JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1982, 15, 123-139 NUMBER 1 (SPRING 1982)

COLLATERAL GAINS AND SHORT-TERM MAINTENANCE IN READING AND ON-TASK RESPONSES BY INNER-CITY ADOLESCENTS AS A FUNCTION OF THEIR USE OF SOCIAL REINFORCEMENT WHILE TUTORING

R. DOUGLAS GREER AND SUSAN ROVET POLIRSTOK

TEACHERS COLLEGE, COLUMBIA UNIVERSITY
AND HERBERT LEHMAN COLLEGE

Two experiments are reported concerning the effects of the differential use of verbal approval by problematic adolescents serving as tutors in a remedial reading program for an inner-city school. The experiments, each with 3 tutors and 15 tutees, used a combined multiple baseline and ABCBC design. Data showed that tutors' approvals as well as tutors' and tutees' on-task and reading responses were low and stable during baseline. Tutors were trained to use verbal approval for tutees' on-task behavior. Tokens were presented and withdrawn to control the tutors' use of approval. During phases in which tutors' approvals were raised via token dispensation, tutor reading and on-task scores increased in a nonexperimental setting. Tutee reading scores also increased as a function of tutor approvals. The second experiment replicated these findings and, in addition, (a) tested the validity of changes in reading responses via standardized tests, (b) isolated and compared the covariance between variables in all phases, and (c) provided data on tutee attention to tutors as a possible natural reinforcer for the short-term maintenance found in both studies. Data are discussed as evidence that tutors had acquired the ability to recruit reinforcement from the classroom for appropriate behavior.

DESCRIPTORS: collateral behaviors, tutoring, maintenance, social reinforcement, adolescents

Society continues to be concerned about the low reading scores of students in inner-city schools, particularly in those neighborhoods besieged by crime, poverty, and illiteracy. One way to increase reading skills and maximize individualization while minimizing costs is to have students tutor each other (Harris & Sherman, 1973; Keller, 1968; Robertson, DeReus, & Drabman, 1976). Such a procedure has the added potential advantage of utilizing peer reinforcement which is part of the natural community of reinforcers in a classroom.

Peer and cross-age tutoring have been two of the most consistently effective procedures reported both in the experimental analysis literature (McGee, Kaufman, & Nussen, 1977) and in the educational research literature (Devin-Sheehan, Feldman, & Allen, 1976). A few studies have assessed and found effects for the tutors in addition to the usual effects found for tutees (Dineen, Clark, & Risley, 1977). Remedial students have been used to tutor other remedial students in at least one case (Davis, 1972). One of the unresolved issues concerns the extent of the effect for the tutor. A second issue concerns what specific tutor behaviors are most effective in changing both tutor and tutee behaviors.

Cloward (1967) proposed that tutors might benefit even more than tutees, and Greer and Polirstok (Note 1) suggested that serving as a tutor might be an optimum procedure to use with the most problematic and oppositional young adolescents. Unfortunately, many of these students do not have the necessary reading skills to tutor even younger remedial reading students. However, teaching them to use social reinforcement techniques for tutee on-task be-

Reprints can be obtained from R. Douglas Greer, Departments of Psychology and Special Education, Box 76, Columbia University's Teachers College, 525 West 120th Street, New York, New York 10027.

havior might have the potential for developing new social behaviors for tutors. These new social behaviors might then result in the recruitment of reinforcers in the natural community for appropriate rather than inappropriate behaviors (Stokes, Fowler, & Baer, 1979).

