of 156 training-item trials and 26 probe-item trials. A cycle was usually completed in a session. Each probe item was presented only once per cycle, but the seven training items were presented many times. The sequence of probe items was rearranged by shuffling the stimulus cards seven times. This occurred in approximately the middle of each condition. This general procedure remained in effect for the entire study.

First baseline. Table I summarizes procedures across the experimental conditions. During baseline, the subjects were taught to name the training items, and probes were presented without training (after six training items) to see if they would ask questions about them. On training-item trials, the experimenter prompted naming. That is, he held up an item, named it, and told the subject to identify it ("Say _____"). Praise and a token were given for correct identification, while an incorrect response or silence were ignored. Tokens were initially given for each correct response, but this ratio was decreased over a few sessions to a token for every second correct response, on the average (VR 2). About five prompts were initially given for each training item. When the subject had correctly named an item without help for five consecutive trials, a second training item was introduced and presented until both items were correctly identified for five consecutive randomly sequenced presentations of each item. A third item was then introduced, and the three items were presented until these three were correctly identified for five consecutive randomly sequenced presentations of each item. The fourth, fifth, sixth, and seventh items were introduced in the same manner. Criterion for naming items was reached when the subject could correctly identify the seven training items, randomly presented, with an accuracy of 80% for two consecutive cycles. The number of cycles to be

Table 1
Experimental Procedures and Stimulus Items Across Experimental Conditions

<i>Condition</i>	<i>Training Stimuli</i>	<i>Probe Stimuli</i>
First Baseline	Teach names of letters ("Say ____") and numbers or letters and colors	Present, reinforce question-asking, do not prompt.
First Question-Modelling	Model "What letter?" for letters only, present nos. or colors without "What letter?" Reinforce correct naming.	Present, reinforce question-asking, do not prompt.
Question-Training	Teach "What letter?" response to half of letters only, model "What letter?" for other half letters. Present nos. or colors without modelling "What letter?" Reinforce correct naming.	Present, reinforce question-asking, do not prompt.
Second Question-Modelling	Same as first Question-Modelling (Discontinue requiring subject to ask question)	Present, reinforce question-asking, do not prompt.
Second Baseline	Same as first Baseline (Discontinue modelling "What letter?")	Present, reinforce question-asking, do not prompt.
Control Sessions (Pre-emptive reinforcement)	Same as first and second Baseline	Present Probe, Reinforce before question-asking can occur.
Third Baseline	Same as first and second Baseline	Present, reinforce question-asking

included in each subject's baseline (for the multiple baseline design) was counted from the achievement of this criterion.

Probe items were presented after every six training-item trials, whether or not the criterion for introduction of a new training item had been reached. They were presented for 15 sec or until the subject asked any question about them, and the experimenter said nothing while the probe was presented. The consequences for question-asking were praise, the answer to the question, and a token. All responses to probes other than questions were ignored. In addition, if the subject made an incorrect verbal response to a probe during the last 5 sec of its presentation, the stimulus was not removed until 5 sec after the last incorrect response. (This delay was scheduled to prevent the presentation of the next training item from acting as a reinforcer for incorrect responses to probes.)

Terry began the next condition immediately after he reached the 80%-correct criterion for naming training items. Jack stayed in the base-

line condition for four cycles after he had reached criterion.

First question-modelling. The question-modelling condition was identical to baseline, except that the experimenter asked "What letter?" when presenting each letter training item. As in baseline, nothing was said while presenting the probes, but the subjects were praised and given a token if they asked a question. Terry began the next condition after six cycles of question-modelling; Jack remained in this condition for 10 cycles.

Question-training. The third condition involved requiring the subjects to ask a question to produce a trial and the opportunity to earn reinforcement. They were taught to ask "What letter?" when half of the training items presented in the cycle were held up with the blank side facing them. These items thus resembled the probes, in that the subject could not name them. When asked "What letter?", the experimenter answered the question (e.g., "This is U."), and if the subject named the letter, he

received praise and a token. The subjects were initially taught to ask "What letter?" as a chain. To do this, the experimenter first held up a training item backwards and said, "Say what". If the subject repeated "What", the item was turned around, he named it and received a token. Incorrect responses were ignored. When the subject reliably said "What" for these items without a prompt for five consecutive trials, the experimenter said, "Say what leh-". Praise and tokens were then given only for this response. In a similar fashion, the final syllable was added after performance on "What leh-" became reliable for five consecutive trials.

The remaining half of the letter-training items were presented for naming as during the modelling condition, preceded by "What letter?". The training-color items (for Terry) and the training-number items (for Jack) were always presented without a model question-training procedure. Probes were presented as during baseline and modelling conditions.

