JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1990, 23, 387–396 NUMBER 3 (FALL 1990)

LONG-TERM FOLLOW-UP OF ECHOLALIA AND QUESTION ANSWERING

R. M. FOXX AND GERALD D. FAW

CHOATE MENTAL HEALTH AND DEVELOPMENTAL CENTER

A long-term follow-up of echolalia and correct question answering was conducted for 6 subjects from three previously published studies. The follow-up periods ranged from 26 to 57 months. In a training site follow-up, subjects were exposed to baseline/posttraining conditions in which the original trainer and/or a novel person(s) presented trained and untrained questions. Four subjects displayed echolalia below baseline levels, and another did so in some assessments. Overall, echolalia was lower than in baseline in 80.6% of the follow-ups. Five subjects displayed correct responding above baseline levels. No clear differences were noted in correct responding or echolalia between the trainer and novel-person presentations or between trained and untrained questions. In a follow-up in a natural environment conducted by a novel person, lower than baseline levels of echolalia were displayed by 3 subjects; 2 subjects displayed lower than baseline levels in some assessments. Two subjects consistently displayed correct responding above baseline, and 3 did so occasionally. Issues related to the study of maintenance are discussed.

DESCRIPTORS: echolalia, maintenance, language, mentally retarded, follow-up measures

Several studies using cues-pause-point procedures have shown that mentally handicapped individuals can be taught to use their established repertoires of labeling skills to answer questions rather than to echo (McMorrow & Foxx, 1986; McMorrow, Foxx, Faw, & Bittle, 1987) and that these skills generalize to untrained questions (Foxx, Faw, McMorrow, Kyle, & Bittle, 1988). Although these effects were encouraging, no information was provided regarding their durability (e.g., see Foxx, 1990; Foxx, Bittle, & Faw, 1989; Koegel & Rincover, 1977; Lovaas, 1987).

To address this issue, subjects from these studies were evaluated after several years to determine how their current levels of immediate echolalia and correct responding to trained and untrained questions (i.e., generalization stimuli) compared to their pretraining levels (i.e., baseline conditions). Hence, whenever possible, the original trainer conducted the follow-ups in the same rooms used during baseline. Durability of responding across persons was examined by having a novel individual also present the questions. Finally, prior to the training site follow-up, the effects of presenting the questions in a different setting were examined by having a novel person randomly present them in the subjects' living environments.

Although our major focus was to assess the durability of the cues-pause-point effects, we also sought to add to the knowledge base regarding maintenance. The paucity of reports of long-term behavior change has hindered the development of a technology of maintenance and has raised concerns regarding whether behavior change programs can produce durable results (e.g., Kohler & Greenwood, 1986). Hence, maintenance, one of the most meaningful measures of a program's lasting significance (Baer, Wolf, & Risley, 1968), has remained an underresearched and little understood area.

THE FOLLOW-UPS

The follow-ups for each study are presented separately because different assessments were conducted in each. The follow-up lengths ranged from 26 to 57 months. The same target behaviors (i.e., stimulus/response pairs), recording and reliability methods, and response feedback used in each study were employed.

We thank Linda Bowers for her assistance.

Requests for reprints or complete lists of the questions and correct responses may be obtained from Richard M. Foxx, Choate Mental Health and Developmental Center, 1000 North Main Street, Anna, Illinois 62906.

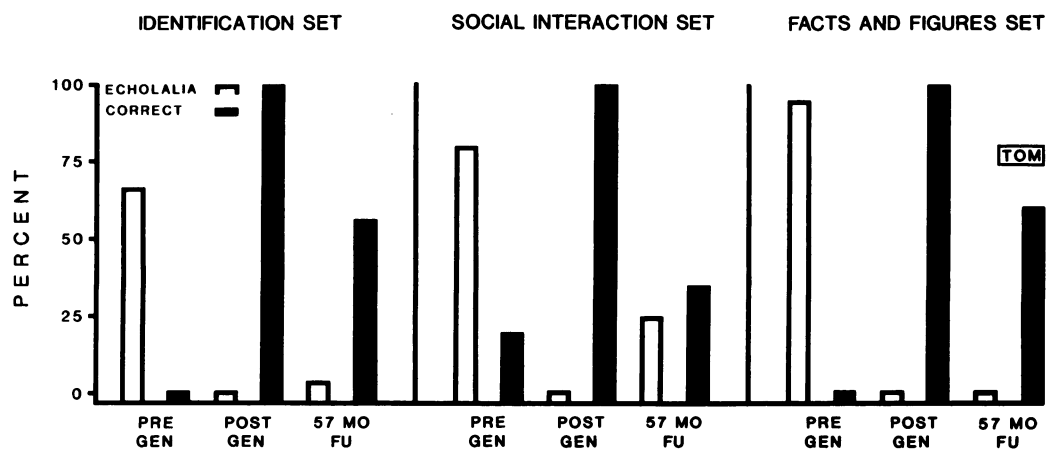

Figure 1. The percentage of echolalic and correct responses to questions in three content sets. PRE GEN refers to prebaseline, POST GEN to posttraining, and 57 MO FU to follow-up performance. The identification set questions were presented by a security officer, peer, and staff member, whereas only the peer and staff member conducted the social interaction and facts and figures sets.

