ASSESSING AND TRAINING TEACHERS IN THE GENERALIZED USE OF BEHAVIOR MODIFICATION WITH AUTISTIC CHILDREN¹

ROBERT L. KOEGEL, DENNIS C. RUSSO, AND ARNOLD RINCOVER

UNIVERSITY OF CALIFORNIA, SANTA BARBARA; THE JOHN F. KENNEDY INSTITUTE AND THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE, AND UNIVERSITY OF NORTH CAROLINA, GREENSBORO

This study investigated the feasibility of developing reliable, valid criteria for measuring and training the skills necessary to teach autistic children. The behaviors of 11 teachers and 12 autistic children were recorded in a series of different teaching situations. Teacher-training was initiated at different times for different teachers. The results showed: (1) it was possible to assess empirically whether a teacher was correctly using defined behavior-modification techniques; (2) generally, for any given session, systematic improvement in the child's behavior did not occur unless the teacher working in that session had been trained to use the techniques to a high criterion; (3) all 11 teachers were rapidly trained to use these techniques; and (4) the teachers learned generalized skills effective with a variety of children and target behaviors.

DESCRIPTORS: response generalization, behavior therapy, behavioral instruction, contingency management, stimulus generalization, discrimination training, teacher training, teacher behavior, generalization, autistic children

In the past 10 years, a relatively large number of investigations have shown that behavior-modification procedures can be effective in the treatment of autistic children (*e.g.*, Hewett, 1965; Koegel and Rincover, 1974; Lovaas, 1966; Lovaas, Freitas, Nelson, and Whalen, 1967; Metz, 1965; Risley and Wolf, 1967; Wolf, Risley, and Mees, 1964). In reviewing the literature, Kozloff (1975) and Lovaas and Koegel (1973) pointed out that: (1) behavior modification has produced considerable treatment gains in autistic children; and (2) these results have been replicated in numerous experimental laboratories. However, it is not at all certain that any given teacher or therapist could replicate these results (Lovaas, Koegel, Simmons, and Stevens-Long, 1973). This conclusion was based primarily on the fact that there are currently no published criteria for measuring whether or not a given individual is adept at using behavior-modification techniques with autistic children. Criteria for teacher certification in special education generally do not include a behavioral assessment of a teacher's ability to teach these children. Furthermore, most autistic children are still excluded from school programs on the grounds that the teachers are

unable to provide instruction for such children. The assumption generally made by school administrators is that the teachers are teaching correctly, and the children are unable to learn.

As a result, it seems timely to assess whether or not teachers are (or even can) use the procedures that have been demonstrated in the literature to be effective in teaching such children. The present experiment was designed to: (1) assess empirically whether or not a child's teacher is using behavior-modification procedures correctly; and (2) assess the necessity and feasibility of training teachers to teach autistic children effectively.

¹This investigation was supported by the United States Public Health Service Research Grant MH 11440 and MH28210 from the National Institute of Mental Health, and by State of California Research Grant 42-00000-000832, Title VI-B of the United States Elementary and Secondary Education Act. Many people assisted in this investigation. We are grateful for the help of Polly Dimalanta, Nancy Woodfield Everett, Glen Dunlap, Frank Schlosser, and Peggy Pulleyn. We particularly appreciate the helpful suggestions Laura Schreibman provided throughout this investigation. Reprints may be obtained from Robert L. Koegel, Social Process Research Institute, or Speech Department, University of California, Santa Barbara, California 93106.

METHOD

Subjects and Setting

Eleven teachers participated in this investigation. Ten were under 30 yr of age. Four resided in cities of Southern California; the other seven in New Jersey. At the time of this study, nine teachers had at least 2 yr of previous teaching experience with autistic children. None of the 11 teachers had any formal training in behavior modification but all had read in the area. All the teachers were employed in special-education classrooms. However, the selection was not random, in that all of the teachers requested training, and all of them generally appeared eager to participate. No teacher was denied participation.

