*GENERALIZATION AND MAINTENANCE OF TREATMENT
GAINS OF BEHAVIORALLY HANDICAPPED STUDENTS
FROM RESOURCE ROOMS TO REGULAR CLASSROOMS
USING SELF-EVALUATION PROCEDURES*

GINGER RHODE, DANIEL P. MORGAN, AND K. RICHARD YOUNG

DEPARTMENT OF SPECIAL EDUCATION
UTAH STATE UNIVERSITY

This study provided six behaviorally handicapped elementary school students with a short-term resource treatment to bring their behavior under the control of a combination of treatment procedures emphasizing self-evaluation. Once acceptable levels of appropriate behavior were maintained with only minimal external reinforcement and students were accurately self-evaluating their own work and behavior, generalization and maintenance of behavior gains were sought by introducing a reduced form of the self-evaluation procedures in the students' regular classrooms. A multiple baseline across pairs of subjects design was used to examine individual student's behavior. Analysis of the results of the study indicated that students transferred and maintained high levels of appropriate classroom behavior in their regular classrooms, once self-evaluation procedures were extended into those settings. For four of the six students, all extra-training components were faded. Only two students required a modified form of the original intervention to maintain behavior gains in their regular classrooms.

DESCRIPTORS: self-evaluation, generalization, children, classroom behavior

The success of special education programs is often measured by the extent to which handicapped students' academic and behavioral improvements transfer from their special education classrooms to mainstream regular classroom settings (Reynolds, 1979). Facilitating generalization and maintenance of gains from special education settings to regular classrooms is a key ingredient in the overall success of education programs for handicapped students (Asher & Taylor, 1981; Gottlieb & Leyser, 1981). There is little evidence to date, however, that gains made in special education programs are con-

sistently transferred to regular classroom settings. Typically, when handicapped students are removed from their regular classrooms for treatment in resource rooms or other special education settings, they exhibit academic and behavioral improvements in the special education settings but do not maintain the improvements after returning to their regular classrooms (Walker & Buckley, 1972; Wildman & Wildman, 1975).

In recent years, research in psychology and special education has begun to focus on the use of self-management training as a means of facilitating generalization and maintenance of treatment gains (Drabman, Spitalnik, & O'Leary, 1973; Robertson, Simon, Pachman, & Drabman, 1979; Turkewitz, O'Leary, & Ironsmith, 1975). Self-management training is an appealing means of promoting behavioral improvements in nontreatment or regular classroom settings, since students who receive self-management training depend less on their teachers for

The authors acknowledge the students and faculty of the Edith Bowen School for their cooperation in conducting the research reported in this manuscript. Appreciation is also expressed for the research support made possible by a grant from Utah State University's College of Education/Bureau of Research Services which made the study possible. Reprints may be obtained from Daniel P. Morgan, Department of Special Education, UMC 65, Utah State University, Logan, Utah 84322.

guidance, reinforcement, and control than they did prior to training (Workman & Hector, 1978). Treatment gains that are regulated by self-management procedures have also been reported to be more resistant to extinction than those established by externally regulated procedures alone (Johnson, 1970; Kanfer & Duerfeldt, 1967; McLaughlin, 1976; Rosenbaum & Drabman, 1979). The present study was designed to extend the results of previous research by modifying and extending three related studies (Drabman *et al.*, 1973; Robertson *et al.*, 1979; Turkewitz *et al.*, 1975) which emphasized a type of self-management—self-evaluation.

In the studies by Drabman *et al.* (1973), Robertson *et al.* (1979), and Turkewitz *et al.* (1975), self-evaluation was instituted through the use of a "matching" procedure whereby students were initially taught to evaluate their classroom behavior on the same rating scale that their teachers used. For the students, matching meant closely approximating the teacher's rating of their behavior. Matches were called for less frequently as the studies progressed, with the students continuing to approximate their teacher's ratings consistently. Eventually, the teacher component of the procedure was faded, leaving the primary responsibility for evaluating their behavior to the students themselves.

In the studies by Drabman *et al.* (1973) and Turkewitz *et al.* (1975), generalization and maintenance of improvements were studied in after-school "adjustment" classes, whereas the study by Robertson *et al.* (1979) examined them within the students' self-contained special education classroom. Of the three studies (Drabman *et al.*, 1973; Robertson *et al.*, 1979; Turkewitz *et al.*, 1975), only Turkewitz *et al.* observed students in their regular classrooms to determine whether improved behavior had generalized. In examining students' behavior in their regular classrooms in comparison to control students, Turkewitz *et al.* found that improvements for experimental students had not generalized to regular classrooms. No attempts were reported to prompt or teach the use of self-evaluation pro-

cedures in students' regular classrooms in the Turkewitz *et al.* research. The study by Robertson *et al.* (1979) examined generalization and maintenance of improvements from morning to afternoon, on days when self-evaluation procedures were not in effect, and when a substitute teacher was present. Of the three related studies (Drabman *et al.*, 1973; Robertson *et al.*, 1979; Turkewitz *et al.*, 1975) none programmed for the occurrence of generalization and maintenance in other settings. As Stokes and Baer (1977) have pointed out, generalization is not a passive phenomenon that can be expected to occur on its own but must be actively programmed.

Self-management training appears to be a viable means of promoting generalization in students with learning and behavior problems, but few studies have demonstrated a practical application of self-management training with behaviorally handicapped students in a well-controlled study in naturalistic settings. In fact, Walker (1979) stated that one of the characteristics of behaviorally handicapped students is a lack of self-control. The goal of the present study was to investigate the effectiveness of a self-management procedure in facilitating the generalization of appropriate classroom behavior of behaviorally handicapped students from a special education program to a regular classroom setting. This research differed from previous research in that it took place in a regular elementary school and in that specific procedures designed to promote generalization and maintenance of treatment gains were specifically programmed for use in students' regular classrooms.