Recently there has been considerable interest in generalization (Stokes & Baer, 1977; Johnston, 1979). However, generalization may be a subcomponent of an even more generic question associated with interventions. Generalization, strictly defined, may be only one type of a multiple effect which might occur as a result of treatment. Another part of this multiple effect involves the influence of a given treatment on collateral behaviors (Hersen & Barlow, 1976). Collateral behaviors may be defined as those behaviors that are topographically dissimilar to the behaviors treated. They have not been classified as generalization traditionally, yet they may be changed as a result of a treatment for a separate target behavior. Drabman, Hammer, and Rosenbaum (1979) have proposed a generalization taxonomy that would include and subdivide collateral behaviors into classes of generalization effects. Regardless of how collateral behaviors are classified, behavior analysts have noted the importance of monitoring them (Drabman et al., 1979; Kazdin, 1973; Lovaas & Simmons, 1969; Risley, 1968; Sajwaj, Twardosz, & Burke, 1972).

Two experiments are reported in which adolescent students in an inner-city remedial reading program were trained to use social reinforcement while tutoring. The tutors' rates of differential social reinforcement for their tutees' on-task behaviors were controlled by a token economy in order to test the effects of tutors' use of reinforcement on the collateral behaviors of the tutors themselves.

EXPERIMENT 1

METHOD

Participants and Settings

Tutors. The tutors were three ninth-grade males (ages 14, 15, and 16) from an urban junior high school who had experienced the academic and discipline problems usually associated with delinquent or predelinquent populations. Each tutor (Tutors A, B, and C) had been given two 5-day suspensions from school for breaking school rules during the preceding semester. The tutors were consistently off task in their remedial reading course and usually did not attempt daily assignments. Tutor A's Stanford Achievement Test Score (SAT) in reading was 7.3, while that of Tutor B was 7.8, and that of Tutor C was 3.9.

Tutees. Fifteen tutees were selected randomly from eighth graders assigned to the Reading Resource Center, a class in a program devoted to remedial reading instruction. Five students were randomly assigned to each tutor. The mean SAT reading score for all 15 tutees was 5.8.

Settings. The experiment took place in an urban school in which 80% of the students read at least one year below grade level. Over 35% of the student body spoke Spanish in their homes. The eighth-grade tutee class met during a separate time period and in a different classroom from that of the remedial reading class for the ninth-grade tutors. In tutoring phases, tutors met with their eighth-grade tutees' reading class during what was formerly the tutors' study period. The collateral behaviors of tutors were monitored in the tutors' remedial reading class in which none of the special conditions of the tutoring setting was in effect. All tutors and tutees were together only in the tutoring setting. Each tutor worked with his group of five tutees at one table. Tutors circulated among their tutees dispensing approval statements to tutees who were engaged in on-task behaviors.

Response Definitions and Data Collection Procedures

Data were collected for (a) the daily reading responses of tutees in the tutoring setting, (b) tutors' approvals of tutees, and (c) tutors' on-task and reading responses in their own re-

medial reading class in which no experimental conditions were in effect.

Daily reading responses. Tutor and tutee daily reading responses consisted of the number of correctly written responses to reading comprehension items. Reading items were drawn from the Science Research Associates' (SRA) Reading Kit (Thurstone, 1963) titled "Reading for Understanding-Junior" (RFU). Students were assigned specific reading items based on their SAT reading scores and their respective scores on an SRA Placement Test (Staats & Butterfield, 1965). A "self-check" answer key was available for the tutees to use under teacher supervision.

Reading task assignment and data collection for the students consisted of the following procedures: (a) 20 reading items for each student's level of ability were assigned daily; (b) students progressed to the next level of difficulty when they achieved 18 correct responses out of 20 items (90% criterion); (c) when the students scored less than 18 correct, they continued to receive items at the same ability level; (d) each student checked his own answers while being monitored by the teacher who then logged the data; and (e) all incorrect items were reviewed by the teacher after the period of time allotted for student responses.

Tutor on-task behavior. On-task behaviors consisted of following the class rules. The class rules in both the tutors' class (nontutoring setting) and the tutees' class (tutoring setting) included: (a) Sit in your seat. (b) Do not talk louder than a whisper and then only about your work. (c) Raise your hand to receive help. (d) Use your dictionary to check word meanings and enter the words and their meanings in your own vocabulary log book. (e) If you finish all 20 items, read a book from the class library.

