EFFECTS OF CRITERION-LEVEL PROBING ON DEMONSTRATING NEWLY ACQUIRED DISCRIMINATIVE BEHAVIOR

M. ELIZABETH STELLA AND BARBARA ETZEL

THE UNIVERSITY OF KANSAS

The purpose of the study was to determine whether probes of the final criterion-level discrimination administered during and after training provided an accurate measure of acquisition. Training and probe stimuli were designed to make training and probe trials initially very discriminable and then progressively less discriminable as training progressed. Initially, the discrimination required on probe trials was more difficult than the discrimination required on training trials. However, this difference in difficulty was gradually eliminated as training stimuli were topographically altered and made identical to probe stimuli by the end of training. Results showed that while correct responding was maintained throughout training, error patterns occurred on all probe trials administered during training. Error patterns developed regardless of whether probe trials occurred only at the beginning of training sessions (temporally discriminable probes) or were randomly interspersed in the training sessions (temporally indiscriminable probes). Probe error patterns seemed to be controlled by the stimulus properties of training and probe trials. Thus, probes did not measure acquisition as it occurred during training. Probe error patterns were maintained when probes were administered after completion of training. This final measure of acquisition did not agree with the demonstration of acquisition provided by the final training trial. The results suggest that probe trials can measure a different stimulus-response relationship from that trained when training starts with an easier or known discrimination and probes involve a final or criterion test of a more difficult or unknown discrimination. Stimulus control of correct responses versus error patterns is discussed.

Key words: probes, error patterns, acquisition, stimulus control, generalization, errorless learning, stimulus shaping, verbal counting response, children

A probe is defined as "a change in conditions at some arbitrary point in an experiment made to evaluate or test for the conditions currently in control" (Verhave, 1966, p. 529). When

probe trials that test for acquisition of the final, criterion behavior are administered before and after training, the probes provide a means of measuring change in behavior. However, such designs leave some doubt that acquisition or behavior change was the result of training rather than other simultaneous events or maturation of subjects (Campbell & Stanley, 1963; Hersen & Barlow, 1976; Kazdin, 1978; Kratochwill, 1978). Thus, probing throughout training has been suggested to improve experimental control (Bijou, Peterson, Harris, Allen, & Johnston, 1969) and is frequently carried out (e.g., Doran & Holland, 1979; Fields, 1981). When probes are administered at various points throughout training, they provide a measure of the transition that occurs as behavior changes and/or stimulus control is transferred from one element to another.

In order to provide an accurate measure of transition states, experimenters expect that probes themselves will have no effect upon the behavior. That is, only the independent vari-

This research was partially supported by a grant to the University of Kansas by the National Institute of Child Health and Human Development (#HD002528) through the Bureau of Child Research. Fellowship support to M. Elizabeth Stella was received from a training grant (#HD000247 and #HD07173) to the Department of Human Development of the University of Kansas by the National Institute of Child Health and Human Development. Support was also received from the Kansas Research Institute for the Early Childhood Education of the Handicapped, USOE #300-77-0308.

The authors wish to thank Drs. Donald M. Baer, James A. Sherman, Judith M. LeBlanc, Lois Dixon, Ted R. Ruggles, and Shirley Kramer for their helpful comments on this and an earlier version of this manuscript.

Reprints may be obtained from M. Elizabeth Stella at the Kansas Neurological Institute, 3107 West 21st Street, Topeka, Kansas 66604, or from Barbara Etzel at the Department of Human Development, 130 Hawthorn Hall, University of Kansas, Lawrence, Kansas 66045.

able, not the probe, should be affecting behavior change (Sidman, 1960). However, probes that by their own presence influence behavior either positively or negatively (i.e., are reactive) may confound the effects of the independent variable. Sidman (1960) has recommended that experimenters use probes only in ways that are known to be nonreactive.

Despite this recommendation that properties of probes should be well understood before they are used, few studies have directly examined probe methodology. Little is known about how probes should be designed and implemented in order to provide an accurate measure of acquisition or behavior change without impeding, facilitating, or otherwise altering the effects of the independent variable (e.g., training procedures). Two probe procedures have been identified in the literature and experimentally manipulated to demonstrate how behavior is affected by changes in these procedures. The first variable is how probes are reinforced and the second is when probes are introduced.

Differences in reinforcement schedules between probe and training trials have been shown to affect responses on probe trials. Olenick and Pear (1980) found that a fixed-ratio schedule of reinforcement for correct responses on training trials combined with continuous reinforcement for correct responses on probe trials produced the best performance, and hence the best demonstration of learning, as measured by probe trials. Huguenin and Touchette (1980) reported that differential reinforcement on probe trials resulted in a different measure of stimulus-response relationships than when nondifferential reinforcement was scheduled for probes. They concluded that "the influence of test variables on the behavior of interest may overwhelm the effect of the training variable" (p. 84). The second procedural variable shown to affect responding is when in the training sequence the probes are introduced. Fields (1981) showed that, under certain conditions, early introduction of probes facilitated transfer of control from a familiar stimulus to another more difficult (unfamiliar) stimulus. The reactive nature of probes in these studies was clearly demonstrated to have more positive than negative effects upon responding.

Other researchers have reported more severe or negative problems with probes, such as

error patterns developing on probe trials (Lovaas, Schreibman, Koegel, & Rehm, 1971; Spradlin, Cotter, & Baxley, 1973). These error patterns interfered with the demonstration of the criterion stimulus-response relationship. Instead of measuring the trained response that had been previously demonstrated, probes measured something different and error patterns were substituted for the trained response.

This review of literature indicates that probes can be reactive when they are discriminated from training trials. Probes can facilitate the effects of training when certain reinforcement schedules are used and when introduction of probes is well timed. It is not clear what procedures or conditions cause probes to impede acquisition, nor is it clear when probes measure a different stimulus-response relationship than that trained. In addition, little is known about what conditions lead to the development of error patterns on probes. Error patterns are known to have very powerful effects on acquisition when they are controlled by training stimuli (Sidman & Stoddard, 1966, 1967; Stoddard & Sidman, 1967), but little is known about the effects of error patterns controlled by test or probe stimuli.

The present research was designed to (1) examine procedural variables that make probe and training trials discriminable (e.g., the number of training stimuli that are probed, temporally discriminable probes, temporally indiscriminable probes), (2) examine the extent to which probing throughout training actually measures the acquisition process, and (3) examine how error patterns can affect the ability of probes to measure the acquisition process.