METHODS

Students

Six behaviorally handicapped elementary students were selected from a group of students who were referred by their teachers for excessive classroom behavior problems. To determine eligibility for placement in the resource room as

behaviorally handicapped, students were observed in their regular classrooms on at least five occasions for 15 min per observation and were administered individual standardized intelligence and achievement tests. Standardized tests used to document placement eligibility included the Peabody Individual Achievement Test, the Woodcock Reading Mastery Tests, the Key Math Diagnostic Arithmetic Test, and the Slosson Intelligence Test. Parents and teachers provided additional information in the form of the Walker Problem Behavior Identification Checklist, a School Problem Behaviors Checklist, a Parent Information Survey, and anecdotal reports.

Final selection of students who were included in the study was based on the diagnostic criteria established by Utah State Office of Education for behaviorally handicapped students and was mutually agreed to by teachers, the students, parents, and the school principal. Individualized Education Programs (IEP's) were also developed for each of the six students participating in the study, in accordance with Public Law 94-142 (Education of All Handicapped Children Act).

Grade placements for participating students were as follows: 2, first grade; 1, second grade; 1, fourth grade; and 2, fifth grade. The mean age of students at the beginning of the study was 8.7 yr, with a range from 6.5 to 10.9. Prior to participation in the study, three of the students were enrolled full-time in regular classrooms (Subjects 2, 3, and 5). Two students were enrolled full-time in a self-contained classroom for behaviorally handicapped students (Subjects 4 and 6), and one participated part-time in a regular classroom and part-time in a self-contained classroom for behaviorally handicapped students (Subject 1).

At the time of placement, the students were experiencing moderate to severe problems in their present school placements. Problems included high rates of disruptive classroom behavior most of the time, refusal to work and complete assigned tasks, frequent inappropriate talking in class, frequent out-of-seat occurrences,

noncompliance with teacher requests, aggression toward peers, bizarre talk, and daydreaming.

Numerous interventions had already been attempted for three of the six students. For these three individuals, intervention included exposure to varied behavior management programs in previous and present schools (Subjects 4 and 6), previous placement in a self-contained classroom for behaviorally handicapped students in a different school (Subject 4), weekly mental health therapy sessions (Subject 6), placement under the jurisdiction of a probation officer (Subject 6), and prescription of medication (Subject 4, who was taken off medication 2 wk before the study began). Although school records indicated behavior problems had been evident for some time for the remaining three students (Subjects 2, 3, and 5), no previous intensive intervention had been attempted.

Setting

The study took place in the Edith Bowen School on the Utah State University campus. The Edith Bowen school is an elementary school serving approximately 175 students in kindergarten through sixth grade. Treatment took place in several classroom settings within the school. The resource room treatment occurred in a separate room from the students' regular classrooms. The 8 m $\times$ 12 m room contained work tables and chairs as well as several large tables on which instructional materials were stored. During resource room sessions, six of the work tables and chairs were placed in two rows of three each, so that students could work individually at the tables on assigned tasks. During the generalization phase of the study, treatment occurred in the six different regular classrooms in which the students were placed for the majority of the school day. The term "regular" classroom refers to students' current school placements, whether they were regular education classrooms, a self-contained classroom for the behaviorally handicapped, or a combination of the two. Each of these settings was similar in size (approximately 12 m $\times$ 18 m) and

other physical characteristics (e.g., desks for 25 students, learning centers, quiet reading areas).

Procedures: Phase I

Initial baseline and the first phase of the study took place in a resource room setting. The six students were exposed to a baseline period and six subsequent treatment conditions in the resource room over a 15- to 17-wk period before being returned to their current classroom placements on a full-time basis. The 3 hr of training each week took place over four school days. On two of the days, sessions were each 1 h long; on the other two days, sessions were each 30 min. The purpose of the resource room treatment was to increase students' appropriate classroom behavior to acceptable levels and to teach them to use self-evaluation procedures for generalizing and maintaining the behavioral changes in their regular classrooms. Successful treatment in the resource room was a prerequisite for implementation of the self-evaluation procedure in the regular classrooms. The experimental conditions for baseline and phase I are described below.

Condition 1: resource room baseline. Individualized academic instruction was implemented in the resource room using curriculum materials suited to the needs of individual students. The students spent most of their time on math and reading tasks; two students also practiced handwriting, and one student received help with his regular classroom spelling assignment. Drill and practice assignments were selected from a wide variety of instructional materials. The resource room teacher was instructed to use basic classroom management procedures (i.e., contingent praising and ignoring, as well as reprimands when deemed necessary) to manage classroom behavior. No form of token reinforcement was used. It should also be noted that in addition to data collected in the resource room, data were also collected on the students in their regular classrooms during this condition and throughout the remainder of the study.

Condition 2: token reinforcement/systematic verbal feedback. Classroom rules for the resource room were introduced and discussed with students. The teacher and students modeled and role played examples and nonexamples of the rules. Feedback, correction, and clarification by the teacher were provided to ensure that students understood expectations for them in that setting.