Observations were conducted on a 20-sec observe/10-sec record basis. In order to be recorded as having an on-task interval, a student had to follow the above rules throughout an entire 20-sec observation period. During each 10-sec record interval, it was noted whether the student had been on or off task. Any occurrence of an off-task behavior during a given observational interval resulted in a score of off task for the student observed for that particular interval.

On-task data were collected for tutors in their reading class in which no experimental conditions were in effect. Observe and record cues were sounded through earplugs, begun 7 min after the tardy bell, and continued for 30 min of the 45-min period. The primary and reliability observers first monitored Tutor A for a 20-sec interval, then recorded their observations during a 10-sec record interval; Tutor B was observed for the next 20-sec interval, and then C was observed. Observers then returned to observing Tutor A and in this manner rotated the tutors monitored throughout the 30-min observation period. No data were collected for tutee on-task behavior, although the criteria used for approval of tutee on-task behavior by tutors were the same as used for defining tutor on-task responses.

Tutor social reinforcement of tutees. Tutor approvals consisted of positive evaluative verbalizations given to tutees during sampled observation intervals in the tutoring classroom. Examples of specific approval statements are delineated in Polirstok and Greer (1977).

Tutor approvals were observed in the tutoring setting using the same 20-sec observe/10-sec record observation procedure used for observing on-task behavior in the nontutoring setting. However, only one observer monitored each tutor. Tape-recorded observe cues (beep) and record cues (bell) were audible to all class members. Beep and bell cues used for observation and record signals were begun 2 wk prior to the experiment to habituate students in the experimental setting. Observers were rotated across each tutee-tutor group from day to day. Approval data consisted of the number of approvals recorded during 20-sec observation intervals for each tutor during the 30-min periods. A tape recorder was placed at the table where each of the tutor-tutee groups was located. Tape

recordings of tutor-tutee interactions were made to check the reliability of tutor approval observations. The tape recorder was placed in each tutee group 2 wk prior to the experiment in order to habituate class members to the recorder's presence.

Experimental Design

There were five phases in this ABCBC design. During the untrained tutoring phase (A), tutors were assigned to help five tutees to follow classroom rules. The tutors were given no special instructions in order to determine their natural use of social reinforcement. Data taken on the number of approval statements made by tutors showed that approvals were few in number. Therefore, five training sessions were devoted to teaching tutors individually to use approval statements and to ignore the off-task behavior of their tutees. Training sessions occurred after phase A and before the B phase. The training procedures are described in detail in Polirstok and Greer (1977). After the tutors had learned to use high rates of contingent approval, the second phase of the experiment was begun.

During the second phase, B, (tokens for contingent approvals), tutors were given a token in the form of a point for each contingent approval that they gave to a tutee. The tokens were exchangeable for a variety of items: (a) additional shop classes, (b) certificates for food from a fast-food chain, (c) opportunities to use films and audio cassettes from *The Classics Comics Library,* and (d) a letter to the tutors' parents praising the tutors' work.

During the third phase (C), tokens for tutors' approval statements were withdrawn. This phase was termed a tutoring alone condition.

During the fourth phase (B), tokens were once again given to tutors for their approval statements to tutees. This phase was a reinstatement of the initial B condition. After the last session in this phase, tokens were withdrawn once again.

Two weeks after the conclusion of the final token phase, follow-up observations were made to test for maintenance effects. The conditions in effect throughout the 2-wk hiatus on data collection and throughout the follow-up observations were those of tutoring alone (C).

In addition to the within-subjects feature of the design, the study simultaneously incorporated a between-subjects control in the form of a multiple baseline. This aspect of the design is shown in Figures 1-3.