EXPERIMENT I

The purpose of this experiment was to determine whether probes administered throughout training provide a measure of acquisition under conditions where probes may be discriminated from training trials on the basis of temporal and stimulus differences.

METHOD

Subjects and Setting

Seven preschool children from the Edna A. Hill Child Development Laboratory in the Department of Human Development at the University of Kansas were assigned to experi-

mental groups, each containing both males and females. The chronological ages of the subjects were balanced across groups: Group I, four subjects, mean age of 3 yr 11 mo; Group II, three subjects, mean age of 3 yr 10 mo. Otherwise, the assignment of subjects to groups was random.

Sessions were held in a small experimental room (approximately 2.28 m by 2.74 m) connected to an adjoining observation room by a one-way mirror. The experimenter and subject sat side by side on child-sized chairs at a child-sized table. Two cups and plastic tokens were placed on the table between the child and the experimenter. A small tape-recorder recorded the subjects' verbal responses.

Procedures

Pretraining. Pre-entry skills were assessed and established by training if necessary. These sessions were conducted to ensure that all subjects had similar entry-level skills. These consisted of counting one to four stimuli in a set and counting four different sets of stimuli in a left-to-right progression. Subjects were brought individually to the experimental room and invited to sit in one of the chairs at the table. The experimenter then showed the child the two cups placed on the table. One cup contained a predetermined number of plastic tokens and was designated as the child's cup. The experimenter explained that when the child responded correctly, the tokens would be transferred from the experimenter's cup to the child's cup. Next the child was told to choose one of three small toys. To earn the toy (e.g., small plastic animal), the subject had to earn all tokens. Children learned by contacting the contingencies (i.e., no toy earned) that no more than two errors were allowed per session.

Stimuli (Figure 1) consisted of one rabbit, two butterflies, three flowers, and four balls drawn in black ink on white cards measuring 29 cm by 11 cm. The children were trained to count the pictures (presented in random order), beginning with the first (far left) position and proceeding to the right until each of the four sets of stimuli had been counted. For example, the correct verbalization for the trial depicted in Figure 1 would be: *one-two; one; one-two-three; one-two-three-four*. Correct responses were reinforced (praise and tokens delivered) and errors were corrected by the experimenter (see Table 1). Pretraining continued until a criterion of three consecutive trials counted correctly was met.

Pretest. Once criterion was met on pretraining, the pretest was given (Table 1) to obtain a baseline or measure of control by the probe stimuli. The single trial of the criterion pretest (Figure 2) consisted of the simultaneous visual presentation of four standard musical notes: the quarter note or one-count stimulus, the half note or two-count stimulus, the dotted half note or three-count stimulus, and the whole note or four-count stimulus. These stimuli were drawn in black ink on white cards measuring 29 cm by 11 cm. The elliptical bases of the note measured 3.8 cm across, and the line extending up at the right of three of the notes measured 6.5 cm. At the beginning of the pretest session, the subject chose a toy as usual and was told that tokens could be earned for working hard (defined as complying with the experimenter's instruction: "Do what I ask you to do."). The pretest trial was then presented (Figure 2). The experimenter placed the pretest stimulus card in front of the subject and asked the subject to "point to and count" each note. The correct counting re-

PRETRAINING STIMULUS CLUSTERS

Fig. 1. Pretraining stimuli consisted of four sets of familiar pictures. The correct counting response for this trial would be: "1-2; 1; 1-2-3; 1-2-3-4."

CRITERION PRE- AND POSTTEST

Fig. 2. Criterion pretest and posttest trial presented four musical notes: dotted half note; quarter note; half note; whole note. The correct counting response would be: "1-2-3; 1; 1-2; 1-2-3-4."

sponse in this pretest was: *one-two-three; one; one-two; one-two-three-four*. If a child did not count or was silent, the instruction was repeated. Following the child's response, four tokens (one for each stimulus) were transferred from the experimenter's cup to the child's cup, and the experimenter reinforced the child's cooperation in coming to the experimental situation (e.g., "Thank you."). The toy was exchanged for the tokens and the session ended.

Training. Each training session also began with the child choosing a toy. After that, children either received a probe trial followed by training or proceeded directly to training.

(1) *Probes:* For the subjects of Group I, the first trial of every training session was always a probe trial (Table 1). The subjects of Group II received no daily probes. An example of the probe stimuli presented to Group I subjects is presented in Figure 3. A probe trial consisted of all four criterion-level stimuli pre-

Table 1
Summary of Procedure for All Conditions of Experiment 1

Condition	Stimuli	Correct Counting Response	S ²⁺	Correction Procedure?	Performance Criterion Required?
I Pretraining	a) 1 rabbit b) 2 butterflies c) 3 flowers d) 4 balls	"one" "one-two" "one-two-three" "one-two-three-four"	Praise and tokens contingent upon correct counting response	Experimenter modelled correct response and child required to respond correctly	Correct response on three consecutive trials
II Pretest	a) dotted-half note (♩.) b) quarter-note (♩) c) half note (♩) d) whole note (♩)	"one-two-three" "one" "one-two" "one-two-three-four"	Comment (e.g. "Thank you") and tokens contingent upon compliant behavior	None	None
III Daily Probe (1 trial per session for Group I subjects)	Same as for Pretest and Posttest	Same as for Pretest and Posttest	Comment (e.g. "Thank you") no tokens or praise delivered	None	None
IV Program Training (Minimum of 65 trials)	a) one turtle changed into a quarter note b) two turtles in a pond changed into a half note c) three turtles in a pond changed into a dotted half note d) four turtles behind a door changed into a whole note	"one" "one-two" "one-two-three" "one-two-three-four"	Same as for Pretraining	Same as for Pretraining	90% correct on each training session

Table 1—(continued)

Condition	Stimuli	Correct Counting Response	S ^{R+}	Correction Procedure?	Performance Criterion Required?
V (a) First Posttest	Same as for Pretest	Same as for Pretest	Same as for Pretest	None	a) If no errors made, then no more trials presented
(b) Repeat Posttest	Same as for Pretest and first Posttest	Same as for Pretest and first Posttest	Same as for Pretest and first Posttest	None	b) If one error was made on the first Posttest trial, then a second identical trial was presented exactly as the first Posttest trial. The session was then terminated and no further trials presented. If two or more errors were made on the first Posttest trial, the experimenter showed the last five program trials, reminded the subject about the trials where the pictures changed shapes (Program Reminder Procedure) and then presented the second Posttest trial as described above.

sented on a white card of the same size described for pretest stimuli. The stimuli were randomized across the four stimulus positions. Following the choosing of the toy, the card containing probe stimuli was placed in front of the subject and the instruction ("Point and count") was given. Following the subject's response, the card was removed as the experimenter said "Thank you" but delivered no tokens or praise.