McMorrow and Foxx (1986)

Tom, a 26-year-old man (IQ of 40, Peabody Picture Vocabulary Test) who had been diagnosed as both autistic and mentally retarded, was evaluated. He was taught to use hand-printed word cards (he could verbally label a variety of printed words when prompted to do so) to answer correctly three sets of 10 questions from three content areas (i.e., identification, social interaction, and facts and figures). Echolalia was replaced with correct responding, and the latter remained high as the various procedural components and response cues (i.e., cards) were systematically eliminated. Generalization tests were conducted by a uniformed security officer, a peer, and an unknown staff member prior to baseline and after training.

The 57-month follow-up was designed to repeat the pre- and posttraining generalization assessments. Different individuals conducted the assessments because the original persons had left the area. The security officer asked the set of 10 identification questions in his office. A peer and a novel staff person asked all 30 questions in a facility library. During the previous 57 months, Tom had been discharged and reinstitutionalized several times. He was residing in the facility during the follow-up.

Figure 1 shows that in the pregeneralization trials, Tom displayed high levels of echolalia. He answered only 20% of the social interaction questions correctly. During the postgeneralization trials, he never echoed and answered all questions correctly. In the 57-month follow-up, the mean decrease in echolalia across sets from the pregeneralization trials was 70 percentage points (i.e., 78.6% to 8.6%). Correct responding averaged 51.4% at follow-up versus 5.7% in the pregeneralization trials. Echolalia decreased and correctness increased on each set during follow-up compared to the pregeneralization trials, but echolalia was higher and correctness lower than in the postgeneralization trials.

McMorrow, Foxx, Faw, and Bittle (1987)

This study sought (a) to determine whether lower functioning individuals could be trained to use different types of response cues (e.g., objects rather than word cards), (b) to examine generalization more thoroughly, and (c) to examine responding within and across different training settings. Two mentally handicapped women participated in two experiments conducted in 1985. Dot was 24 years old (IQ of 21, Stanford-Binet) and Eva was 18 years old (IQ of 35, Stanford-Binet). Photos were used in Experiment I as training stimuli, and objects were used in Experiment II.

Experiment I. Training was conducted sequentially on two six-question sets for Eva but on only