Reliability

Reliability observations were made twice weekly in the nontutoring setting (tutor on-task behaviors). The three observers were paraprofessionals who typically assisted in the Reading Resource Center; thus, the paraprofessionals served as relatively unobtrusive observers. Each observer also served as a data observer for approval statements made by the tutor in the tutoring setting. Reliability observers and the data observer were trained to use the observational procedures described earlier.

For "approvals" data, reliability observations were obtained from the daily tape recordings of each of the three tutorial groups. Two tapes each week were randomly selected for reliability checks. Interobserver reliability (observer agreement) was computed by dividing the total number of paired observer agreements by the total number of agreements plus disagreements only for those intervals in which a behavior was recorded by one observer. Reliability for approvals data (tutor setting) ranged from 81% to 94% with a mean of 88%. Reliability for on-task data (nonexperimental setting) ranged from 76% to 95% with a mean of 90%.

RESULTS

Tutor Approval

The three major dependent variables for the study were tutee reading scores, tutor on-task behavior, and tutor reading scores. However, data on the amount of approvals given by tutors were also taken in all phases except during the first tutoring alone phase. The approval data

showed that when tutors were assigned without special training, their administration of contingent approvals was low and stable. During the untrained tutoring baseline, the mean number of approvals emitted per session was 1.33 for Tutor A, 1.21 for Tutor B, and 1.6 for Tutor C. During the first administration of tokens for tutor approvals, the per session mean approvals were 13.52 for Tutor A, 8.8 for Tutor B, and 7 for Tutor C. During the second administration of tokens for tutor approvals, the means were 16.8 for Tutor A, 14.6 for Tutor