(2) *Program*: Group I subjects' probe trials were immediately followed by program training; Group II subjects (who were not probed daily) began each training session with program training trials. Five training trials were given daily. Thus, once all four stimuli were introduced in the training program, the subject responded to four stimuli on each of the five trials, for a total of 20 responses per session.

The training program was designed to minimize training errors and guarantee that all subjects would acquire the criterion-level task in a predictable fashion during training. Stim-

ulus-fading and stimulus-shaping procedures (Schilmoeller, Schilmoeller, Etzel, & LeBlanc, 1979) were used to transfer control of responding from discriminable stimuli to the more difficult unfamiliar criterion-level stimuli, as illustrated in Figure 4. Each program training trial was preceded by a two- or three-sentence story segment (read by the experimenter) that commented on the changes occurring in the

EXPERIMENT I: GROUP I DAILY CRITERION PROBE

Fig. 3. Examples of the probe trial received daily by Group I subjects who were probed daily on all four stimuli in Experiment I. The correct counting response for this trial would be: "1-2; 1-2-3; 1-2-3-4; 1."

visual stimuli (e.g., the disappearance of the turtles in the ponds) but did not instruct or comment on the counting responses that the child was to make. After the story segment was read, the experimenter told the subject to point to and count each stimulus (Table 1). In addition, as each stimulus was introduced (Figure 4: Trials 1, 5, 10, and 16), the experimenter modeled the correct counting response. As a result of the training program, subjects learned to emit a counting response first to easily discriminable pictures that were eventually changed into the more difficult criterion-level stimuli by the last program trial. Thus, the correct verbal counting response on the last program trial (Figure 4: Trial 65) was: *one-two; one-two-three; one; one-two-three-four*.

Praise ("good" or "that's right") and a token

for each correct counting response followed every training trial. Thus, if a child responded to all four stimuli correctly on a training trial, four tokens would be transferred simultaneously to the subject's cup when the response chain was completed. If an incorrect response occurred, the experimenter modeled the correct response and told the subject to respond again. The token(s) for any incorrect response(s) remained in the experimenter's cup and the token(s) for any correct response(s) was transferred to the subject's cup. At the end of the session, the tokens were exchanged for the small toy if 90% of the responses (18 of the 20 training trial responses) were correct.

A criterion of 90% correct was required to progress to a new set of five training trials in the next session. If responding was less than 90% correct on training trials (fewer than 18

Fig. 4. Selected trials from the training program illustrate how stimuli were introduced as meaningful pictures and how the counting cues (dotted lines) were faded (Trials 1 to 21). Stimuli were then shaped to criterion level (Trials 22 to 65).

of 20 responses correct) in one session, those same five trials were repeated in the next session. Sessions were also repeated when a subject returned after an absence of four or more days. Procedures were not altered in any way in these repeated sessions. The standard training program consisted of 65 trials, or 13 sessions of five training trials each. However, two subjects (one from each experimental group) received a longer, preliminary version of the program, one with 80 trials that required 16 sessions to complete. No difference in training or probe performance occurred due to differences in length of the training programs so the results of the two subjects receiving these additional trials were not analyzed separately.

Posttest. The posttest was given in the session following the completion of program training (Table 1). The posttest was identical to the pretest (and probe trials) in stimulus presentation and instructions. After choosing a toy, the stimulus card was placed directly in front of the subject and the subject was told to "Point and count." If the subject responded correctly to all four stimuli on this first posttest trial, four tokens were delivered, the tokens were exchanged for a small toy, and the session was terminated.

If a subject did not respond correctly on the first posttest trial, an effort was made to see if the error(s) could be corrected easily. Two procedures were available, and use of one or the other was determined by the number of errors made on the first posttest trial (Table 1). When only one error occurred (three of the four stimuli were counted correctly), the tokens were delivered as described above, and the stimulus card was removed. The same posttest stimulus card was immediately presented again with no special instructions or feedback. The experimenter merely placed the posttest stimulus card directly in front of the subject as before, and instructed the subject to point and count. Regardless of the response, the stimulus card was then removed. The tokens earned from the first trial were exchanged for a small toy and the session was terminated. A second procedure (Program Reminder Procedure) was used when a child made two or more errors on the original posttest trial. This procedure involved more extensive intervention than the first because two or more errors probably meant that correct responses had not generalized to posttest conditions from training

conditions. Following the first posttest trial (where two or more errors occurred), tokens were delivered as described above. Then the posttest stimulus card was removed and the last five training trials of the program (see Figure 4, Trials 62 and 65 for examples) were shown simultaneously by the experimenter. Pointing to these cards, the experimenter said, "Remember how we read the story about the turtles and all the pictures changed shapes?" The child was not allowed to respond to these training stimuli. Following this reminder, the program stimuli were removed and the posttest trial was presented again with the standard instruction ("Point and count"). Following the subject's response to the second presentation of the posttest trial, the session was terminated as described above.

Design. Subjects randomly assigned to Group I were probed on all four stimuli; subjects in Group II received no probes. A pretest (preprobe) and posttest (postprobe) were administered to all subjects in both groups for within-subject measurement of performance before and after the experimental manipulation, which consisted of one probe trial per session for Group I subjects but none for Group II subjects.

Measurement. The dependent measure was the subject's verbal counting responses. These responses were written down by the experimenter exactly as they occurred. In addition, a tape recording was made of all sessions for scoring by another observer.