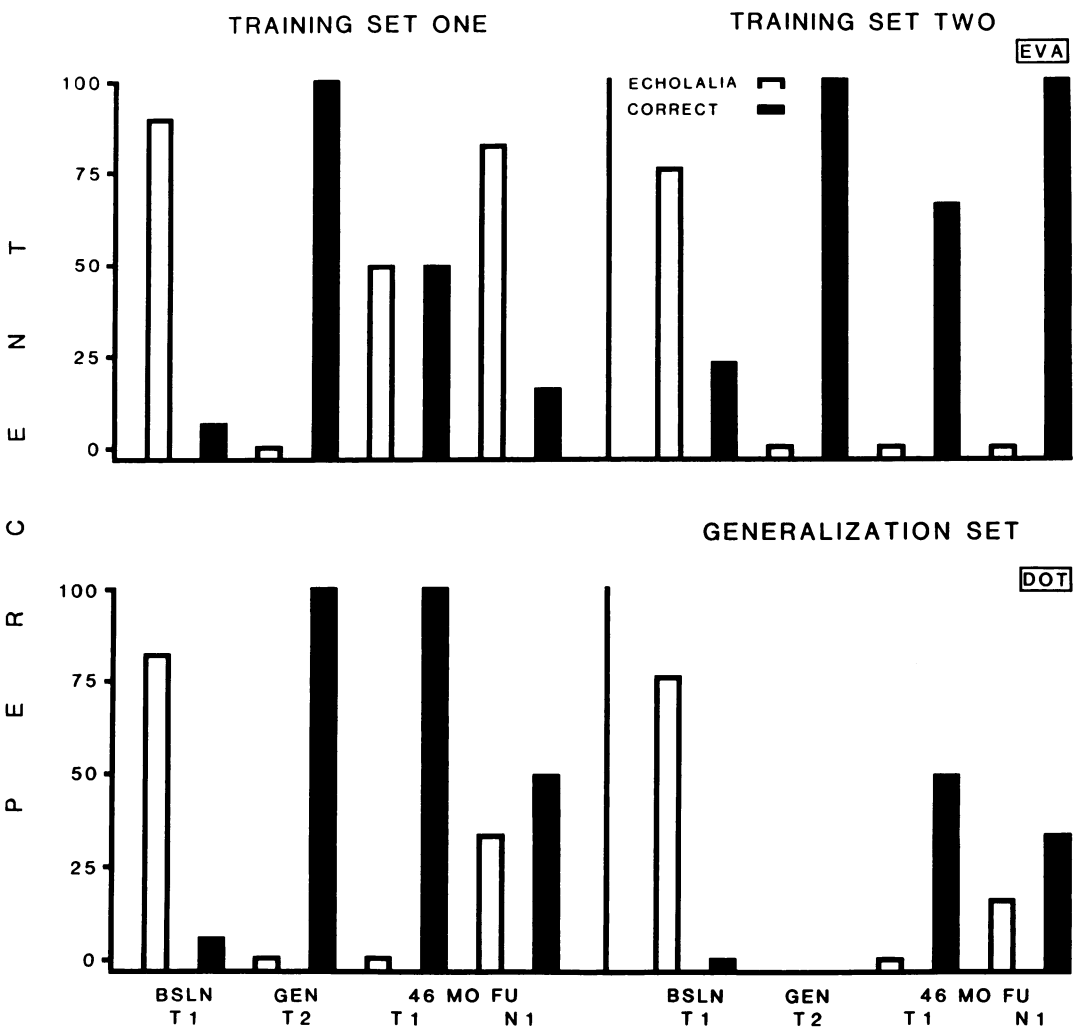

Figure 2. The percentage of echolalic and correct responses to questions in two training sets for Eva and a training and generalization set for Dot. BSLN refers to the baseline conducted by the trainer (T1); GEN refers to a posttraining assessment conducted by a second trainer (T2); and 46 MO FU to follow-up performance assessed by T1 and a novel person (N1). No posttraining assessment (GEN) was conducted in Dot's generalization set.

one set for Dot. Each subject's echolalia was replaced with correct responses that were maintained when the training components were eliminated. Dot's Set 2 served as a generalization set, because her echolalia declined and correct responding improved after training began on Set 1. Following the experiment, generalization trials with no photos present were conducted by a second trainer (T2) on both training sets for Eva and on Set 1 for Dot.

The 46-month follow-up assessments were conducted with each subject in the same fashion as the T2 generalization trials. In one assessment, the original trainer (T1) asked both question sets in the training setting, whereas a novel person (N1) did so in the other assessment because T2 was not available. N1's test was conducted first.

Figure 2 shows that in T1's seven baseline trials, Eva's echoing averaged 90.5% and 76.2%, respectively, on Sets 1 and 2, whereas correct responding averaged 7.1% and 23.8%. In the T2 generalization trials, she did not echo and answered both question sets correctly.

In T1's follow-up, echolalia was 40.5 and 76.2 percentage points lower than baseline in the respective two sets, and correct responding was 42.9 points higher in both. In the N1 follow-up, Eva's Set 1 echolalia and correct responding occurred at near baseline levels. However, she answered all questions correctly and did not echo in Set 2.

In five T1 baseline trials on the training set, Dot echoed 83.3% and averaged 6.7% correct, whereas she never echoed and answered all questions correctly in the T2 generalization assessment and T1 follow-up. In the N1 follow-up, echolalia was 50 percentage points lower than baseline, and correct responding was 43.3 points higher.

Figure 2 shows that Dot's follow-up responding to the untrained stimuli in the generalization set improved from her baseline levels. In the first five T1 baseline trials that corresponded to those in Set 1, echolalia averaged 76.7% and correct responding was 0.[1] Dot's echolalia was 0% and 16.7% with T1 and N1, respectively, and her correct responding was 50% and 33.3%.

Overall, echolalia was either absent or well below baseline levels in all follow-ups except Eva's N1 Set 1. In all follow-ups, correct responding exceeded baseline and was 100% in Eva's N1 Set 2 assessment and Dot's T1 training Set 1 assessment. Echolalia was lower and correct responding was higher with T1 than with N1 in three of the four follow-ups.

Experiment II. Three different settings (i.e., table top, wall, and cluttered room) were used, and there were two sets of five question/response (object) pairs in each setting. Training was sequentially conducted on one set in each setting while the other set remained in baseline. This permitted an examination of generalization within and across settings. The training replaced both subjects' echolalia with correct responses in the three training question sets and also produced increased correct responding and decreased echolalia in the three generalization sets. Responding was essentially unchanged in a series of four generalization posttests (Generalizations A through D) conducted after training. No objects, feedback, or consequences were used in these tests.