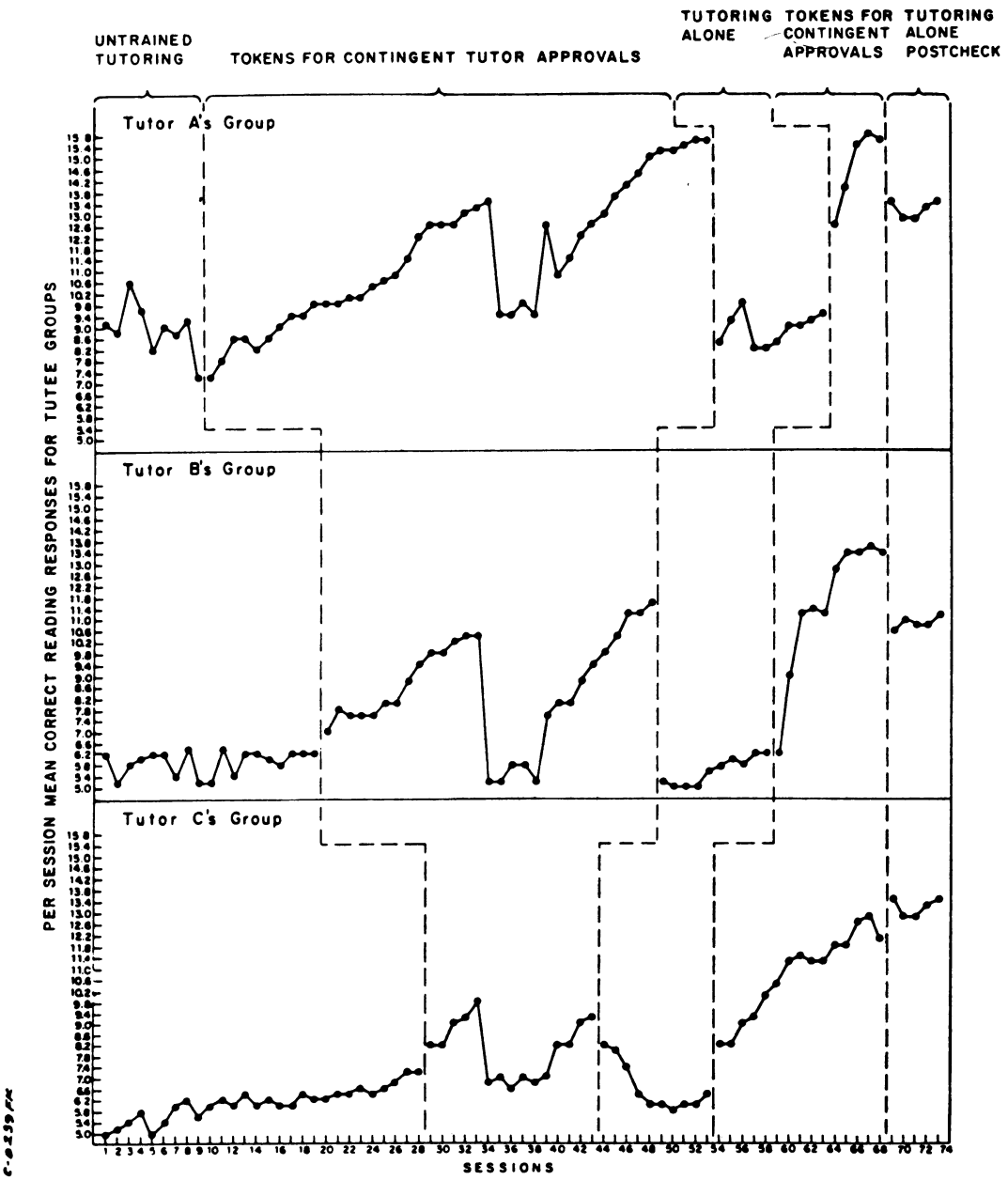

Fig. 1. Per session mean correct reading responses for tutee groups in Experiment 1. Data points for each tutee group session are the total number of correct responses divided by the number of tutees in the tutee group.

B, and 11.2 for Tutor C. For the follow-up (tutoring alone) phase, the means were 15.2 for Tutor A, 11.8 for Tutor B, and 14 for Tutor C. Approval rates were low and stable during the baseline, rose substantially during the phases with tokens for tutor approval statements, and were maintained during the follow-up phase.

Tutee Daily Reading Scores

The effects of the tutors' social reinforcement on the reading means of the tutees are shown in Figure 1. Means are given in the interest of brevity and because the means were representative of individual scores (Hersen & Barlow, 1976; Note 2). During the tokens for approvals phase (B), the groups of Tutor A, Tutor B, and Tutor C increased their daily scores over baseline levels. Scores for all of the groups were comparable to their untrained tutoring baseline during sessions 34-39. Session 34 was the first day after the spring holidays, a factor that apparently affected all the data (see Figures 2 and 3). The untrained tutoring phase means were 8.96 for Tutor A's group, 5.91 for Tutor B's group, and 6.1 for Tutor C's group. The means for the first tokens for contingent tutor approvals increased to 11.61 for Tutor A's group, 8.47 for Tutor B's group, and 7.85 for Tutor C's group. The means during the tutoring alone phase declined to 8.8 for Tutor A's group, 5.6 for Tutor B's group, and 6.6 for Tutor C's group. For the final administration of tokens for tutor approvals, the means again increased to 14.7 for Tutor A's group, 11.6 for Tutor B's group, and 10.8 for Tutor C's group. The means for the follow-up phase were 13.1 for Tutor A's group, 10.9 for Tutor B's group, and 11.9 for Tutor C's group.

Tutors' On-Task Behavior in a Separate Reading Class

The incidences of the on-task behavior of the three tutors as it occurred in the nonexperimental setting during all phases are shown in Figure 2. These data represent a test of the influence on the tutor of experimental conditions in effect only during the tutees' classes. Baseline (A) data show a low incidence of on-task behavior for Tutors A and B. Tutor C demonstrated a higher incidence of on-task behavior during baseline than did Tutors A and B.