Points were given by the teacher for appropriate classroom behavior (i.e., following the rules) and correct academic work. Points could be exchanged for small toys, candy, and other snacks at the end of each session. Students were told that the ratings they received for work and behavior were to be converted to a corresponding number of points (e.g., a rating of "5" would convert to 5 points). Students were rated on a 0 to 5 rating scale based on the following criteria:

- 5 = *excellent*—Followed all classroom rules entire interval; work 100% correct.
- 4 = *very good*—Minor infraction of rules (i.e., a talk-out or out-of-seat occurrences), but followed rules rest of interval; worked almost entire interval; work at least 90% correct.
- 3 = *average*—Did not follow all rules for the entire time, but no serious offenses. Followed rules approximately 80% of the time (i.e., inappropriate behavior may have involved talking-out or out-of-seat occurrences).
- 2 = *below average*—Broke one or more rules to extent that behavior was not acceptable (e.g., aggressive, noisy, talking) but followed rules part of the time. Work approximately 60 to 80% correct.
- 1 = *poor*—Broke one or more rules almost entire period or engaged in higher degree of inappropriate behavior most of the time. Work between 0 and 60% accurate.
- 0 = *totally unacceptable*—Broke one or more rules entire interval. Did not work at all or work all incorrect.

Each day students were also given a point card for use during each resource room session. The classroom rules were printed on the back of each card. The teacher marked the cards with a colored pen to indicate her award of 0 to 5 points each for academic work and classroom behavior. Resource room sessions were initially divided into 15-min intervals, and students were given points and feedback for their academic work and classroom behavior at the end of each interval. Between the times that points were awarded, the teacher circulated among students to correct work, provide assistance, and praise individual work and the group as a whole for following classroom rules.

Condition 3: matching self-evaluations with the teacher/100%. The teacher asked students to rate their own academic work and behavior by awarding themselves points on the 0 to 5 point scale. The teacher also continued to rate students' work and behavior independently by awarding points in a similar fashion. Students matched their self-evaluations with the teacher's evaluations of their work at the end of each 15-min interval. The teacher would first ask each student why he thought he should earn the number of points he had given himself. The teacher then explained her rating. If a student was within one point of the teacher's ratings (either higher or lower), he kept the number of points he had given himself. If the two ratings exactly matched (even if they were 0), the student earned a bonus point. Bonus points could be earned for both work and behavior for perfect matches. If there was more than a one point difference (higher or lower) between a teacher and student rating, no points were earned for that interval. Thus, beginning with this condition, both the teacher and the student rated the student's work and behavior and marked the point cards. The teacher's ratings were recorded with a colored pen and student ratings were recorded with a pencil.

Condition 4: elimination of matching/(50%). Only half of the students were selected to match their self-evaluations with the teacher at the end

of each interval. Names were drawn out of a container to determine which students would match. Students who did not match self-evaluations received the number of points they had given themselves. Students who matched received points in the same manner as condition 3. Although the teacher did not publicly match her ratings with all students, she continued to rate all students privately and recorded the points separately.

Condition 5: elimination of matching/(33⅓%). This condition was similar to conditions 3 and 4 except that only two of the students were selected to match self-evaluations with the teacher. Reinforcement intervals were extended from 15 to 20 min. Five points for academic work and five points for behavior could still be earned for each 20-min reinforcement interval.

Condition 6: elimination of matching/(16⅔%). Only one of the students was selected to match with the teacher every 20 min. Verbal feedback concerning their self-evaluations continued.

Condition 7: self-evaluation in resource room/30 min (elimination of matching). Students kept the number of points they gave themselves, while the teacher continued to monitor students' ratings and privately record her own ratings for them. Since some students began to give themselves the maximum number of points possible even when their work and behavior did not warrant it, an occasional "surprise" match between the teacher and students was initiated every two or three days. Backup reinforcers continued to be faded by extending the reinforcement intervals from 20 min to 30 min. Thus, self-evaluation took place twice each hour, rather than three times. Five points for academic work and five points for behavior could be earned for each 30-min reinforcement interval.

Procedures: Phase II

Procedures used to achieve generalization and maintenance of the treatment gains were introduced to one pair of students at a time. The pair

of students whose classroom behavior and academic work first reached appropriate and stable levels were introduced to generalization and maintenance conditions first. An average of at least 80% appropriate classroom behavior in the resource room over a 4-day period was required before students were returned on a full-time basis to their regular classrooms. Additional pairs of students were introduced to the Phase II conditions approximately 1 wk apart.

During the week preceding the date that students were projected to return to their regular classrooms on a full-time basis, their teachers received a general, individual orientation from the first author. It should be noted that the first author assumed the duties of the resource teacher at this point, when the resource teacher moved out of the country. If the resource teacher had remained, the program-related duties carried out by the first author would have been completed by the resource teacher.

Teachers were informed of the anticipated date for their students' full-time reentry into their classrooms. Procedures to be used in regular classrooms and expectations for teachers in carrying out the program components were discussed. Teachers were advised that consistency in following specific program instructions was important, particularly in regard to self-evaluation intervals, providing appropriate feedback, and supplying students with information concerning condition changes in the program as they occurred. Teachers were also encouraged to praise students' appropriate behavior frequently and to ignore inappropriate behavior. Classroom rules for each classroom were reviewed and clarified to confirm teachers' expectations.

The 0 to 5 rating scale was also discussed with teachers, and copies of the rating criteria were distributed. Written instructions reviewing the regular classroom teachers' role in implementing each aspect of the self-evaluation program were also provided to teachers prior to each condition change for their reference and review. On the

days that condition changes took place, the first author visited the classrooms to make certain that procedures were being followed as designed and to answer any questions teachers might have regarding program implementation.

An attempt was made to judge the accuracy of teachers' ratings for academic work and behavior at least once each week for all students. Independent ratings by the first author were compared to the teachers' ratings for that day. Discrepancies were discussed with teachers, and decisions regarding future rating of areas of disagreement were made.