Reliability. Reliability was assessed during training and criterion-test sessions. Reliability or interobserver agreement on the content of the verbal responses was assessed for each subject during pretest, one training session (chosen randomly), and posttest sessions. Percentages of agreement were calculated, using the formula

$$\frac{\text{Agreements}}{\text{Agreements} + \text{Disagreements}} \times 100. \text{ Overall agreement on subjects' counting responses was 97\%.}$$

RESULTS

Figure 5 presents correct responding for subjects in Groups I and II and during their pretest, daily probes, and posttest trials (bar graphs), and during training (line graphs), as well as an analysis of their error patterns or response strategies on probes and tests (letter code above bars). During pretests, the criterion-

EXPERIMENT I: INDIVIDUAL TRAINING AND PROBE DATA

Fig. 5. Pretest, training, probe, and posttest data for all subjects in Experiment I. Group I subjects were probed at the beginning of every training session on all criterion-level stimuli. Group II subjects received no daily probe.

level stimuli did not control correct counting responses. All subjects either counted consecutively ("C") across the four stimuli (i.e., "one; two; three; four") or repeatedly ("R") counted each stimulus as "one" (i.e., "one; one; one; one"). Subjects using the repeated count strategy were more likely to have one correct response during pretest; the one-count stimulus would always be counted correctly using this strategy. Such instances of correct responding of course did not reflect acquisition.

Correct responding during program training was similar in Groups I (filled circles) and II (unfilled circles). Program sessions were repeated (as indicated by asterisks) when responding fell below criterion level (90%) or when subjects returned after an absence. The probe experience did not differentially affect responding of Group I subjects in their training trials. All subjects averaged better than 90% correct across all sessions. The criterion-level stimuli appearing above program train-

ing data (line graphs) in Figure 5 indicate the session in which each stimulus completed the shaping process and reached criterion level. Because three of the four stimuli reached criterion level several trials and sessions before training ended, stimulus control of correct responding by criterion-level stimuli was demonstrated repeatedly in the last few training sessions. Stimuli were presented at criterion level for the following number of training trials: two-count (half note) stimulus, 29 trials (Trials 36 to 65); three-count (dotted half note) stimulus, 19 trials; four-count (whole note) stimulus, nine trials. The one-count (quarter note) stimulus reached criterion level on the final training trial and therefore had no overtraining and fewer demonstrations of control at criterion level. However, all subjects responded correctly to all stimuli on the final training trial, so stimulus control of correct responding by all criterion-level stimuli was demonstrated before the posttest was administered.

Group I's error patterns and number correct on daily probe trials are also presented in Figure 5. Subjects persisted in consecutively or repeatedly counting on all probe trials, resulting in an occasional correct response (to the quarter note or one-count stimulus). Subject 2 used an extension of the consecutive count error pattern on daily probes. This pattern consisted of always counting to five ("*one, two, three, four, five*") for the stimulus in the far left position, counting to four ("*one, two, three, four*") for the stimulus in the second position, counting to three for the stimulus in the third position, and counting to two for the stimulus in the last position. This pattern resulted in occasional correct responses, but no consistent correct responding to any stimulus occurred. Thus, no improvement in responding on probe trials occurred across sessions for any subject in Group I. If Group I's probe trials were to demonstrate when acquisition or transfer of correct responding to criterion-level stimuli occurred, then responses to probes (slashed bars under line graphs) should have increased after each stimulus reached criterion level in training.

All subjects received at least one posttest trial, when all four stimuli were presented at criterion level. Subjects probed throughout training on all four stimuli consistently generalized error patterns evident in the probe

trials for the first posttest and did not generalize responses learned during training until trained to do so by the story-reminder procedure (star above second posttest bar). All Group I subjects required this program reminder and then were given a second posttest trial. This procedure resulted in correct responding on the second or repeated posttest trial by all subjects, except Subject 2 who showed the same error pattern on the repeated posttest as in the probes and earlier posttest. Thus, all subjects in Group I could respond correctly to all criterion-level stimuli during training (i.e., correct response on final training trial where all stimuli were at criterion level). However, correct responses did not occur under posttest conditions until after intervention occurred (the program reminder).

Group II subjects, who received no probes during training (Subjects 5 to 7), responded correctly under posttest conditions. Thus, the same learning that had occurred during training was maintained during the posttest trial. Although Group II subjects showed error patterns during their pretests, they did not display those same error patterns during the posttest. Instead, subjects generalized responses exhibited on the final training trial to the posttest. Only one subject (S.5) made an error in the first posttest, and that error was spontaneously corrected (self-corrected).

To summarize, results indicated that probes did not provide an accurate measure of acquisition as it occurred during training, but probes did not interfere with correct responding on training trials. In addition, the error patterns typically seen on the probe trials were maintained under posttest conditions. The occurrence of probe error patterns on the posttest may have been controlled by the physical similarity of probe and posttest stimuli, since all probe stimuli were more similar to the posttest stimuli than all training stimuli.

EXPERIMENT II

Experiment II was designed to determine whether error-pattern development and error-pattern generalization would be reduced or eliminated if a different form of probe was used. It was hypothesized that a reduction in the number of stimuli probed daily might reduce the development of error patterns on probes. A reduction in number of stimuli

probed daily would reduce the similarity of probe and posttest trials, thus decreasing the probability of occurrence of probe error patterns (if they developed) under posttest conditions.

METHOD

Subjects and Setting

Five preschool children from the Edna A. Hill Child Development Laboratory in the Department of Human Development at the University of Kansas were assigned to experimental groups each containing males and females. The chronological age of the subjects was balanced across groups: Group I, two subjects, mean age of 3 yr 10 mo; Group II, three subjects, mean age of 3 yr 10 mo. None participated in Experiment I; otherwise assignment of subjects was random. Sessions were held in the small experimental room described in Experiment I.

Stimulus Materials and Procedures

Pretraining, pretest, program training, and posttest stimuli and procedures were identical to those described in Experiment I. (As in Experiment I, two subjects, S.9 and S.12, received a longer, preliminary version of the training program.)

Daily probes. Examples of probe stimuli are presented in Figure 6. All probe trials consisted of the simultaneous presentation of two criterion-level stimuli presented on white cards of the same size described for pretest and posttest stimuli. One group of subjects received probe stimuli that consisted of the quarter note or one-count stimulus and the dotted half note or three-count stimulus (Figure 6, Group I: top stimulus card). A correct response to this probe would be: *one; one, two, three*. The probe stimuli used with the second group of children consisted of the half note or two-count stimulus and the whole note or four-count stimulus (Figure 4, Group II: lower card). The two stimuli were randomly presented in either the first or second position on the stimulus card. As in Experiment I, one probe trial was presented at the beginning of each training session.