The mean number of intervals that each tutor was on task per session was calculated for each phase. During the untrained tutoring baseline, the means were .88 for Tutor A, .21 for Tutor B, and 8.13 for Tutor C. For the first tokens for tutor approvals phase, the means increased to 14.65 for Tutor A, 12.18 for Tutor B, and 13.26 for Tutor C. For the tutoring alone phase, the means declined to 10.7 for Tutor A, 8.2 for Tutor B, and 8 for Tutor C. The means for the second administration of tokens for tutor approvals again increased to 15.8 for Tutor A, 14.1 for Tutor B, and 15.86 for Tutor C. The means for the follow-up (tutoring alone) phase were 15 for Tutor A, 11.8 for Tutor B, and 16.4 for Tutor C.

Tutors A and B show effects when baseline and the first tokens for approvals phases are compared. However, the presence of the ascending baseline combined with the precipitous drop beginning with session 34 of the first tokens for approvals phase does not give a strong indication of a collateral effect for Tutor C. All three tutors show effects when the tutoring alone (C) and tokens for approvals (B) phases are compared. The follow-up data in the postcheck phase (C) show short-term maintenance of treatment effects. The presence of the rising baseline data for Tutor C indicated a possible modeling effect. Thus, the decision was made to withdraw Tutor C from the tokens for approvals condition (B) prior to withdrawing Tutors A and B. The subsequent fall in Tutor C's scores and the continuing rise of the scores of Tutors A and B indicated that a modeling phenomenon, if present before, was not being maintained.

Tutors' Correct Reading Responses

A further test of the collateral effects of treatment is shown in Figure 3. The correct reading