Twelve autistic children also participated. They were randomly selected either from a participating teacher's classroom, or from one of the training sites. Five of the children were enrolled in an experimental classroom at the University of California at Santa Barbara, four were enrolled in a pilot classroom for autistic children in Orcutt, California, the other three were enrolled in the Princeton Child Development Institute in Princeton, New Jersey. The children ranged in chronological age from 5 to 13 yr. Ten were reported to be untestable on standardized IQ tests; the other two had reported IQ scores of 28 and 32. The results of standardized social maturity tests placed all of the children between 2 and 4 yr. All the children had been excluded from existing public school classrooms, and had been diagnosed as autistic by agencies not associated with this study. They engaged in little appropriate play or social behavior and had almost no appropriate speech. Six children were essentially mute, and the others were primarily echolalic. They evidenced little, if any, self-help behaviors (*e.g.*, feeding themselves, dressing themselves, *etc.*), and frequently engaged in tantrum and self-stimulatory behaviors. In general, they were probably representative of children in the more severe half, with regard to severity of autism.

During each session, one teacher was seated facing one child at a 50.8- by 76.2-cm table. All the stimuli necessary to teach the behavior assigned, such as different color blocks, different pictures, *etc.*, and a supply of food reinforcers were present on the table.

Target Behaviors

Target behaviors were selected for each child according to the individual level of academic skill. Also, some target behaviors were assigned that had already been learned by a child, so that we could determine if the teacher could maintain correct responding. A total of 27 target behaviors was selected. Each was selected from one of the following categories: self-help skills (*e.g.*, shoe-tying); arithmetic skills (*e.g.*, "1" *versus* "2"); writing skills (*e.g.*, tracing letters); picture labelling (*e.g.*, "bear" *versus* "horse"); abstract language skills (*e.g.*, "big" *versus* "small"); speech skills (*e.g.*, verbal imitation—"Say 'Mama'").

Teacher-Training Procedure

The teachers first were asked to read a training manual that described examples of correct and incorrect use of five categories of behavior-modification procedures. Each teacher also was shown videotapes illustrating correct and incorrect use of each of the procedures. The teacher then attempted to teach a child a new target behavior. About once every 5 min, the trainer would interrupt the session and provide feedback as to whether or not the teacher was conducting the teaching according to the operational definitions described below. If errors occurred in any category, the trainer (either an author or a university student familiar with these procedures) so informed the teacher, and then modelled the procedure according to its operational definition. Feedback was very brief, usually taking only a few seconds, in order to produce minimal interference with the session. For example, the trainer might say, "You are using good prompts, but your S^Ds are too long. Try saying, 'Touch red.'" Approximately once every half hour, the teach-

ers were given more elaborate feedback. These sessions continued until the trainer was subjectively satisfied with the teacher's performance. All the teachers completed training within less than 25 hr (usually either five days of 5 hr per day, or 10 days of 2.5 hr per day). None of the teachers reported either training schedule to be excessively demanding.

Design

The design of this experiment was a modified multiresponse baseline (Birnbrauer, Peterson, and Solnick, 1974). For a predetermined number of baseline sessions (1, 3, or 4 sessions) each teacher attempted to teach one to four randomly assigned target behaviors. In each session, one teacher worked with one child on one behavior. During successive sessions, a teacher might be assigned to a different child and/or a different target behavior. After training, the teacher again attempted to teach one to three behaviors. The

tasks assigned both before and after training were always different from those used during training. During each session, we simultaneously measured the correctness of the teacher's procedures and the learning of the child to assess whether changes in the children's behavior were a function of changes in the teacher's behavior and whether the training produced the teachers' generalized use of behavior-modification procedures across (autistic) children and tasks.

Assessment of Teacher Use of Behavior Modification Procedures

In each session, at least one observer recorded data on the teachers' use of five aspects of behavior-modification procedures. Each observer, whether familiar or naive with respect to the conditions of the experiment, had completed at least one college level course in behavior modification. The observers were given the following instructions:

Table 1

The assignment of target behaviors (B) and children (Ch) for each session of each teacher in this experiment.