Condition 8: self-evaluation in the regular classroom/30 min. Procedures used in this condition were the same as those used in condition 7 in the resource room. The regular classroom teachers asked the students to rate their academic work and behavior every 30 min of a selected hour period during which academic work was performed, using the same self-evaluation cards they had used in the resource room. Students continued their ratings on the same 0 to 5 point scale they had used in the resource room. Feedback on accuracy of self-ratings was given to the students by their regular teachers.

Regular teachers privately recorded their own ratings for students and conducted occasional surprise matches with the students every 2 or 3 days at their own discretion as the resource teacher had done in the later conditions of Phase I. When surprise matches took place, students could either earn bonus points, keep the number of points they had given themselves, or lose all their points, in accordance with the previous matching procedures.

During Phase II, except when surprise matches took place, students kept the number of points they gave themselves. Students exchanged their points once each day by bringing their self-evaluation cards to the resource room after lunch.

Condition 9: fading self-evaluation in the regular classroom/60 min. Students rated themselves once every 60 min in the regular class-

room on their self-evaluation cards during the designated self-evaluation hour. In this way, less reinforcement was earned, and longer work periods were required for earning it. Points could still be exchanged once each day for reinforcement.

Condition 10: points exchange/VR 2 days. This condition was similar to Condition 9, except that students exchanged points on only two randomly selected days each week. The days of the week were written on slips of paper and drawn from a container to determine on which days the students would be allowed to exchange points. Students were informed ahead of time that they would be allowed to keep and exchange their points on an average of every 2 days but that they would not know before each day's ratings were completed whether they would be allowed to exchange points for that day. Surprise matches occurred approximately once each week. Verbal praise and feedback by teachers for accuracy of self-evaluations continued.

Condition 11: no points exchange. Students were asked to continue to self-evaluate their academic work and behavior on their self-evaluation cards but were told they would no longer trade their points for reinforcers. Students were told by their teachers that they no longer needed points to help them. Teacher praise and feedback for accurate ratings continued.

Condition 12: self-evaluation verbally/60 min. Students discontinued marking on their self-evaluation cards. Instead they were asked by the teacher to rate their academic work and behavior verbally at the end of the designated 60-min period using the same 0 to 5 point system they had been using. Teacher praise and feedback on accuracy of ratings continued.

Condition 13: self-evaluation verbally/VR 2 days. This condition was similar to condition 12, except that students were asked to self-evaluate their academic work and behavior verbally on an average of every 2 days.

Condition 14: no verbal self-evaluation. Stu-

dents no longer verbally self-evaluated their academic work and behavior. However, the benefits and possible outcomes of continued "private" self-evaluation was discussed by the teachers with the students, and students were encouraged to continue self-evaluating on their own.

Booster sessions. If appropriate classroom behavior was lower than 80% for three successive school days, booster sessions were given by the first author to prevent further deterioration. Administration of a booster session began by removing the student to a quiet part of his classroom such as the teacher's office. Classroom rules were reviewed and discussed with the student by the first author. The student was asked which rules he had been following well and which ones had given him difficulty. Regular classroom teacher comments were incorporated with notes made by classroom observers to provide the student with external feedback regarding his work and behavior. Modeling and role playing were used to generate and practice alternatives to the problems the student had been experiencing. Finally, the student was asked to verbalize his plans for handling problem situations the next time they occurred. Generally, a booster session lasted 10 to 15 min.

Data and instrumentation. Two primary dependent measures were included in the study: (a) mean percentage of time engaged in appropriate classroom behavior per 10-min observation period and (b) mean percentage of time engaged in inappropriate classroom behavior per 10-min observation period. Table 1 provides a specification of the behavioral definitions used in this study.

During the resource room training program, each student was observed daily for at least one 10-min interval in both his regular classroom and the resource room and in his regular classroom during Phase II. Randomly selected classmates were also observed daily in the regular classrooms during the time of resource room training and throughout the remainder of the

Table 1
Observation Code and Definitions

I. *Appropriate Classroom Behavior*

1. *Attend*—The student is looking at the teacher when the teacher is talking or presenting information to the individual or class, looking at materials in the classroom that have to do with the lesson, or looking at a peer who is presenting related academic information to the class. Attending behavior is characterized by eye contact with (and head and body orientation in the direction of) the appropriate classroom objects (teacher or task).
2. *Work*—The student is engaged in or is completing teacher-assigned tasks. Work responses are characterized by nonverbal, motor movements if a written response is required. If the student is reading, progressive eye movement and page turning are evidenced.
3. *Volunteer*—The student raises his hand to offer information or otherwise offers an appropriate response related to the ongoing academic activity in response to a teacher's question or suggestion.
4. *Reading Aloud*—The student is observed to be orally reading during the reading period or when asked to do so by the teacher. The student can be reading any form of printed material ranging from books, charts, blackboard, or word cards.
5. *Answering Questions*—The student will answer questions when called on to do so by the teacher.
6. *Asking the Teacher a Question*—After raising his hand to gain teacher attention, the student asks the teacher a question when she calls on him.
7. *Other*—The student exhibits appropriate classroom behavior as determined by classroom rules in operation in the classroom.