The two 43-month follow-ups procedurally replicated generalization Tests B and C conducted, respectively, by T1 and N1, except that another question set of 10 somewhat abstract questions (e.g., "What is white and up in the sky?") was not assessed in follow-up. N1 was unavailable for the follow-up. In Test B, T1 again presented all sets in the same room used in the experiment, whereas a novel person (N2) did so in Test C. Test C was conducted first.

Figure 3 shows Eva's overall mean echolalic and correct responding in the Set 1 baseline trials for the three training settings. Echolalia averaged 92.7% and correct responding 4.8%. In the T1 (Generalization B) and N1 (Generalization C) posttests, echolalia was 0 and correct responding 66.7%. In the follow-up, both types of responding were comparable to baseline. Similar overall effects were seen with the three generalization sets.

Dot's mean overall echolalic and correct responding in baseline trials on the three training sets was 75% and 3.3%, respectively. In both the T1 (Generalization B) and N1 (Generalization C) posttests, echolalia was 6.7%, whereas correct responding averaged 66.7% and 60%, respectively. In the follow-up, echolalia averaged 46.7% with T1 and 40% with N2, whereas correct responding averaged 26.7% with both. In the generalization sets, echolalia averaged 70% and correct responding 10% in baseline and 6.7% and 33.3% in both posttests. Echolalia averaged 33.3% and 40%, respectively, in the T1 and N2 follow-ups, whereas correct responding was 13.3% in both.

These follow-up results were less encouraging than those from Experiment I. Eva echoed about the same as baseline, whereas Dot displayed lower than baseline levels. Only Dot's correct responding on both sets exceeded baseline levels, but it was

[1] Only the first five generalization set trials were used in computing the baseline mean because the training set had only five baseline trials. This method of calculating the baseline mean for a generalization set was used in those follow-ups that contained such a set. When more than one generalization set was involved (see McMorrow et al., 1987, Experiment II), the number of trials used to calculate the generalization sets' baseline mean was based on the number of baseline trials in the initial training set.

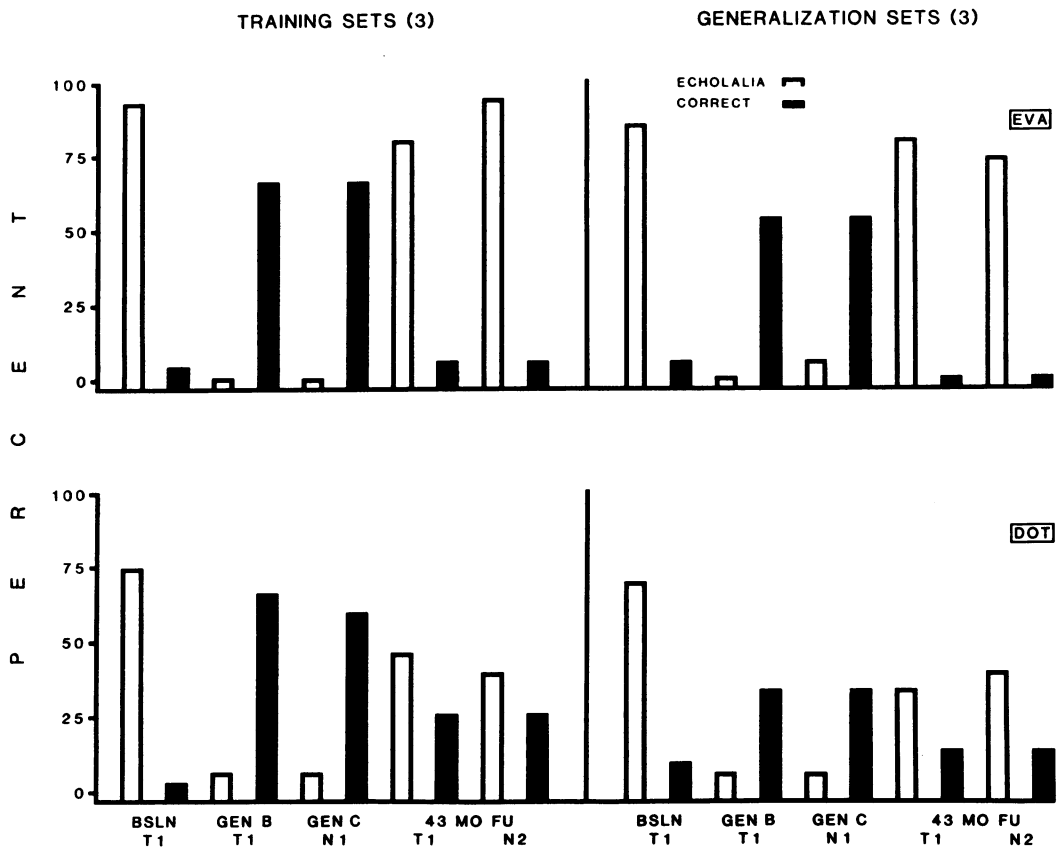

Figure 3. The percentage of echolalic and correct responses on three trained and three untrained (generalization) question sets. BSLN refers to the baseline conducted by the original trainer (T1); GEN B and C were posttraining generalization assessments conducted, respectively, by T1 and a novel person (N1). The 43-month follow-up (MO FU) was conducted by T1 and a different novel person (N2).

unimpressive. No clear differences were apparent for either subject regarding their echolalia on training versus generalization sets or to different questioners (T1 vs. N2). Although correct responding was higher on the trained sets, it differed little from the untrained ones. Correct responding was the same regardless of who questioned them.