Design. Subjects randomly assigned to Group I were probed with the one- and three-count stimulus and were never probed with the two- and four-count stimulus. Since two stimuli were probed and two were never

EXPERIMENT II DAILY CRITERION PROBE

GROUP I:

GROUP II:

Fig. 6. Example of probe trial received daily in Experiment II. Group I subjects were probed at the beginning of every training session on one- and three-count stimuli (correct counting response would be "1; 1-2-3"). Group II subjects were probed daily on two- and four-count stimuli (correct counting response would be "1-2; 1-2-3-4").

probed, effects of repeated probing could be assessed within each subject by comparing posttest performance on the two stimuli probed daily with that of the two stimuli not probed daily. Subjects randomly assigned to Group II received the opposite experience; probes presented the two- and four-count stimulus but never presented the one- and three-count stimulus. This counterbalancing of which stimuli were probed daily (across Groups I and II) provided a control for any effects due to stimulus differences such as difficulty. In addition, Groups I and II provided information on the effects of probing less than the criterion response. That is, responses to all four stimuli were required at the final criterion level (posttest), but only two of these four stimuli were responded to during probes. Groups I and II were then used for comparison purposes to determine if responding to all four stimuli during probes (as in Experiment I) was more or less disruptive than responding to an abbreviated version of the criterion task (in Experiment II).

Measurement and reliability. Verbal counting responses were recorded as described in Experiment I. Reliability was assessed as in Experiment I and was 99%.

RESULTS

Figure 7 presents complete pretest, training, probe, and posttest data for all subjects in each experimental group. Data are presented for stimuli probed (slashed bars and filled circles) and not probed (dotted bars and unfilled circles).

All subjects showed no correct responding or chance-level responding during pretests. During training, program sessions were repeated (as indicated by asterisks) when total responding fell below criterion level (90%) or when subjects returned after an absence. For example, Subject 8 was probed daily on the

one- and three-count stimulus. During training, stimuli that were being probed (filled circles) and stimuli that were never probed (unfilled circles) usually maintained at a combined average of 90% or higher responding. However, when responding to unprobed stimuli fell, bringing total correct responding below criterion level in Session 13, the same training trials were repeated in Session 14.

Different histories of probing did not affect responding during training. All subjects averaged better than 90% correct on program training trials. This nearly errorless responding was maintained for stimuli in the program that were probed daily as well as stimuli that were never probed. For subjects probed daily on the two- and four-count stimulus (Group II), correct responding was maintained at levels similar to those of Group I. Group II was included to control for stimulus differences

Fig. 7. Correct responding during pretests, training, probes, and posttests for all subjects in Experiment II. Group I subjects were probed at the beginning of every training session on one- and three-count stimuli; Group II subjects were probed daily on two- and four-count stimuli.

such as difficulty, but the similarity in correct responding between Groups I and II indicated no such difference.

All subjects demonstrated the skill of counting all criterion-level stimuli correctly during training, so probes did not interfere with correct responding on training trials. Also, any lack of correct responding on probe trials could not be attributed to lack of acquisition during training. Most subjects persisted in consecutive counting on all probe trials, which resulted in no demonstration of acquisition on probe trials. Use of this error pattern again resulted in an occasional correct response of the one-count stimulus. Subject 9 was the exception and showed more correct responding on probe trials. Subject 9 counted the three-count stimulus correctly on the probe trial in Sessions 2 to 4 and Sessions 6 to 9. However, the three-count stimulus was counted incorrectly beginning with Session 10 when the one-count stimulus was counted correctly (through Session 18 and Posttest). Thus, Subject 9 averaged 47% correct across probe trials, demonstrating some acquisition. However, for most subjects in Groups I and II, the probe trials did not provide a measure of acquisition, since all subjects demonstrated correct responding to criterion-level stimuli on training trials but did not demonstrate this acquisition on probe trials.

If probe trials were performing their function of tracking acquisition of correct responding to criterion-level stimuli, then probe responding should have increased and stabilized at 50% correct in the session after the three-count stimulus reached criterion level in training and increased to 100% in the posttest session. Since subjects continued to use error patterns on all probes, acquisition was not demonstrated on the probes.

For subjects who had a history of being probed on the one- and three-count stimulus (Group I), performance on the first posttest was mixed. Subject 8 did not demonstrate acquisition on either of the stimuli having a probe history (one- and three-count stimulus) or those with no probe history (two- and four-count stimulus). Instead, the error pattern associated with the two-stimulus probe was maintained under the four-stimulus posttest condition. When generalization of correct responses was trained by giving a reminder of

program training and the posttest trial was presented again, Subject 8 still responded incorrectly to the one-count stimulus, resulting in 50% correct on stimuli with a history of probes but responded correctly to stimuli not having a probe history. Subject 9, who had demonstrated some correct responding on probes, made an error on the first posttest trial but self-corrected (SC). When the posttest trial was presented again (with no program reminder), responding was 100% correct for probed and unprobed stimuli.

Subjects probed daily on the two- and four-count stimulus (Group II) were consistently more correct in their responses on the first posttest trial than subjects in Group I. Even though Group II subjects demonstrated no acquisition on probe trials, they did demonstrate acquisition on posttests.

Thus, most subjects probed with only two of the four training stimuli did demonstrate some acquisition on posttests. In three of four cases, errors that occurred on posttests were corrected by the subjects themselves with no assistance from the experimenter. However, neither subject in Group I was 100% correct on the first posttest on stimuli with probe histories. This suggested that Group I had a greater probability of interference in demonstration of acquisition of stimuli probed daily than did Group II subjects. This could have been due to the fact that the three-count stimulus was in the first (far left) position on the posttest trial. Since subjects in Group I were probed daily on this stimulus, the fact that it was responded to first on the posttest may have increased the probability that these subjects would respond with error patterns commonly used during probe trials. In short, the first stimulus (i.e., the three-count stimulus) responded to may have controlled the error-pattern counting chain rather than the correct counting chain. The same logic predicts that subjects in Group II would be more likely to respond to the first posttest stimulus (three-count stimulus) with the correct response acquired during training, since this stimulus never controlled a probe error pattern. Group II's results confirmed this prediction; once the chain of correct responses began and was used on the first two unprobed stimuli, the probability of responding correctly to the two remaining (probed) stimuli increased.

Experiment I vs. Experiment II. As in Experiment I, the results of this experiment demonstrated that criterion-level probes administered at the beginning of each training session did not interfere with correct responding on training trials and did not provide an accurate measure of acquisition as it occurred during training. Subjects did respond correctly to criterion-level stimuli on training trials, but that control or stimulus-response relationship did not occur (transfer or generalize) on probe trials.