	Base	After Training						
Teacher 1	Ch ₁ B ₁	Ch ₃ B ₂	Ch ₁ B ₃					
Teacher 2	Ch ₂ B ₄	Ch ₃ B ₅	Ch ₁ B ₆					
Teacher 3	Ch ₁₀ B ₇	Ch ₁₁ B ₅	Ch ₉ B ₉	Ch ₁₁ B ₃	Ch ₈ B ₁₀	Ch ₉ B ₇		
Teacher 4	Ch ₂ B ₁₆	Ch ₂ B ₈	Ch ₁ B ₅					
Teacher 5	Ch ₆ B ₁₁	Ch ₇ B ₁₉	Ch ₆ B ₁₀	Ch ₇ B ₂₀	Ch ₉ B ₈	Ch ₆ B ₂₁	Ch ₆ B ₁₀	
Teacher 6	Ch ₃ B ₁₇	Ch ₂ B ₁₈	Ch ₂ B ₅					
Teacher 7	Ch ₄ B ₁₁	Ch ₅ B ₁₂	Ch ₄ B ₁₃	Ch ₈ B ₁₄	Ch ₈ B ₁₅	Ch ₄ B ₈	Ch ₅ B ₅	
Teacher 8	Ch ₁ B ₂₂	Ch ₁ B ₂₃	Ch ₁ B ₂₄					
Teacher 9	Ch ₁₂ B ₇	Ch ₁₂ B ₇	Ch ₁₂ B ₇	Ch ₁₂ B ₇	Ch ₁₂ B ₇	Ch ₁₂ B ₇		
Teacher 10	Ch ₃ B ₂₅	Ch ₁ B ₂₆	Ch ₃ B ₂₃					
Teacher 11	Ch ₃ B ₂₇	Ch ₃ B ₈	Ch ₂ B ₁₀					
	1	2	3	4	5	6	7	
				Task				

SCORING INSTRUCTIONS

Observe a session for 30 seconds, then record during the next 30 seconds. Use the definitions given below to make a decision about the adequacy of each aspect being assessed. If you do not see a procedure being used in a given period, leave the appropriate space blank.

S^Ds: The Teacher's Instructions to the Child

1. The S^D should be clear and *discriminable*, that is, it should stand apart from anything else that the teacher says. A good S^D has a distinct onset and a distinct offset.
2. The S^D should be *appropriate* to the task. If the teacher is teaching the child to point to a red card, he should not mistakenly say, "Point to the blue card", or "Hand me the red card".
3. The S^D should be *consistent* with that given on the previous trial. Exception: the S^D can appropriately vary when a discrimination task is being taught, and the child has reached criterion on each of the individual stimulus items.
4. The S^D should be *uninterrupted*.
5. When the S^D is presented, the child should be *attending*. The child should not be engaging in off task behavior. The child should be sitting quietly, not engaged in any overtly disruptive behavior when the verbal component of the S^D is presented. The child should be looking either at the task at hand or at the teacher.

Prompts: Guiding the Child to Respond Correctly

1. The prompt must be *effective*, that is, it must evoke a correct response.

Shaping: Reinforcing Successive Approximations to the Target Behavior

1. Each reinforced response should be at least as good as the last one (score shaping with reference only to responses actually observed during the scoring period).

Consequences: The Teacher's Behavior after the Child Responds

1. Consequences should be delivered *immediately* after the response. Immediately is defined as within three seconds.

2. Consequences should be *contingent*, that is, reinforcers only for correct responses, and extinction only for incorrect responses.

3. The consequence should be *unambiguous*. A "No" said with a smile, or a "Good boy" given with a frown are ambiguous.

4. Consequence should be *consistent*. Reinforcers should follow each correct response, during acquisition. Similarly, punishment, if used, should follow every instance of bizarre or inappropriate behavior.