Foxx, Faw, McMorrow, Kyle, and Bittle (1988)

Because McMorrow et al. (1987) did not demonstrate functional control of generalization, Foxx et al. (1988) attempted to do so. We also sought to measure the strength and spread of any generalization effects.

Three subjects participated: Casey, a 38-year-old male with severe mental retardation (estimated IQ of 27, Vineland Social Maturity Scale); Steve, a 16-year-old male with severe mental retardation (IQ of 35, Stanford-Binet); and John, a 43-year-old male with profound mental retardation (estimated IQ of 20, Vineland Social Maturity Scale). All displayed immediate echolalia, and Steve sometimes displayed perseverative speech by saying "no" repeatedly. Training was sequentially introduced across subjects on one set of five question/response (object) pairs in one setting, and generalization was assessed on a different set in another setting. (Although periodic probes also were employed, they are not germane to this follow-up because no posttraining measures were taken.) Echolalia was replaced with correct responding in the training settings, and generalized improvements occurred in

the generalization set. Following the study, four posttests (A through D) were conducted in a novel room with no objects present.

The 26-month follow-up procedurally replicated Posttests A and B and, as such, permitted a direct comparison to the initial baseline that was conducted in an empty room. (This empty room baseline condition was labeled "baseline barren" in Figure 1 in the original study.) This was accomplished by having the original trainer (Posttest A) and a novel person (Posttest B) ask the training and generalization set questions in the same room used in the posttests. The original trainer (T1) for Casey and John was available, whereas Steve's trainer had left the area, as had the novel person who conducted everyone's Posttest B. As a result, T1 and N2 conducted Casey's and John's follow-ups, and two novel individuals (N2 and N3) conducted Steve's follow-ups. Tests involving novel persons were conducted first.

Figure 4 shows that no one correctly answered any training set questions during baseline. Echolalia averaged 53.3% and 84.4%, respectively, for Casey and John, and Steve's maladaptive speech averaged 80%. In Posttests A and B, only Steve displayed maladaptive speech (20% in both), and the subjects' correct responding ranged from 60% to 80%.

In the 26-month follow-up, Casey's echolalia was 0 and John's was 60% and 20% with T1 and N2, respectively, whereas Steve's maladaptive speech was 80% with both N2 and N3. In both assessments, Casey and Steve scored 20% correct and John scored 0.

No one responded correctly to any generalization set questions during baseline. Casey and John's echolalia averaged 40% and 86.7%, respectively, whereas Steve's maladaptive speech averaged 88%. In Posttests A and B, maladaptive speech and echolalia were either low or nonexistent, and correct responding ranged from 20% to 80%.

In the follow-up, Casey's echolalia was below baseline with T1 and at baseline with N2, whereas John's was much lower with both. Steve's maladaptive speech exceeded baseline with both assessors. All had 0 correct in every assessment, the same as baseline.

Casey's and John's overall follow-up results for echolalia were positive, whereas Steve's results for maladaptive speech were not. There were no clear differences in responding between questioners or sets for any subject. The data for correct responding were quite unimpressive.

Natural Environment Follow-Up

A novel person presented the subjects with the follow-up questions in their living environments prior to the training site assessments. This assessment resembled the subjects' first baseline trial, because it was their initial exposure to the questions in this setting and the questioner was a novel person. As a result, it was compared to the baseline trial.

Table 1 shows that Tom, Casey, John, Eva, and Dot (Experiment I) had lower levels of echolalia at follow-up than at baseline, whereas Steve's maladaptive speech and Eva's and Dot's echolalia in Experiment II were at or above their initial baseline trials. Overall, echolalia was lower than baseline in 11 of the 17 comparisons (64.7%) with a mean decrease of 50 percentage points (range, 16.6% to 100%). Echolalia averaged 43.7% (range, 0% to 93.3%) versus 73.7% in the baseline trial (range, 40% to 100%).

Table 1 also shows that correct responding was higher than baseline in all sets for Tom, Eva (Experiment I), and Dot and higher on the training sets for Eva (Experiment II), Casey, and Steve; John showed no change. Correct responding was higher than baseline in 12 of 17 comparisons (70.6%) with a mean increase of 23.6 percentage points (range, 6.6% to 70%). The overall means were baseline, 6.3% (range, 0% to 33%) and follow-up, 22.5% (range, 0% to 70%).