II. *Inappropriate Behavior*

1. *Talking Out*—The student speaks without permission or interrupts the teacher and another student who are talking to each other.
2. *Out-of-Chair*—Movement of the student from his chair when not permitted. Such movement may include leaving the chair to open the window, remove items or threaten to remove items from the teacher's or other students' desks, name-calling, and moving around the room.
3. *Modified Out-of-Chair*—Movement of the student from his chair with some part of the body still touching the chair (excluding sitting on feet).
4. *Noise*—The student creates any audible noise other than vocalization.
5. *Rocking*—The student lifts one or more of his chair legs from the floor while he is seated in his chair.
6. *Noncompliance*—Failure by the student to initiate the appropriate response as requested by the teacher.
7. *Aggression*—The student makes movement toward another person so as to come into contact with him, whether directly or by using a material object as an extension of the hand.
8. *Other*—The student clearly violates school or classroom rules or engages in behavior which prevents him from engaging in learning tasks and which are not otherwise specifically defined. Such behavior must be determined by the rules in operation in students' classrooms. Examples of such behavior may include engaging singly in activities or tasks not approved by the teacher or related to the assigned academic tasks (i.e., combing hair, writing on desk, looking at or handling objects within the immediate area surrounding the student's desk or work area, not appropriate to the academic task at hand).

study. Thus, mean percentages of the two categories of coded behavior for each condition of Phase I and Phase II were calculated for students and their randomly selected classmates, based on observations in the regular and resource rooms.

Observations were made for one child at a time on a continuous 10-sec observe-and-record schedule. During each 10-sec interval, a student's behavior was coded with an "I" for inappropriate in the designated row on the observation form for any 10-sec interval in which one or more of the behaviors specified as inappropriate

in the observation code was exhibited. An "A" denoted any 10-sec interval in which appropriate classroom behavior was demonstrated for the entire 10-sec interval. Percentages of the two categories of coded behavior for each 10-min observation period were calculated by dividing the number of 10-sec observation intervals coded with an "I" for inappropriate behavior or an "A" for appropriate classroom behavior by the total number of observation intervals and multiplying the obtained number by 100 for that 10-min observation period.

Observers were given a list of target students and the times at which they were to be observed for one week at a time. When a second observer was present in the regular classroom, observers were not aware of which students (target students or their classmates) other observers were observing. Observers were asked to calibrate their stopwatches at the beginning of observation intervals whenever another observer was present and to avoid any form of communication with each other in regard to their observations. When two observers were present at the same time, they might or might not have been observing the same student. This method was used to prevent "reactivity of reliability assessment" (Kazdin, 1977). The list of students with which observers were provided weekly was constructed using random selection without replacement to determine which student was to be observed during each 10-min interval.

Observer agreement checks were made on an average of at least four times during each of the 14 conditions of baseline, Phase I, and Phase II for one 10-min observation interval. Agreement was calculated by dividing the number of perfect agreements on the occurrences of coded behavior categories by the total number of agreements plus disagreements for two observers, times 100.

Observer selection and training. Six weeks before the first phase of the study was scheduled to begin, five observers were selected for training from those who responded to posted notices on the Utah State University campus. Observers were selected from the pool of respondents based on their availability for making observations during scheduled hours as needed and their understanding that if hired they were responsible for making scheduled observations or arranging for one of the other observers to assume their responsibilities in the event that an unanticipated absence occurred. Observers were paid minimum wage for the hours they spent in training and observation. All the observers were undergraduate students, and only one had previous observation experience.

Observer training began with discussion and

explanation of the observation recording form and observation code to be used in the study. Modeling and role playing of examples and nonexamples of behaviors included in the observation code and of situations similar to those expected in the study were demonstrated for observer trainees, and the way in which behaviors were to be coded was explained in detail. Variations in classroom rules for the resource room and the regular classrooms were also explained and then modeled and role played. Observer trainees practiced recording behavior in the regular and resource classrooms and were given feedback and clarification as needed by the first author until reliability was consistently attained at acceptable levels. Observers were required to achieve at least 85% agreement with reliability checkers on three successive trials to participate in the study.

To help control for observer drift, the first author conducted reliability checks with each observer at least once each week during their scheduled observations. The observation code was discussed with observers and revised as unanticipated and unusual behaviors occurred throughout the study, so that the group would continue to observe and record consistently. Observers were encouraged to bring up any questions they had concerning their observations with the first author. To minimize the presence of observer bias, observers were not informed of the study purpose until the research was completed.

Experimental design. A multiple baseline across pairs of subjects design was used to examine the effects of using the self-evaluation procedure in obtaining generalization of appropriate classroom behavior from the resource room to regular classroom settings. The students' behavior in the regular classroom during the initial baseline and throughout the resource room training in Phase I served as the baseline period against which Phase II performance in the regular classroom was evaluated. When stable levels of appropriate behavior over time did not occur in the regular classroom during Phase

I, a significantly less intense form of the initial self-evaluation procedure was introduced into the regular classroom to one pair of students at a time. This method of programming for generalization is consistent with the suggestion made by Stokes and Baer (1977) that generalization may be claimed if the extent of the intervention in the generalization setting (i.e., regular classrooms) is clearly less than that of the initial intervention in the treatment setting (i.e., the resource room).

RESULTS

The students were introduced to Phase II in pairs: Subjects 1 and 2, 3 and 4, 5 and 6. To facilitate inspection of data for the students across study conditions, the data for Subjects 1, 3, and 5 are presented in Figure 1; the data for Subjects 2, 4, and 6 are presented in Figure 2.