5. Consequences should be *effective*. They should be selected individually for each child. Reinforcers should be something the child will approach, and punishers something the child will avoid.

Discrete Trials

1. Each trial must have a distinct onset and offset, and a discrete intertrial interval. An intertrial interval consists of any noticeable period of time in which no on-task responding is required of the child.

During every 30-sec recording interval, observers scored each of the five categories as "correct" (fulfilled all aspects of the definition of the procedure for all of the trials occurring in that interval), "incorrect" (did not fulfill the definition in some way during any trial(s) within the interval), or "NA" (not applicable—procedure was not observed in the 30-sec period).

The observers simply made a check mark in the appropriate box for each of the five procedures. An overall index of the teacher's behavior was then computed for each session. First, the number of "correct" 30-sec intervals was totalled for each procedure (summing across categories). Second, the number of correct plus incorrect intervals was totalled. Then, the first total was divided by the second, and multiplied by 100 to give a percentage. The average of the five percentages was computed in order to give an overall index of the teacher's behavior during each session.

Measurements of the Child's Performance

In each session, data were recorded continuously on the correctness of the child's perform-

ance of the target behavior. On each trial, an observer recorded whether the child's response was correct, incorrect, prompted, or an approximation of the target behavior. If an approximation was performed, the observer noted whether or not it was a closer approximation to the target behavior than the last reinforced response. If a prompt was used to bring about a correct response from the child, the observer noted whether the prompt was faded over the course of training.

In addition to these trial-by-trial measures of student improvement, we obtained a summary measure of whether or not a child was learning in each session. A "+" was recorded if the child's responding during the last 10 trials of the session was improved (*i.e.*, a closer approximation, a reduced prompt, or a higher percentage of correct responding) as compared to the first 10 trials of the session. Conversely, a "0" was recorded if responding during the last 10 trials showed no improvement, or a deterioration, when compared to the first 10 trials.

Reliability

Reliability of recording the teachers' use of behavior-modification procedures was assessed by the following procedure. Two observers (usually one trainer and one naive observer) independently (data sheets shielded from each others' sight) recorded data for 24 of the 47 sessions in this experiment. In addition, two naive observers independently recorded data from 16 videotaped sessions shown to them in a randomized order. An agreement or disagreement between observers was recorded for each of the five procedures in each 30-sec interval. An agreement was counted if both observers marked the same square on their data sheets, and a disagreement was counted if only one observer marked a given square. Reliability for each session was calculated by dividing number of agreements (summing across categories) by the total number of agreements plus disagreements for that session. The mean per cent agreement for these sessions was 94.6%. The range (82% to 100%)

of reliability percentages was identical for both the *in vivo* and videotape recordings.

For child responses, one trainer and one naive observer independently and continuously recorded the child's behavior in 24 (of 47) randomly selected sessions. The observers were in agreement if, on a given trial, they both recorded the response as correct, incorrect, prompted, or a successive approximation. Reliability was calculated for each session by dividing the number of agreements by the number of agreements plus disagreements in that session. The mean reliability for these measures was 91.7%, with a range of 84% to 100%. For the summary (improvement *versus* no improvement) measure recorded in each of these 24 sessions, the observers were in complete agreement.

RESULTS

Measurement of Teacher Behavior Before and After Training

Data for all 11 teachers, showing their use of behavior-modification procedures during baseline and posttraining sessions, are presented in Figure 1. Each point represents a composite score based on the average per cent correct use of the five categories of behavior-modification procedures during a given session. The ordinate shows the per cent correct use of the behavior-modification procedures; sessions are presented on the abscissa.

During the baseline sessions, all 11 teachers evidenced low percentages. The percentages for 10 of the 11 teachers were consistently below 58%, and in two instances were as low as 0%. The remaining teacher achieved scores of 63%, 91%, and 71% during three baseline sessions.