The results also indicated that the degree to which correct responding occurred during the first posttest was directly affected by the extent of the subjects' probing history and the similarity of probe stimuli to the posttest stimuli. In Experiment I, those subjects who were probed daily on all four stimuli experienced probes that were more similar to the posttest trial (which also presented four stimuli). In Experiment II, subjects were probed daily on only two of the four stimuli, and so probes were less similar to the four-stimulus posttest trial. Posttest conditions were more likely to maintain probe error patterns for subjects probed on all four stimuli (Group I of Experiment I) than for subjects probed on only two stimuli (Experiment II). Thus, designing probe trials to differ from posttest trials by presenting fewer stimuli and by requiring less than the final criterion response resulted in increased accuracy in performance on the posttest trial.

The fact that error patterns occurred on probe trials and not on training trials in both experiments could have resulted because of differences in the visual stimuli, the occurrence of the story line during training, and/or the timing of the probe (as the first trial of every session).

EXPERIMENT III

The purpose of this experiment was to determine if temporally unpredictable probe trials increased the probability that correct responses would occur under probe and posttest conditions. Other researchers have suggested that the timing of probe trials might affect probe responding (Garcia, 1974; Peterson, 1968; Schroeder & Baer, 1972). Therefore, probes were interspersed randomly among

training trials in order to reduce their temporal predictability.

METHOD

Subjects and Setting

Twelve preschool children (five males and seven females) with a mean age of 3 yr 11 mo served. None had participated in the other experiments. Subjects were randomly assigned to experimental conditions so that each group contained three subjects each, both males and females. In addition, the chronological age was balanced across age groups so that the mean age was nearly equivalent across groups (Group I, 3 yr 8 mo; Group II, 4 yr 0 mo; Group III, 4 yr 0 mo; Group IV, 4 yr 1 mo). Otherwise, assignment of subjects to groups was random. Experimental sessions were held in the small experimental room described in the first experiment.

Stimulus Materials and Procedures

The stimuli used for pretraining, pretest, program, posttest, and probe trials were identical to those described in Experiments I and II. All procedures were identical to those described in Experiments I and II, with the exception of the number of pretests and posttests given and the timing of probe trials.

Pretests. To demonstrate that the subjects came to the experiment with the skill of counting unfamiliar stimuli in a predictable manner, a pretest, as described in Experiments I and II, was given before and after pretraining. Administering the pretest before and after pretraining clarified whether pretraining per se affected the type of error pattern demonstrated on pretests (i.e., did responding change from the first to the second pretest as a function of what occurred during pretraining?).

Posttests. To show whether subjects changed their response patterns on repeated posttest sessions, all subjects (except S.16 and S.22 who were not available for repeated posttests) were given a minimum of four posttest sessions, each consisting of one criterion-level trial. Procedures for the first four posttest sessions were identical to those used in Experiments I and II. If subjects responded correctly (100%) under these minimal-instruction conditions for four consecutive posttest sessions, no further posttest sessions were given. If subjects did not respond correctly (less than 100%) on at least

four of the first five posttest sessions, the generalization of correct responding was trained by a program-reminder procedure given exactly as described in Experiment I. If subjects did not respond correctly on the posttest trial following the program reminder, the last five program trials were retrained (story read, subject responded, differential feedback provided) in the following session. When these last five program trials were retrained (using exactly the procedures described in Experiment I for normal training sessions), the probe trial usually included during normal training sessions was omitted. Each retraining session was followed by another posttest session. This alternation of retraining and posttest sessions continued until correct responding began to occur on posttests. Retraining was then discontinued and posttests were given until four consecutive posttest sessions were 100% correct.

Daily probes. Daily probe trials were administered exactly as described in Experiments I and II (i.e., stimuli presented, subject instructed to "Point and count," subject responded, verbal "Thank you" made). However, probe trials did not always precede program training trials. Instead, the one probe trial presented daily was randomly interspersed among training trials to reduce the predictability of when it would occur. One probe trial was included in every session for subjects scheduled to receive probes. Probe stimuli varied across groups. One group was probed on all four stimuli, one group never received probes, and other groups were probed on only two stimuli (same as in Experiments I and II).

Design

Three subjects were assigned to one of four probe conditions. The four probe conditions were the same as those described in Experiments I and II. Subjects assigned to Groups I and II were probed daily on two of the four stimuli: the quarter note (one-count stimulus) and the dotted half note (three-count stimulus), or the half note (two-count stimulus) and the whole note (four-count stimulus). Thus, subjects in Groups I and II served as their own controls for differences due to effects of probing or not probing stimuli during training. Subjects assigned to Group III were probed on all four stimuli throughout train-

ing, whereas subjects assigned to Group IV were never probed during training (control group).

Measurement and Reliability

Verbal counting responses were measured. Interobserver agreement on the counting responses was 99%.

RESULTS

Figure 8 presents test and training data for individual subjects in two of the four experimental groups (Group I: probed on one- and three-count stimulus; Group II: probed on two- and four-count stimulus). All six subjects responded to pretest stimuli with a consecutive-count ("one, two, three, four") error pattern prior to pretraining (first pretest). After completion of pretraining, all subjects responded to the second pretest using the same error pattern; pretraining did not change their response strategy.

Responding averaged better than 90% correct on program training trials. Acquisition on program training trials was similar across subjects regardless of probe history. Although the high levels of correct responding during program training indicated that all subjects learned to respond correctly to stimuli under training conditions and probes did not interfere with training, trained responses did not generalize to probe trials. Subjects in Groups I and II continued to show error patterns on probe trials after stimuli reached criterion level in training, even though the probe trials were randomly interspersed among the training trials. Some subjects showed one error pattern during early training sessions and another error pattern before training terminated. However, this change did not reflect a change to more appropriate probe responding; the second error pattern was just as inappropriate as the first and resulted in few correct responses on probes. (All correct responses on probe trials could be attributed to the error pattern rather than to any demonstration of differential responding to a certain stimulus.)