These results seemed encouraging given that (a) between 26 to 57 months had elapsed since the subjects had been trained, (b) the questions were presented away from the training site by an unfamiliar person, and (c) no response cues were present. In the majority of assessments, echolalia was well below the subjects' initial baseline trial even though they were not necessarily answering correctly. Indeed, although correctness was higher than in baseline, it averaged only 22.5%. The subjects' use of incorrect responses in a situation in which

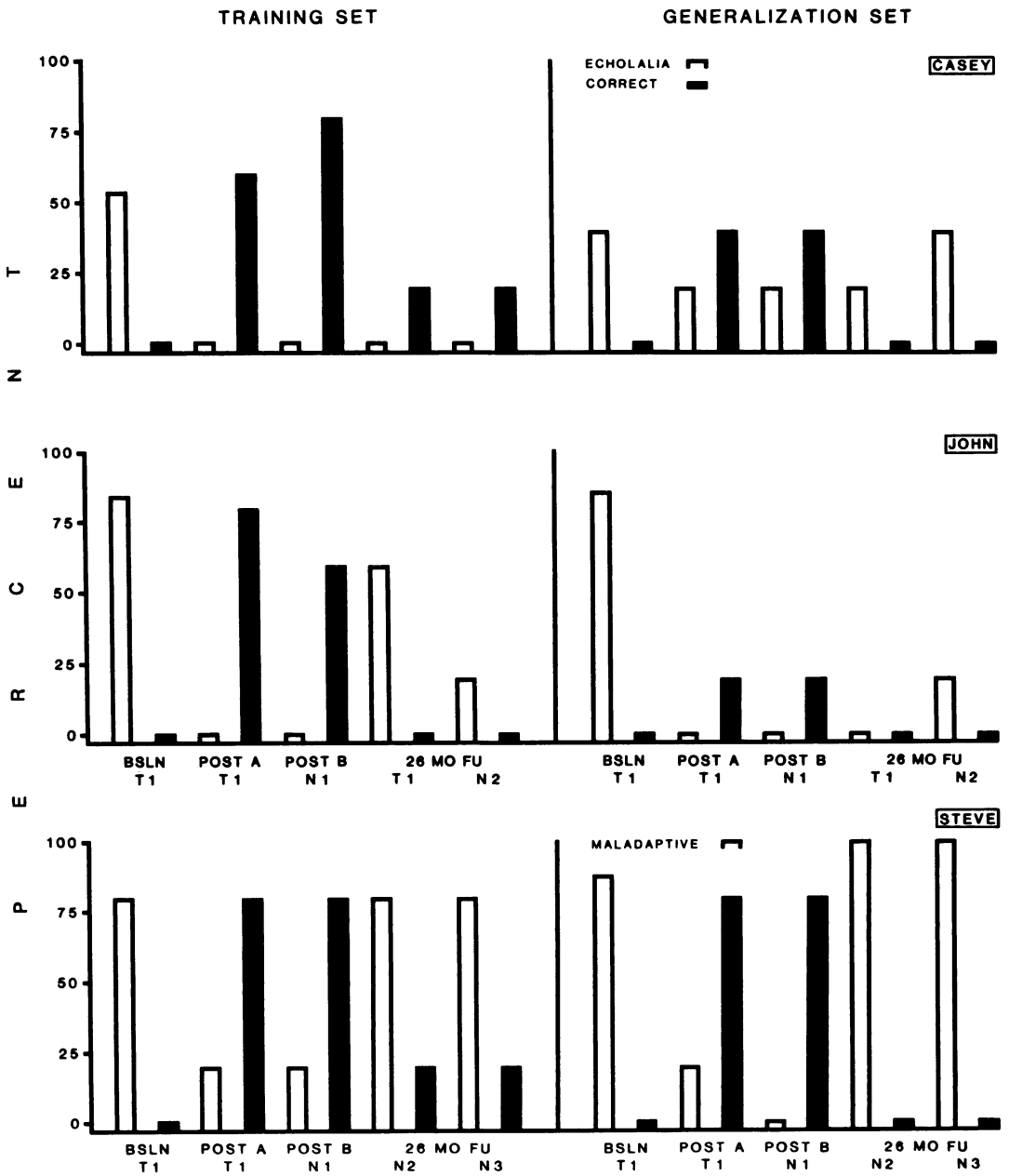

Figure 4. The percentage of echolalic (Casey and John), maladaptive (Steve), and correct responses to a trained and untrained question set. BSLN refers to the baseline conducted by the trainer (T1); Post A and B refer to posttraining generalization assessments conducted by T1 and a novel person (N1), respectively. Casey and John's 26-month follow-up (MO FU) was conducted by T1 and a different novel person (N2), whereas two novel persons (N2 and N3) conducted Steve's follow-up.

the correct ones may not have been in their repertoires was of interest, because such situations generally evoke high levels of echolalia (e.g., see Carr, Schreibman, & Lovaas, 1975).