In the posttraining sessions, the 11 teachers scored between 90% and 100% correct use of behavior-modification procedures for 25 of the 26 sessions. For each teacher, this change represents both a decrease in the frequency of incorrect intervals, and an increase in the frequency of correct intervals, the number of recording intervals per session remaining constant. All 11

Fig. 1. The session-by-session per cent of correct use of behavior-modification procedures by teachers before and after teacher training.

teachers showed considerable increases in their correct use of behavior-modification procedures after the training program. Furthermore, after training, teacher behavior generalized to new tasks.

Measurement of Children's Behavior Before and After Teacher Training

To determine whether correct performance of the five techniques of behavior modification was

sufficient for teaching autistic children, we assessed whether there was an increase in the child's (unprompted) correct responses during each session. During 20 of the 21 baseline sessions, the children showed either no improvement or a decrease in correct responding. In marked contrast, after the teachers were trained, the children improved their level of correct responding during each of 26 posttraining sessions.

Figure 2 shows all of the data for four representative teachers. Per cent correct responses performed on the particular task being taught to

a given child during a given session is shown on the ordinate, and blocks of 10 training trials are shown on the abscissa. Single vertical lines separate different sessions; double vertical lines signify the time at which the teacher was trained. Since different behaviors were being taught to different children in each session, the number of trials varied (from 20 to 100 trials) over sessions.

The first teacher (Teacher 3 in Figure 2) attempted to teach three different behaviors to three different children during the baseline sessions. In all three instances, the data are similar;

Fig. 2. The per cent of correct responding by the children before and after teacher training. Per cent correct responses are plotted for each block of 10 trials within a session.

the children showed no improvement. In fact, during these sessions all children evidenced a lower percentage of correct responding at the end of the session than at the beginning. The results for the other three teachers during baseline sessions are similar. Out of a total of 14 baseline sessions, 13 showed either no improvement or a reduced level of correct responding by the end of the session.

After the four teachers had been trained, the children's correct responding increased. Every child showed large increases in the per cent of correct responses during each of 12 sessions conducted after teacher training. In four sessions, the teachers introduced new prompt, shaping, or reinforcement-schedule fading techniques (marked in the figure by asterisks). Although temporary drops in correct responding occurred in these cases, each child quickly regained at least 90% correct responding.

Generalized Use of Behavior Modification Procedures

The teachers were not specifically trained to teach the particular behaviors used in the post-training assessment; these results show that the training procedures produced generalized correct applications of the procedures taught during training. In addition, it may be important to note that posttraining Sessions 2 and 3 for two teachers (Teachers 5 and 7) were conducted two and four months after training, leading us to feel optimistic about the durability of the results of training.

DISCUSSION

The results showed the following. (1) The assessment procedure reliably differentiated between teachers with respect to their correct use of defined behavior-modification procedures. Generally, when teachers failed to use the procedures correctly, then efforts to teach the children produced no measurable improvement. Conversely, when the teachers showed consistently high percentages of using the procedures

correctly, their teaching was effective in producing gains in the children's responding. (2) The training procedure was effective, for all teachers, in increasing the correct use of the procedures. Further, after training, the teachers' correct use of these procedures generalized to new target behaviors and new children.

The training procedure was a package involving the use of modelling, feedback, and training manuals. We did not do a component analysis and as such cannot be sure which components contributed to each aspect of the results. It is possible that certain components influenced acquisition of the techniques and others influenced generalization. A component analysis of this or any other teacher-training program might lead to interesting results (Bandura, 1976; Kazdin and Moyer, 1976).

There was no attempt to determine correlations between intermediate levels of teacher proficiency (*e.g.*, 50%, 70%, *etc.*), and the rates of change in correct responding of the children. A 90% proficiency on the five procedures was almost always sufficient to ensure effective modification of the behavior of these severely autistic children. It may, however, be interesting to note that, during one of the baseline sessions for Teacher 9, there was no improvement in child performance with a high level of correct teacher performance. Thus, other categories of teacher behavior may be important to measure and some categories may be more important than others. Possibly some other variable, such as consistency in teacher behavior across sessions, may to some extent influence the results (note that this teacher worked with the same child in every session). As Parsons, Baer, and Baer (1974) and Herbert and Baer (1972) pointed out, there can be value in improving the overall performance of someone who is already moderately competent.