Subjects probed daily on the two- and four-count stimulus (Group II) were very successful in immediately generalizing correct responding from training to posttest trials. However, on

EXPERIMENT III: INDIVIDUAL TRAINING AND PROBE DATA

Fig. 8. Correct responding during pretests, training, probes, and posttests for two groups of subjects in Experiment III. Probes were randomly intermixed among training trials for all subjects. Group I was probed on one- and three-count stimuli and Group II was probed on two- and four-count stimuli.

the first posttest, subjects probed daily on the one- and three-count stimulus (Group I) responded incorrectly to stimuli with a probe history, while stimuli with no probe history were counted correctly. Subject 14 responded correctly to probed and unprobed stimuli in all subsequent posttests. Subjects 13 and 15

persisted in errors in subsequent posttests and required special intervention before error patterns were replaced by correct responses.

Figure 9 presents data for the three subjects who were probed daily on all stimuli (Group III) and for the three subjects who received no daily probe (Group IV). All subjects in both

EXPERIMENT III: INDIVIDUAL TRAINING AND PROBE DATA

Fig. 9. Correct responding during pretests, training, probes, and posttests for the two remaining groups of subjects in Experiment III. Probes were randomly intermixed among training trials for subjects in Group III (probed on all criterion-level stimuli). Group IV received no daily probe.

groups used error patterns on pretests. Four subjects (S.19, 21, 22, 23) changed their pretest error pattern following pretraining, but the

second pattern was as inappropriate as the first. Subjects in Group II continued to show error patterns on probe trials, even after stim-

uli reached criterion level and controlled correct responses on training trials. Responses on training trials averaged above 90% correct for all subjects.

During posttests, Group III subjects were more likely than any other group to make errors. Two of three subjects in Group III performed very poorly on the first posttest and on the repeated posttests. Only Subject 21 generalized some correct responses to the first posttest. However, repeated posttest trials with a program reminder and program retraining were required for all three subjects before correct responding occurred consistently under the posttest conditions. Group IV subjects consistently responded correctly on all posttests.

When posttest data were compared across all four groups (Figures 8 and 9), subjects probed on the two- and four-count stimulus (Group II) or never probed (Group IV) generalized and maintained correct responding during posttests. All subjects in Group II (S.16, 17, and 18) and Group IV (S.22, 23, and 24) were 100% correct throughout posttests. Subjects probed daily on the one- and three-count stimulus (Group I) were less successful during posttests and two of the three subjects required intervention (program reminder or retraining) before error patterns were replaced by correct responding on posttests. Since subjects in Groups I and II were probed daily on two of the four stimuli, the differential success on posttests of these two groups cannot be attributed to the number of stimuli probed daily. Since acquisition during training was equally successful, the differential posttest success of these two groups cannot be attributed to specific stimulus differences such as differences in stimulus difficulty. Rather, differences could again be attributed to the fact that the first stimulus responded to during all posttests (three-count stimulus) had a history of being responded to incorrectly by the subjects of Group I but not by the subjects of Group II. This demonstration replicates the findings of Experiment II. When the first stimulus responded to during posttests had controlled errors in earlier probes, these error patterns were maintained on posttest trials (Group I). When the first stimulus responded to during posttests had not been probed and thus did not control errors (Group II), correct responses acquired during training generalized to posttest trials.

This analysis of stimulus control also predicted the errors made by the subjects of Group III (probed on all four stimuli) in posttests. As illustrated in Figure 9, subjects probed daily on all stimuli were more likely to respond with probe error patterns on posttests, since the first stimulus responded to during posttests controlled probe errors. Because all four stimuli controlled errors on probes, the subjects of Group III were more likely to have no correct responses during early posttests than were the subjects of Group I (i.e., the probe trials received by Group III subjects were identical to their posttest trials, since all four stimuli were always presented, whereas Group I subjects had probes that were slightly different from their posttests, in that only two stimuli were presented during probes and four stimuli were presented during posttests). This difference in number of stimuli probed daily may have increased the probability of correct responding during posttests for subjects in Group I.

The posttest results also indicated that when error patterns were maintained under posttest conditions, repeated posttest exposure did not improve performance. Only when some form of direct instructional intervention (program reminder and retraining) was given were error patterns replaced by correct responding.

DISCUSSION

The results of these experiments showed that, under certain conditions, (1) probe trials did not interfere with correct responding on training trials; (2) probe trials did not always provide a measure of acquisition that agreed with measures obtained on training trials (i.e., probes appeared to underestimate acquisition); (3) error patterns that developed on probe trials were maintained under posttest conditions and resulted in a different estimation of acquisition than that obtained on training trials.

Implications for Use of Frequent Probing in Experimental Designs

Although probing frequently throughout training may strengthen an experimental design (Bijou et al., 1969), the results of this experiment show that a measure of acquisition may not always be obtained by probes that occur too frequently. The experimental control that frequent probing is designed to demon-

strate may not be realized when error patterns prevent probes from providing an accurate measure of acquisition.

Frequent probing has generally been used successfully in the study of generalized imitation (e.g., Baer, Peterson, & Sherman, 1967; Garcia, 1974; Guess & Baer, 1973; Peterson, 1968) and therefore is not always a problem. The successful use of frequent probing to measure generalized imitation contrasted to the unsuccessful use of frequent probing to measure acquisition of a new discrimination in the present study may suggest that frequent probing should be avoided as a design feature when certain conditions exist. The present study may suggest some conditions where frequent probing is counterproductive. For example, one condition where frequent probing should be avoided might be when training and probe trials initially require discriminations or responses of unequal levels of difficulty. A second condition might be when certain stimulus-response relationships are not available to the subject at all times due to the nature of training (i.e., certain relationships are not presented until late in training). A third condition could be when very complex discriminations and/or chains of responses are required at the final criterion level of performance. Of course, further research would help define other conditions that make frequent probing undesirable as a feature of an experimental design.

Issues of Stimulus Control

Although the results of these experiments suggest that under certain conditions probes seem to provide an underestimation of acquisition, an alternative interpretation is possible. Probe trials may have been demonstrating another form of stimulus control. In addition to the discrimination that was established directly, a second discrimination may also have been learned. For example, during training, certain stimuli were presented under certain conditions that have been referred to as setting events (Bijou, 1976) or contextual stimuli (Rilling, 1977). Under the training-stimulus and setting-event conditions, a specific stimulus-response pattern was reinforced. During probes, different stimuli were presented under slightly different setting events (e.g., no story line preceded the probe trial, no corrective feedback was given). Since the stimuli and setting events

were not identical to those that occurred during training trials, the probability that the trained response would occur was reduced, and the probability that a different response (error pattern) would occur was increased. Thus, the subjects' responses after one probe trial and one training trial may have been governed by an "if-then" rule of conditionality and coexisted in a conditional relationship. The rule of conditionality could be described as follows: If training stimuli and setting events exist, then respond with patterns reinforced under training conditions; if training stimuli and setting events do not exist, then respond differently (error pattern). In the present study, this conditional relationship would explain why error patterns persisted on probe trials but did not interfere with correct responding on training trials.