Second question-modelling. Question-training was discontinued; the procedures were the same as those of the first question-modelling condition.

Second baseline. Next, question-modelling was discontinued; the procedures were the same as those of the first baseline.

Preemptive reinforcement. Then, praise and tokens were given before the subject could ask a question for probes. The probe was presented, the subject immediately received praise and a token, and the probe was not removed for 15 sec. If the subject asked a question or made an incorrect response to the probe during the last 5 sec of its presentation, the stimulus was not removed until 5 sec after the last response. The procedure for naming training items was the same as during the baseline conditions.

Third baseline. Finally, baseline procedures were reinstated once again.

Recording and reliability. Responses to probes were scored as correct, incorrect, or no response. A correct response was the question "What letter?" for letter, color, or number

probes. Other questions such as "What?", "What's that?", or the like, would also have been acceptable. Because of their articulation difficulties, Jack's "Wha deh-tuh?" and Terry's "Wha leh-tow?" were scored as "What letter?". All other responses to probes were incorrect; silence throughout the 15-sec presentation of the probe was scored as no response.

Training-item trials were scored as correct if the subject correctly named the item without a prompt. Correct responses sometimes did not correspond to the normal pronunciation of the letter, number, or color. Correct responses for some of the items were defined as follows: for Jack, "kuh-A" = K, and "se-hen" = 7; and for Terry, "breh" = red, "car" = R, "deps" = X, and "da-ree" = green. All other responses to these items (other than closer approximations to normal pronunciation) were scored as incorrect. Only the subjects' first response per item was scored. A training-item trial was terminated and scored as incorrect if the subject did not respond within 20 sec.

Responses to the question-training trials (training items presented with the blank side facing the subject) were scored as correct or incorrect. A correct response was to ask the training question, "What letter?" without a prompt. All other responses were incorrect.

The experimenter scored all responses to the training and probe items during each session. At least once during each condition, a session was tape-recorded and scored by the experimenter and an independent observer. There were 10 checks for Jack (cycles 3, 7, 18, 24, 32, 35, 37, 43, 47, 51) and eight for Terry (cycles 6, 16, 21, 24, 26, 32, 36, 40). The cycle to be taped was selected before it occurred, and each taped cycle was included in the reliability estimate. Each judge had a list of the training and probe items presented during the cycle, in order, and compared the subjects' responses to that list. These tape recordings did not include the experimenter's response to the subjects' response (*i.e.*, praise, ignoring, or tokens). Agreement was scored for both training and probe items

only if both judges agreed on the subject's response.

RESULTS

Reliability estimates for responses to training and probe items were calculated by dividing the number of items agreed on by the total number of items scored. Reliability scores were computed for each session, and average scores obtained by totalling the individual scores and dividing by the number of sessions. Reliability scores averaged 96% for Terry and 99% for Jack, with a range of 91% to 100% (Terry) and 95% to 100% (Jack).

To compare the scoring of responses during a session with the scoring of the tapes, the experimenter scored each reliability session once during the session and once from its tape. The experimenter's scoring was identical for 97.5% of the subjects' responses.

Questions

Figure 1 shows the percentage of probes per cycle to which the subjects responded with the question "What letter?". Neither subject asked a question throughout the first baseline. During the first modelling condition, Jack did not ask any questions but Terry asked "What that?" once for one probe letter.

During the question-training condition, both subjects asked the question "What letter?" for 95% of the *question-training* items (training items presented with the blank side facing the subject) by the fourth cycle of training. Both subjects also began asking the same question for letter and number or color probes. Figure 1 shows a steep acquisition curve for question-asking on probes; question-asking occurred for letter probes before it occurred for color or number probes. (It should be emphasized that this rapid acquisition of question-asking for the probes was unprompted, but was reinforced after it occurred.)

Both subjects continued to ask questions for probe items during the second modelling and

second baseline conditions. However, during preemptive reinforcement, question-asking quickly decreased to zero for both subjects. It recovered completely during the third baseline. Terry resumed question-asking on the first probe for which reinforcement was not immediately given, but Jack required one prompt (the experimenter said "What letter?" while presenting the probe) before he resumed question-asking.

At no time did the subjects display inappropriate question-asking by asking questions about the training items that were presented for naming.

Training-Item Trials

Both subjects learned to name the seven training items to criterion. The arrow in Figure 1 indicates the cycle when this occurred, which was cycle 11 for Terry and cycle 14 for Jack. They continued to name above 90% of the items correctly for the rest of the study.