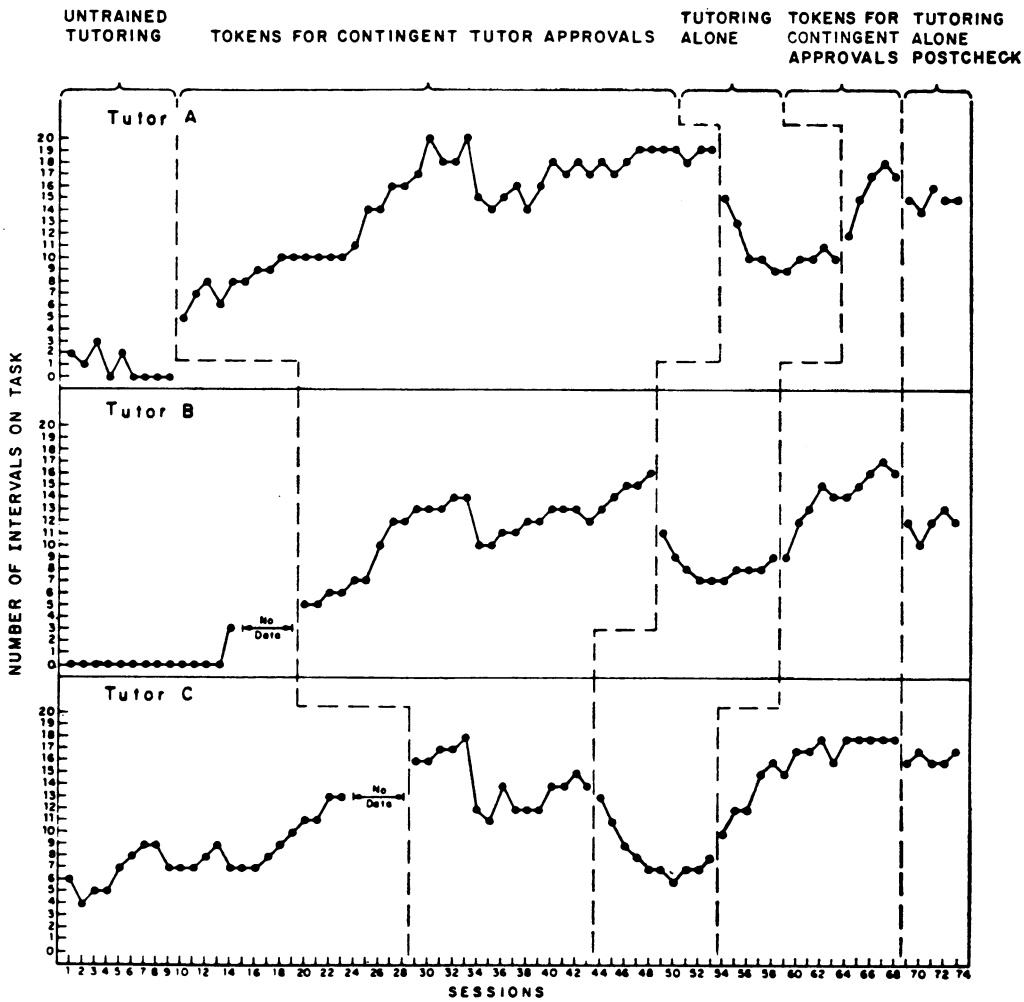

Fig. 2. Tutors' on task in the nonexperimental setting for Experiment 1.

responses of Tutors A and B were extremely low and stable during baseline, while those of Tutor C were not as low nor as stable. The first tokens for approval phase showed a clear treatment effect for Tutors A and B and a less clear effect for Tutor C. The tutoring alone phase showed a drop in achievement for all three tutors compared to tokens for approvals condition (B), but scores returned to baseline levels only for Tutor C. The second tokens for approvals phase (B) showed the highest scores for Tutor B and Tutor C. Tutor A did not achieve his highest score during this phase, a finding that might be attributable to the abbreviated number of sessions (five) in the second experimental phase for Tutor A. Follow-up phase data indicated maintenance of treatment effects for all three tutors when points for number of tutor approvals were withdrawn a second time.

During the untrained tutoring phase, the mean daily reading responses were .22 for Tutor A, 0 for Tutor B, and 3.52 for Tutor C. For the first tokens for approvals phase the means were 9.53 for Tutor A, 6.39 for Tutor B, and 7.93 for Tutor C. After tokens for tutor approvals were removed, the means were 8.5 for Tutor A, 3.9 for Tutor B, and 6.3 for Tutor C. When tokens for tutor approvals were dispensed again, the daily reading means were 14.5 for Tutor A, 11.7 for Tutor B, and 13.6 for Tutor C. For the