How this conditional relationship was established is an issue of stimulus control and how it was maintained is an issue of reinforcement control (Ray & Sidman, 1970). Error patterns on pretests indicated that the criterion-level stimuli and setting events present during the pretest controlled subjects' responses. That is, when instructed to point and count, subjects counted these stimuli as they would any other set of stimuli—by assuming or guessing that each stimulus would receive one count (as stimuli do in the preschool setting). The correct response, of course, was very different from this generalized response; a chain of counting was required for three of the four pretest stimuli. The error patterns seen in the initial pretests and which continued on probe trials were probably generalized from other stimuli and settings to the experimental setting. Thus, the subjects' learning history affected performance on the first exposure (pretest) and subsequent exposures (probes) to criterion-level stimuli, and could explain how stimulus control of error patterns developed so quickly on probe trials. This analysis would suggest yet another condition under which frequent probing may be undesirable (i.e., when stimuli and/or setting events increase the probability that a previously learned but inappropriate response would occur in the experimental setting). Although the development of a stimulus-response relationship (error patterns) on probes may have been influenced by the subjects' past learning history,

how the error pattern was maintained even after subjects were trained to respond differently to criterion-level stimuli is an issue of reinforcement control. Even though reinforcement was not directly programmed on pretest and probe trials, error patterns did not extinguish. It is not possible to determine what maintained the error patterns. However, several setting events could have maintained the error patterns that developed as suggested above. For example, the comment ("Thank you") following each probe, the lack of correction following each probe, and/or the termination of the trial following a response may have reinforced and maintained the error patterns.

The analysis presented above suggests that (1) subjects' past learning history and certain setting events influenced the type of stimulus-response relationships (stimulus control) that developed on probe trials, (2) training and setting events affected the type of stimulus-response relationships that developed on training trials, (3) the interaction of training and probes established conditional control, and (4) certain setting events reinforced and maintained that control. If the development of stimulus-response relationships is as sensitive to setting events as this analysis suggests, then probes can be regarded as a weak or strong experimental tool, depending upon what the experimenter is trying to measure. Probes are potentially an unsatisfactory way to measure behavior change if error patterns or some form of conditional control develop and interfere with the demonstration of behavior that probes are designed to measure. However, the results of the present study suggest that probes may provide a method of examining setting events that influence stimulus control. Rather than avoiding the use of probes because they "underestimate" learning, researchers may be able to use probes to examine conditions under which complex conditional stimulus-response relationships develop.

REFERENCES

- Baer, D. M., Peterson, R. F., & 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.
- Bijou, S. W. *Child development: The basic stage of early childhood*. Englewood Cliffs, N.J.: Prentice-Hall, 1976.
- Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., & Johnston, M. S. Methodology for experimental studies of young children in natural settings. *Psychological Record*, 1969, 19, 177-210.
- Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.
- Doran, J., & Holland, J. G. Control by stimulus features during fading. *Journal of the Experimental Analysis of Behavior*, 1979, 31, 177-187.
- Fields, L. Early and late introduction of probes and stimulus control acquisition in fading. *Journal of the Experimental Analysis of Behavior*, 1981, 36, 363-370.
- Garcia, E. The training and generalization of a conversational speech form in nonverbal retardates. *Journal of Applied Behavior Analysis*, 1974, 7, 137-149.
- Guess, D., & Baer, D. M. An analysis of individual differences in generalization between receptive and productive language in retarded children. *Journal of Applied Behavior Analysis*, 1973, 6, 311-329.
- Hersen, M., & Barlow, D. H. *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press, 1976.
- Huguenin, N. H., & Touchette, P. E. Visual attention in retarded adults: Combining stimuli which control incompatible behavior. *Journal of the Experimental Analysis of Behavior*, 1980, 33, 77-86.
- Kazdin, A. E. Methodology of applied behavior analysis. In A. C. Catania & T. A. Brigham (Eds.), *Handbook of applied behavior analysis: Social and instructional processes*. New York: Irvington, 1978.
- Kratochwill, T. R. Foundations of time-series research. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change*. New York: Academic Press, 1978.
- Lovaas, O. I., Schreibman, L., Koegel, R., & Rehm, R. Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology*, 1971, 77, 211-222.
- Olenick, D. L., & Pear, J. J. Differential reinforcement of correct responses to probes and prompts in picture-name training with severely retarded children. *Journal of Applied Behavior Analysis*, 1980, 13, 77-89.
- Peterson, R. F. Some experiments on the organization of a class of imitative behaviors. *Journal of Applied Behavior Analysis*, 1968, 1, 225-235.
- Ray, B. A., & Sidman, M. Reinforcement schedules and stimulus control. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules*. New York: Appleton-Century-Crofts, 1970.
- Rilling, M. Stimulus control and inhibitory processes. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.
- Schilmoeller, G. L., Schilmoeller, K. J., Etzel, B. C., & LeBlanc, J. M. Conditional discrimination after errorless and trial-and-error training. *Journal of the Experimental Analysis of Behavior*, 1979, 31, 405-420.

- Schroeder, G. L., & Baer, D. M. Effects of concurrent and serial training on generalized vocal imitation in retarded children. *Developmental Psychology*, 1972, 6, 293-301.
- Sidman, M. *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books, 1960.
- Sidman, M., & Stoddard, L. T. Programming perception and learning for retarded children. In N. R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 2). New York: Academic Press, 1966.
- Sidman, M., & Stoddard, L. T. The effectiveness of fading in programming a simultaneous form discrimination for retarded children. *Journal of the Experimental Analysis of Behavior*, 1967, 10, 3-15.
- Spradlin, J. E., Cotter, V. W., & Baxley, N. Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, 1973, 77, 556-566.
- Stoddard, L. T., & Sidman, M. The effects of errors on children's performance on a circle-ellipse discrimination. *Journal of the Experimental Analysis of Behavior*, 1967, 10, 261-270.
- Verhave, T. (Ed.), *The experimental analysis of behavior: Selected readings*. New York: Appleton-Century-Crofts, 1966.

Received September 14, 1981

Final acceptance November 12, 1982