Appropriate behavior for baseline was highly variable and generally low for all students in both settings except for that of Subject 6. Appropriate behavior for Subject 6 exhibited an upward trend in the resource room for the last four days of baseline. The resource room data show a quick and dramatic improvement for all students with the introduction of external reinforcement and feedback. The high level of performance in the resource room maintained throughout the six conditions of self-evaluation training. The variability of the data was minimal. Following baseline, Subjects 1, 2, and 4 averaged 92% appropriate behavior across all six Phase I conditions. Subject 3 averaged 96%, Subject 5 averaged 91%, and Subject 6 averaged 90% appropriate behavior in the resource room. Conversely, during the same time in the regular classroom, the mean percentage of appropriate behavior for each student was as follows: Subject 1 (39%), Subject 2 (31%), Subject 3 (34%), Subject 4 (58%), Subject 5 (28%), and Subject 6 (39%). These data indicate that the self-evaluation training program did improve the behavior of all students while in the resource room. During Phase I, the per-

formance of all students in the regular classroom was highly variable. During the last few days of resource room training, the data for Subjects 3 and 5 show an upward trend indicating some possible spontaneous generalization in the regular classroom.

The data for Phase II, self-evaluation in the regular classroom, indicated an immediate shift to high, stable performance for 5 of the 6 students. Subject 6 showed improvement although the data continued to have a wide range of variability. Subjects 1, 3, 4, and 5 maintained high levels of appropriate classroom behavior in their regular classrooms throughout Phase II, even after the teacher involvement and formal self-evaluation treatment had been faded. For these four students, the average percentage of appropriate behavior in the regular classroom during Phase II (93%) was 54% higher than their average percentage of appropriate behavior in the regular classrooms during Phase I.

Modifications in the treatment program were necessary for Subjects 2 and 6 in order to produce acceptable levels of appropriate classroom behaviors in their regular classrooms. Subject 2 received six booster sessions during Phase II and as noted in Figure 2, he was eventually recycled to an earlier Phase II treatment condition (self-evaluate/60 min) since he had initially been successful during that condition. The generalization treatment program for Subject 6 was modified by recycling him to the third Phase I treatment condition (match teacher evaluation/100%). This decision was made following the unstable pattern of behavior during the first Phase II treatment condition (self-evaluate/30 min). Subject 6 continued to match his regular teacher's ratings every 15 min throughout the duration of the study. In addition, Subject 6 received three booster sessions during Phase II. For Phase II as a whole, Subject 6 averaged 78% appropriate behavior in his regular classroom, an increase of 39% over his performance during Phase I.

Matching accuracy. A match between student self-evaluations and teacher evaluations for aca-

Fig. 1. The percentage of appropriate behavior for Subjects 1, 3, and 5 in the resource room and the regular classroom.

demic work and behavior was scored if student self-evaluations exactly matched teacher evaluations or were within one point higher or lower

than teacher evaluations. Based on these criteria, student self-evaluations matched teacher evaluations 93% of the time during Phase I in the

Fig. 2. The percentage of appropriate behavior for Subjects 2, 4, and 6 in the resource room and the regular classroom.

resource room. Average percentages across students for matching during Phase I conditions ranged from 82 to 100%.

The goal during the generalization and maintenance phase of treatment was to fade remaining evaluation (written and verbal) gradually until teacher intervention was minimal and students were maintaining high levels of appropriate classroom behavior. During Phase II, an actual match between teacher evaluations and student self-evaluations took place only occasionally at the discretion of the regular teachers. When planned or "surprise" matches were not conducted, teachers continued to rate and record their evaluations of academic student work and classroom behavior privately.

During Phase II, student self-evaluations matched teacher evaluations 88% of the time and ranged from 79 to 97% across the different conditions. Although the 88% matching for Phase II represented an 8% decrease from the 96% matching present during Phase I, it still represents an acceptable level of performance. It is interesting to note that during Phase II, the two students who required program modifications (Subjects 2 and 6) evaluated their academic work and classroom behavior less accurately than any of the other students.

Social validation. Wolf (1978) stated that experimental procedures need to be socially validated in terms of significance of the goals, appropriateness of the procedures, and the importance of the effects. Van Houten (1979) proposed the inclusion of the social validation of standards of competency for target behaviors. In the present study, social validation took two forms: (a) evaluation of behavior change within the framework of normative peer data (standards of competency) and (b) feedback from individuals who had an interest or involvement in the experimental program (appropriateness and importance of the procedures).

Normative data were collected on the students' regular classroom peers for a comparison to determine the social importance of the effects of this study. On an average of at least once each

day that students were observed, observations were also made on randomly selected classmates in the regular classrooms. The same observation codes and procedures used to observe the target students were used to obtain data for randomly selected classmates' appropriate and inappropriate classroom behavior.

The randomly selected classmates averaged 86% appropriate behavior during the initial baseline period and 90% appropriate behavior during the subsequent 42 days of Phase I. Their average appropriate behavior for Phase II was 94%. Average appropriate behavior for the peers was similar across Phases I and II, indicating that their high level of appropriate behavior was present throughout the study.

Over the 42 days and six treatment conditions of Phase I which followed baseline, the target students' average appropriate behavior in their regular classrooms was 38%, 52% less than their classmates. However, in the resource room the students averaged 92% appropriate behavior which was comparable to the behavior of their classmates in the regular classrooms during the same period of time. During Phase II, students averaged 92% appropriate behavior in their regular classrooms while their classmates averaged 94% appropriate. Figure 3 depicts the mean percentage of appropriate behavior for the target students and their peers throughout all conditions of Phases I and II.

The data in Figure 3 illustrate the effects of the self-evaluation procedures for the group as a whole. Appropriate behavior for the target students in their regular classrooms in Phase I was much lower than that of their classmates and of their own behavior in the resource room. The mean performance for the students during the last two conditions of Phase I in the regular classrooms shows an upward trend. However, the range of scores was still highly variable during that time. After Phase II treatment was introduced, the mean performance of the target students was comparable to that of their classmates. Each condition of Phase II resulted in much less variability in appropriate behavior,

Fig. 3. Mean percentages and ranges of appropriate behavior for the group of students. Normative data in the form of mean percentages of appropriate behavior for randomly selected peers are also included.

with the exception of performance for Subject 2 on day 79, when he exhibited no appropriate behavior.