Incorrect Responses: Probes

During those phases of the study when Figure 1 shows little or no correct response, the subjects generally showed some form of incorrect response rather than no response. However, during an intermediate seven sessions of the first baseline, Jack showed mainly a pattern of no response.

DISCUSSION

In this study, two retarded children were trained, through imitation and reinforcement techniques, to ask a question ("What letter?") for a subset of items (training items with blank sides facing the subjects). They then asked the question for untrained probe items (which were form stimuli, not blanks). Furthermore, although question-asking was trained for letter items only, the use of the question "What letter?" extended to number or color probes. The training procedure was demonstrated to be the effective condition for the emergence of question-asking through a multiple-baseline across-

Fig. 1. Percentage of probes per cycle to which the subjects responded with the question "What Letter?" during the seven experimental conditions, baseline, modelling, training, modelling (M), baseline (BL), pre-emptive reinforcement (PR), and baseline (BL).

subjects design. These findings support previous studies (Guess *et al.*, 1968; Schumaker and Sherman, 1970; Wheeler and Sulzer, 1970; Baer and Guess, 1971) in which the training of a grammatical form for specific examples

resulted in its generalization to untrained stimuli.

Although a question may take many forms, only one example of one of these forms was examined in the present study. It is open to

investigation if questions initiated by interrogative words other than "what" (e.g., "how", "when", etc.), and those formed by inverting the subject and verb of a declarative sentence, would become generative using similar techniques. It is also necessary to investigate the establishment of a wider range of discriminative stimuli for questions. The present study used a specific form of visual stimulus to serve as the stimulus to be questioned, although questions that occur in normal conversation often do not refer to such specific stimuli.

The reversal technique used in this study, preemptive reinforcement, was also used by Baer, Peterson, and Sherman (1967) and Burgess, Burgess, and Esveldt (1970). It has been called differential reinforcement of other behavior 0-sec (DRO 0-sec). Preemptive reinforcement was used in this study as a more meaningful control for the role of reinforcement in the acquisition of generalized question-asking than the usual DRO. DRO, when it involves reinforcing several seconds of nonresponse, may be viewed as an active suppressant of the response, rather than as a relaxation of the reinforcement presumably responsible for the prior development of the response. Preemptive reinforcement, by contrast, does not strengthen nonresponding; it presents a reinforcer and then allows the subject a fixed period of time in which to respond or not respond without differential outcomes. (In the usual DRO, to respond is to delay future occasions for reinforcement.)

This study, like previous studies of generative language development in retarded children, demonstrated the emergence of a response class in a highly structured setting. Future studies in this area should focus on generalization of such language forms to the children's natural settings. Furthermore, the children's institutional environment should be examined to determine if the opportunities and necessity for these lan-

guage forms exist. For example, do these children need to ask questions to obtain what they need? Are there sufficient stimuli in the environment that would prompt a question? Would anyone answer their questions? It would be quite purposeless to use professional resources to teach children to speak, if the environmental conditions that maintain speech are not present.

REFERENCES

- Baer, D. M., Peterson, R. F., and Sherman, J. A. The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior*, 1967, 10, 405-416.
- Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 91-97.
- Baer, D. M. and Guess, D. Receptive training of adjectival inflections in mental retardates. *Journal of Applied Behavior Analysis*, 1971, 4, 129-139.
- Burgess, R. L., Burgess, J. M., and Esveldt, K. C. An analysis of generalized imitation. *Journal of Applied Behavior Analysis*, 1970, 3, 39-46.
- Graubard, P. S., Rosenberg, H., and Miller, M. B. Student applications of behavior modification to teachers and environments of ecological approaches to social deviancy 1, 2. In E. A. Ramp and B. L. Hopkins (Eds.), *A new direction for education: behavior analysis*. Lawrence, Kansas: University of Kansas, 1971. Pp. 80-101.
- Guess, D., Sailor, W., Rutherford, G., and Baer, D. M. An experimental analysis of linguistic development: the productive use of the plural morpheme. *Journal of Applied Behavior Analysis*, 1968, 1, 297-306.
- Panyan, M., Boozer, H., and Morris, N. Feedback to attendants as a reinforcer for applying operant techniques. *Journal of Applied Behavior Analysis*, 1970, 3, 1-4.
- Schumaker, J. and Sherman, J. A. Training generative verb usage by imitation and reinforcement procedures. *Journal of Applied Behavior Analysis*, 1970, 3, 273-287.
- Wheeler, A. J. and Sulzer, B. Operant training and generalization of a verbal response form in a speech-deficient child. *Journal of Applied Behavior Analysis*, 1970, 3, 139-147.

Received 29 November 1972.

(Revision requested 1 February 1973.)

(Final acceptance 10 July 1973.)