Reliability. Reliability was assessed in all follow-ups. In each assessment, an observer stood 1 to 2 m behind the subject and immediately scored his or her response before the assessor provided

Table 1
Natural Environment Follow-Up

Subject	Set	Number of questions	% Echolalia		% Correct	
			Baseline	Follow-up	Baseline	Follow-up
Tom	Identification	10	60	10	0	60
	Social interaction	10	80	30	20	30
	Facts and figures	10	100	10	0	70
Eva (Experiment I)	Train 1	6	83.3	66.7	16.7	33.3
	Train 2	6	83.3	66.7	16.7	33.3
Dot (Experiment I)	Train 1	6	50	33.3	33.3	50
	Generalization	6	83.3	33.3	0	33.3
Eva (Experiment II)	Train (3)	15	93.3	93.3	0	6.7
	Generalization (3)	15	60	86.7	6.7	0
Dot (Experiment II)	Train (3)	15	60	60	6.7	13.3
	Generalization (3)	15	40	53.3	6.7	13.3
Casey	Train	5	60	20	0	20
	Generalization	5	40	20	0	0
John	Train	5	100	0	0	0
	Generalization	5	100	0	0	0
Steve	Train	5	80*	80*	0	20
	Generalization	5	80*	80*	0	0
M			73.7	43.7	6.3	22.5

* Also includes perseverative responses.

feedback. Interobserver agreement between the observer and assessor (the primary scorer) was calculated by dividing agreements by agreements plus disagreements times 100. The mean percentage agreements on all target behaviors were as follows: Tom, 97.5; Eva, 100 (Experiment I) and 94.2 (Experiment II); Dot, 97.9 (Experiment I) and 95 (Experiment II); Casey, 90; John, 95; and Steve, 92.5.

DISCUSSION

The training site follow-ups revealed that all subjects except Steve showed lower levels of echolalia. Overall, 4 of the 6 (Tom, Dot, Casey, and John) averaged well below their baseline levels. Eva's results were mixed, because she displayed only slightly lower than baseline levels on some sets but no echolalia on another set. Thus, although, as expected, echolalia was higher at follow-up than at posttraining, there were some lasting reductions for 5 of the 6 subjects.

Not surprisingly, correct responding also deteriorated from posttraining, although everyone except John was above baseline levels on at least one set. This suggested that these 5 subjects had retained some of the skills they were taught without any direct programming of maintenance.

No clear differences in echolalia were apparent between the training and generalization sets, although correct performance was slightly higher on the former. The subjects' correct responding on the generalization sets was gratifying, because no attempt was made in the original studies to facilitate the transfer of stimulus control from the cues to the generalization set questions, whereas this was actively programmed in the training sets (Foxx et al., 1988).

In the questioner comparisons, the novel person always conducted the test first. This was done because we felt that these tests were the more clinically relevant; it was expected that novel person(s) would exert less stimulus control than the original trainer. However, this sequencing did confound the comparisons.

The follow-up in the natural environment was

conducted in an effort to obtain a clinically significant measure of stimulus generalization. Consider that it was the subjects' first formal exposure to the questions since training ended, and it measured the transfer of stimulus control to a new setting. Although the subjects displayed less echolalia and more correct responding than during their initial baseline trials, they did not perform as well as they subsequently did at the training site. Two factors that may have enhanced performance at the training site were increased stimulus control and practice effects, because this assessment always followed the assessment in the natural environment.

Because no attempt had been made to program long-term maintenance, it is possible that the durability demonstrated in the follow-ups could be attributed to such factors as the subjects' levels of functioning, characteristics of the target skill(s), maturation, and performance levels prior to and following training, as well as extraneous variables operating in the posttraining environments. Factors inherent in the training program could also have been responsible, including the use of stringent acquisition criteria (i.e., overlearning), interspersal training, fading training components, and the use of unfamiliar persons in some assessments. Consider that (a) the original training continued within and across sets of stimuli even when correct responding was at or near 100% (i.e., overlearning); (b) subjects were exposed repeatedly to previously mastered questions as generalization to untrained ones was being assessed (i.e., interspersal training); (c) all treatment components were systematically faded, including positive reinforcement and feedback; and (d) multiple trainers and novel questioners were used. Unfortunately, it would have been difficult, if not impossible, to have isolated the maintenance-enhancing contributions of these procedural factors, because a number of other procedures were simultaneously being implemented in order to produce a training effect (see O'Leary & Drabman, 1971).