The social appropriateness of the procedures and the significance of the goals were assessed by requesting that regular classroom teachers, parents, and the students themselves respond to a brief questionnaire. Teachers and parents both

expressed high levels of satisfaction with the program and indicated that they would support the use of the program again under similar circumstances. Teachers viewed the program as relatively easy to implement while carrying out their normal classroom duties. The students also said they liked the program and enjoyed trading the points they earned for treats and toys. In

general, the students thought that program procedures resulted in improved work and behavior in their regular classrooms and that their teachers said nicer things about their work and behavior after they had participated in the program.

DISCUSSION

The data in Figures 1 and 2 clearly show that student behavior improved after self-evaluation procedures were taught in the resource room and that improved behavior generalized to and was maintained in the regular classrooms using a significantly less intense version of the original self-evaluation procedure. As a group, students' average appropriate behavior at the conclusion of the study was 54% higher in the regular classroom than it has been during Phase I. For the four students who progressed through the program sequence as planned, the average percentage of appropriate behavior (97%) exhibited after all forms of intervention were withdrawn was 63% higher than before they had participated in the program. For Subjects 2 and 6, who required some form of intervention to maintain treatment gains, appropriate behavior in the regular classroom was still considerably higher (39% and 51%) than it had been before participation in the program.

Furthermore, the observed improvements in appropriate classroom behavior for the target students reached levels essentially equivalent to that of their randomly selected classmates. For the four students who progressed as planned through the program and maintained their improved behavior after all forms of intervention were withdrawn, the average percentage of appropriate behavior exhibited in the last study condition was 97% compared to 96% for randomly selected classmates. For one of the two students who required a reduced form of the initial intervention to maintain behavior gains, appropriate behavior was maintained above the study criterion of 80%. The remaining student averaged 73% appropriate behavior. These per-

centages were still considerably higher for the two students than their level of appropriate behavior prior to treatment.

Equally as important as increases in appropriate behavior is the restriction of the variability of students' appropriate behavior. Demonstration of a functional relationship between an independent variable and changes in behavioral variability, according to Johnston and Penny-packer (1980) should not be viewed as an "analytical inconvenience," but rather as a desired occurrence.

During Phase I, all six students showed wide variability in their appropriate regular classroom behavior, a common characteristic among behaviorally handicapped students (Walker, 1979). After the introduction of the self-evaluation procedure into the regular classroom in Phase II, the range of appropriate behavior percentages was much more restricted for five of the students. Thus, the self-evaluation procedure was responsible for reducing the variability of student behavior in their regular classrooms as well as increasing levels of appropriate behavior.

In this study, treatment gains generalized from a special education training setting to regular classroom settings through the use of a modified, less intense form of the original intervention. Although some extratraining manipulations were used to promote the transfer of behavior gains to the regular classrooms, their cost in terms of teacher time and effort were clearly less than that of the initial Phase I intervention. During the Phase I treatment, students were taught to manage their own behavior through the use of self-evaluation procedures. Teaching students to use self-evaluation procedures in the resource room initially called for a substantial number of teacher requirements and amount of teacher time. It should be noted that implementing Phase I procedures in the regular classroom was an unacceptable and impractical alternative due to the extreme behavior of the students and the heavy response requirements that would have been imposed on the

regular classroom teacher. For example, near the beginning of resource room treatment, the resource teacher asked students to self-evaluate every 15 min, privately rated each of the six and then matched evaluations with and gave feedback to all six.

However, the ultimate goal of Phase I procedures was to reduce teacher requirements to proportions manageable by a regular classroom teacher while still maintaining high levels of appropriate classroom behavior by the students. By the end of resource room treatment, the resource teacher asked students to self-evaluate every 30 min, privately recorded her own ratings for them, but matched evaluations with only one student every 2 or 3 days. At this time, the reduced treatment requirements were introduced into students' regular classrooms. Thus, it was by clearly and substantially reducing the program requirements for maintaining treatment gains in the regular classrooms that generalization can be claimed (Stokes & Baer, 1977).

During the last condition of Phase I, the appropriate behavior of Subjects 3 and 5 increased to levels above the 80% criterion. These sudden increases were not anticipated, since spontaneous generalization of results to the regular classroom had not been achieved in the study by Turkewitz *et al.* (1975) and does not typically occur in education settings (Walker, 1979). Some plausible explanations of this spontaneous generalization should be considered. One possibility might be that the result of the treatment produced new appropriate responses in the regular classroom that contacted a natural community of reinforcement (*i.e.*, the teacher) which further strengthened the behavior (Baer, 1981). Since the schedule of reinforcement was thinned to VR 2 days, the appropriate classroom behavior might then have only required minimal reinforcement from the regular classroom teacher in order to show improvement. Another possible explanation is that the length of Phase I (approximately 40 days of training) provided sufficient practice to help the

behavior become fluent. White and Haring (1980) and Baer (1981) suggested that fluency facilitates the maintenance and generalization of behavior because it increases the possibility of the subject contacting natural communities of reinforcement.

In the future, a design in which Phase I conditions are introduced to subjects using multiple baseline across procedures may help investigate the occurrence of unprogrammed generalization. Such research is needed to understand the process and effects of self-evaluation procedures further.