The five categories measured and trained in this investigation are not the only ones that could be studied. Several investigators (*e.g.*, LeLaurin and Risley, 1972; Sailor and Taman, 1972) have suggested that the physical config-

uration of the teaching environment may be an important variable influencing learning. In our own laboratory, we have investigated various procedures that are effective for teaching a group of autistic children in a classroom setting (Koegel and Rincover, 1974). As these or other new functional relationships are discovered, then new categories should also be included in the assessment and training procedures. The contribution that the present study offers at this time is not in providing an all-inclusive list of procedures to measure but, rather, in suggesting that it is important to measure carefully both the behavior of the teacher and the behavior of the child in order to discover functional relationships between the two.

REFERENCES

- Bandura, A. Effecting change through participant modeling. In J. D. Krumboltz and C. E. Thoresen (Eds), *Counseling methods*. New York: Holt, Rinehart & Winston, 1976. Pp. 248-265.
- Birnbrauer, J. S., Peterson, C. P., and Solneck, J. V. The design and interpretation of studies of single subjects. *American Journal of Mental Deficiency*, 1974, 79, 191-203.
- Hewett, F. M. Teaching speech to an autistic child through operant conditioning. *American Journal of Orthopsychiatry*, 1965, 35, 927-936.
- Kazdin, A. K. and Moyer, W. Training teachers to use behavior modification. In S. Yen and R. McIntire (Eds), *Teaching behavior modification*. Kalamazoo, Michigan: Behaviordelia, 1976. Pp. 171-200.
- Koegel, R. L. and Rincover, A. Treatment of psychotic children in a classroom environment: I. Learning in a large group. *Journal of Applied Behavior Analysis*, 1974, 7, 45-59.
- Kozloff, M. *Educating children with learning and behavior problems*. New York: John Wiley and Sons, 1975.
- LeLaurin, K. and Risley, T. The organization of day-care environments: "zone" versus "man-to-man" staff assignments. *Journal of Applied Behavior Analysis*, 1972, 5, 225-232.
- Lovaas, O. I. A program for the establishment of speech in psychotic children. In J. K. Wing (Ed), *Early childhood autism: clinical, educational and social aspects*. London: Pergamon Press, 1966. Pp. 115-144.
- Lovaas, O. I., Freitas, L., Nelson, K., and Whalen, C. The establishment of imitation and its use for the establishment of complex behavior in schizophrenic children. *Behaviour Research and Therapy*, 1967, 5, 171-181.
- Lovaas, O. I. and Koegel, R. L. Behavior modification with autistic children. In C. Thoresen (Ed), *Behavior modification in education*. Chicago: University of Chicago Press, 1973. Pp. 230-258.
- Lovaas, O. I., Koegel, R. L., Simmons, J. Q., and Stevens-Long, J. Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis*, 1973, 6, 131-165.
- Metz, J. R. Conditioning generalized imitation in autistic children. *Journal of Experimental Child Psychology*, 1965, 2, 389-399.
- Parsonson, B. S., Baer, A. M., and Baer, D. M. The application of generalized correct social contingencies: an evaluation of a training program. *Journal of Applied Behavior Analysis*, 1974, 7, 427-437.
- Risley, T. and Wolf, M. Establishing functional speech in echolalic children. *Behaviour Research and Therapy*, 1967, 5, 73-88.
- Sailor, W. and Taman, T. Stimulus factors in the training of prepositional usage in three autistic children. *Journal of Applied Behavior Analysis*, 1972, 5, 183-192.
- Wolf, M., Risley, T., and Mees, H. Application of operant conditioning procedures to the behavior problems of an autistic child. *Behaviour Research and Therapy*, 1964, 1, 305-312.

Received 27 October 1975.
(Final acceptance 17 May 1976.)