Although interventions may eventually be developed that actively facilitate maintenance, standard practice has been to program it (Baer & Wolf, 1970; Favell & Reid, 1988) using a variety of maintenance strategies (Albin, Horner, Koegel, & Dunlap, 1987; Stokes & Baer, 1977; Wacker & Berg, 1986). Yet, it remains to be demonstrated convincingly that specific maintenance strategies will produce long-term effects, although some recent reports have been encouraging (e.g., Dunlap, Koegel, Johnson, & O'Neil, 1987; Mank & Horner, 1987). Indeed, the study of maintenance as a dependent variable remains an elusive and difficult endeavor, in part because the evaluation and demonstration of maintenance are fraught with potential confounding effects. For example, most in-depth evaluations of maintenance effects appear to require a number of assessments over time. However, such repeated assessments could confound the evaluation, because opportunities to perform a skill after training can influence skill maintenance (Mank & Horner, 1987).

In summary, a technology of maintenance needs to be developed. Descriptive reports of long-term follow-up data can be useful in this developmental process, because they can suggest factors related to the success or failure of maintenance that should be systematically analyzed (e.g., Foxx & Livesay, 1984). It is hoped that this report represents a small step towards the development of such a technology.

REFERENCES

Albin, R. W., Horner, R. H., Koegel, R. L., & Dunlap, G. (1987). *Extending competent performance: Applied research on generalization and maintenance.* Eugene, OR: University of Oregon, Center on Human Development.

Baer, D. M., & Wolf, M. M. (1970). The entry into natural communities of reinforcement. In R. Ulrich, T. Stachnik, & J. Mabry (Eds.), *Control of human behavior* (Vol. II, pp. 319–324). Glenview, IL: Scott, Foresman.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* **1,** 91–97.

Carr, E. G., Schreibman, L., & Lovaas, O. I. (1975). Control of echolalic speech in psychotic children. *Journal of Abnormal Child Psychology,* **3,** 331–351.

Dunlap, G., Koegel, R. L., Johnson, J., & O'Neil, R. E. (1987). Maintaining performance of autistic clients in community settings with delayed contingencies. *Journal of Applied Behavior Analysis,* **20,** 185–191.

Favell, J. E., & Reid, D. R. (1988). Generalizing and maintaining improvement in problem behavior. In R.

H. Horner, G. Dunlap, & R. L. Koegel (Eds.), *Generalization and maintenance* (pp. 171–196). Baltimore: Paul H. Brookes.

Foxx, R. M. (1990). "Harry": A ten-year follow-up of the successful treatment of a self-injurious man. *Research in Developmental Disabilities,* **11,** 67–76.

Foxx, R. M., Bittle, R. G., & Faw, G. D. (1989). A long term maintenance strategy for discontinuing aversive procedures: A 52-month follow-up of the treatment of a dually diagnosed deaf man's aggression. *American Journal on Mental Retardation,* **94,** 27–36.

Foxx, R. M., Faw, G. D., McMorrow, M. J., Kyle, M. S., & Bittle, R. G. (1988). Replacing maladaptive speech with verbal labeling responses: An analysis of generalized responding. *Journal of Applied Behavior Analysis,* **21,** 411–417.

Foxx, R. M., & Livesay, J. (1984). Maintenance of response suppression following overcorrection: A 10-year retrospective examination of eight cases. *Analysis and Intervention in Developmental Disabilities,* **4,** 65–79.

Koegel, R. L., & Rincover, A. (1977). Research on the difference between generalization and maintenance in extra-therapy responding. *Journal of Applied Behavior Analysis,* **10,** 1–12.

Kohler, F. W., & Greenwood, C. R. (1986). Toward a technology of generalization: The identification of natural contingencies of reinforcement. *The Behavior Analyst,* **9,** 19–26.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology,* **55,** 3–9.

Mank, D. M., & Horner, R. H. (1987). Self-recruited feedback: A cost-effective procedure for maintaining behavior. *Research in Developmental Disabilities,* **8,** 91–112.

McMorrow, M. J., & Foxx, R. M. (1986). Some direct and generalized effects of replacing an autistic man's echolalia with correct responses to questions. *Journal of Applied Behavior Analysis,* **19,** 289–297.

McMorrow, M. J., Foxx, R. M., Faw, G. D., & Bittle, R. G. (1987). Cues-pause-point language training: Teaching echolalics functional use of their verbal labeling repertoires. *Journal of Applied Behavior Analysis,* **20,** 11–22.

O'Leary, K. D., & Drabman, R. (1971). Token reinforcement programs in the classroom: A review. *Psychological Bulletin,* **75,** 379–398.

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

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

Received September 1, 1989
Initial editorial decision November 9, 1989
Revisions received December 11, 1989; January 10, 1990
Final acceptance February 2, 1990
Action Editor, David P. Wacker