An additional area for future investigation relates to the behavior of the regular classroom teacher. Although the present research did not objectively investigate changes in regular classroom teacher behavior, teachers were subjectively monitored through informal observations by the first author. In this way, an attempt was made to obtain information regarding individual variations in teacher behavior across all experimental conditions. These informal observations suggested that teachers were delivering relatively low rates of positive reinforcement prior to and during the time self-evaluation procedures were in effect in the regular classrooms. The authors doubt that adventitious contingencies regarding teacher behavior were related to data obtained in this study, but related future research should document teacher behavior across conditions to provide a more precise demonstration of the effects of self-evaluation.

Several other findings contribute to the self-evaluation literature. For example, previous research has suggested that accuracy in self-evaluation may not be a critical factor in obtaining desired behavior changes and that the experience of using self-evaluation in itself may serve a useful function in facilitating the maintenance of treatment gains when all intervention has been terminated (O'Leary & Dubey, 1979; Rosenbaum & Drabman, 1979). In the present study, however, maintenance of higher levels of appropriate behavior seemed to be associated

with accuracy of self-evaluation ratings. That is, students who evaluated their academic work and behavior accurately (Subjects 1, 3, 4, and 5) maintained higher levels of appropriate classroom behavior than students (Subjects 2 and 6) who evaluated their work and behavior less accurately. Thus, the findings of the present study lend support to the suggestion that matching accuracy was a critical component of the self-evaluation procedures. A fruitful area of future investigation would involve a manipulation of matching accuracy in an effort to discover the strength of its relationship to improved behavior.

The present study also suggests that handicapped students as young as 6-yr-old can accurately self-evaluate and that behavior changes for young children can be generalized across time and settings and be maintained using a self-evaluation procedure. In the present study, the three youngest students (two 6-yr-olds and one 7-yr-old) had the highest average percentage of appropriate behavior for the generalization and maintenance phase of the study (Phase II).

Finally, it should be noted that the student whose behavior throughout the study was the most variable and least satisfactory (Subject 6), was also the student whose prior history and level of performance at the outset of the study were the most deviant. A more careful examination of the effectiveness of self-evaluation as a function of severity of the behavior problem should also be considered in future related studies.

REFERENCES

- Asher, S. R., & Taylor, A. R. Social outcomes of mainstreaming: Sociometric assessment and beyond. *Exceptional Education Quarterly*, 1981, 1, 13-30.
- Baer, D. M. *How to plan for generalization*. Lawrence, KS: H & H Enterprises, Inc., 1981.
- Drabman, R. S., Spitalnik, R., & O'leary, K. D. Teaching self-control to disruptive children. *Journal of Abnormal Psychology*, 1973, 82, 10-16.
- Gottlieb, J., & Leyser, Y. Facilitating the social mainstreaming of retarded children. *Exceptional Education Quarterly*, 1981, 1, 57-69.
- Johnson, S. M. Self-reinforcement versus external reinforcement in behavior modification with children. *Developmental Psychology*, 1970, 3, 147-148.
- Johnston, J. M., & Pennypacker, H. S. *Strategies and tactics of human behavioral research*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.
- Kanfer, F. H., & Duerfeldt, P. H. Effects on retention of externally or self-reinforced rehearsal trials following acquisition. *Psychological Reports*, 1967, 21, 194-196.
- Kazdin, A. E. Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis*, 1977, 10, 141-150.
- McLaughlin, T. F. Self-control in the classroom. *Review of Educational Research*, 1976, 46, 631-663.
- O'Leary, S. G., & Dubey, D. R. Applications of self-control procedures by children: A review. *Journal of Applied Behavior Analysis*, 1979, 12, 449-465.
- Reynolds, M. C. Interpreting public law 94-142. In S. Sunderlin (Ed.), *The most enabling environment: Education is for all children*. Washington, D.C.: Association for Childhood Education International, 1979.
- Robertson, S. J., Simon, S. J., Pachman, J. S., & Drabman, R. S. Self-control and generalization procedures in a classroom of disruptive retarded children. *Child Behavior Therapy*, 1979, 4, 347-362.
- Rosenbaum, M. S., & Drabman, R. S. Self-control training in the classroom: A review and critique. *Journal of Applied Behavioral Analysis*, 1979, 12, 467-485.
- Stokes, T. F., & Baer, D. M. An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 1977, 10, 349-367.
- Turkewitz, J., O'Leary, K. D., & Ironsmith, M. Generalization and maintenance of appropriate behavior through self-control. *Journal of Counseling and Clinical Psychology*, 1975, 43, 577-583.
- Van Houten, R. Social validation: The evaluation of standards of competency for target behaviors. *Journal of Applied Behavior Analysis*, 1979, 12, 581-591.
- Walker, H. J., & Buckley, N. K. Programming generalization and maintenance of treatment effects across time and across settings. *Journal of Applied Behavior Analysis*, 1972, 5, 209-224.
- Walker, H. M. *The acting-out child: Coping with classroom disruption*. Boston, London, & Sidney: Allyn & Bacon, 1979.
- White, R. R., & Haring, N. G. *Exceptional teaching* (2nd ed.). Columbus: Charles E. Merrill, 1980.
- Wildman, R. W., & Wildman, R. W. The generalization of behavior modification procedures: A review—with special emphasis on classroom ap-

- plication. *Psychology in the Schools*, 1975, **12**, 432-448.
- Wolf, M. M. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 1978, **11**, 203-214.
- Workman, E. A., & Hector, M. A. Behavioral self-control in classroom settings: A review of the literature. *Journal of School Psychology*, 1978, **16**, 227-236.

Received December 21, 1981

Final acceptance December 21, 1982