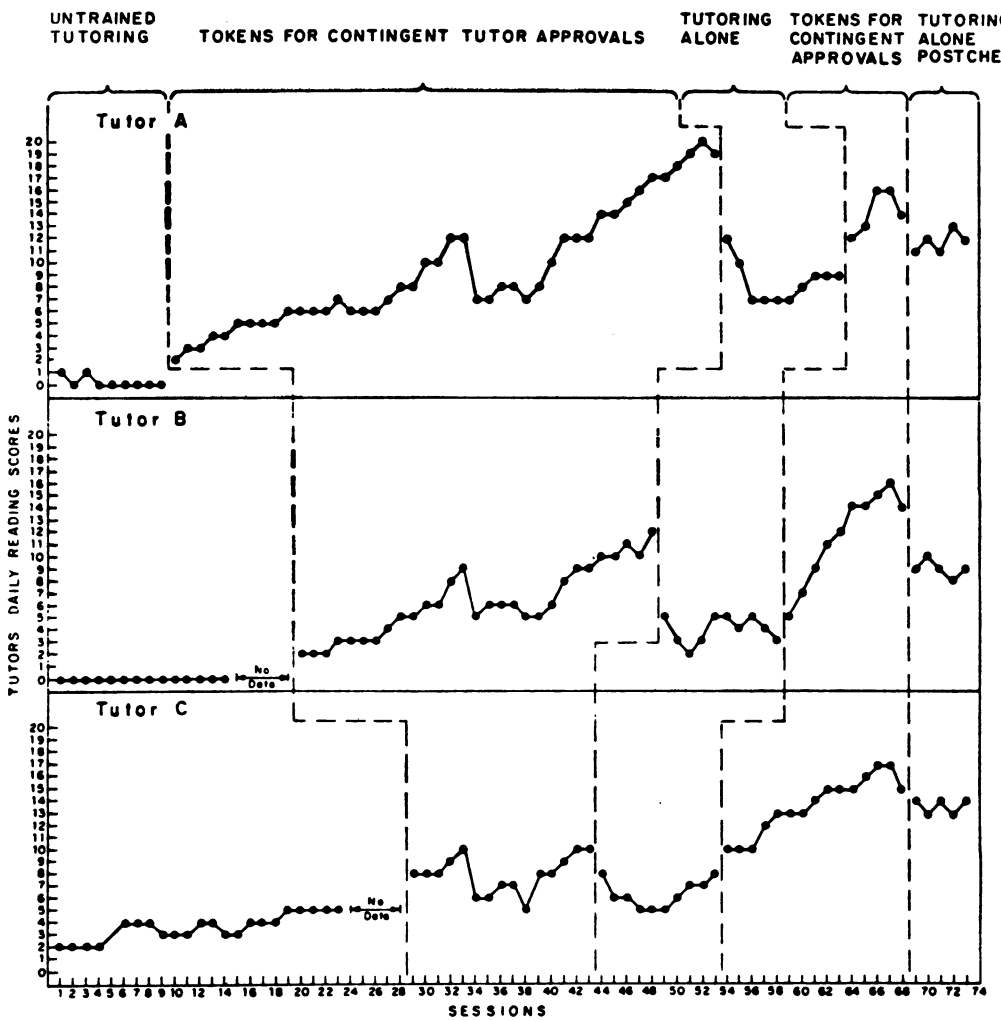

Fig. 3. Tutors' daily reading scores in the nonexperimental setting for Experiment 1.

follow-up phase the means were 11.8 for Tutor A, 9 for Tutor B, and 13.6 for Tutor C.

Discussion

The data showed that tutoring without training did not result in increases in the dependent variables, nor were tutors naturally using approval. However, after five sessions to train tutors in the use of differential social reinforcement followed by dispensation of tokens to tutors for using social reinforcement with the tutees in the tutoring setting, there were increases in the following: (a) the correct reading responses of tutees, (b) the number of on-task intervals of tutors in a separate reading class where no experimental conditions were in effect, and (c) the correct reading responses of tutors in the separate reading class. In general, those behaviors decreased when tokens were initially withdrawn but were maintained without tokens during the follow-up.

There are two important potential contributions from these findings. The fact that for the tutor all the measures other than approval represent nontreated behaviors is of interest because of the suggested potential for multiple educational benefits. A second potential contribution concerns the presence of a maintenance

effect. It should be emphasized that the maintenance test was done only 2 wk after the last tokens for approvals phase. However, a glance at the first tutoring alone phase shows that responses dropped immediately but in the postcheck period taken 2 wk after the last tokens for approvals period, tutor and tutee responses were at levels comparable to treatment. The first experiment offers no clue for the difference in the two C phases. The lack of data on amount of tutor approvals emitted during the first tutoring alone phase precludes strong conclusions concerning the relationship between the dependent variables and approvals per se.

The collateral response effect and the maintenance effect were seen as important enough to warrant a second study. The second study attempted to replicate the collateral effects while expanding the number of variables in such a way as to provide information about potential causes of the maintenance effect, test the validity of the reading responses, and test the degree of relationship between rate of approvals and the other measures.

EXPERIMENT 2

METHOD

Participants and Settings

Tutors. Three eighth-grade boys (ages 13, 15, and 16) were selected to serve as tutors for the second experiment one year after Experiment 1. These students were from the same school and remedial reading program reported in Experiment 1. They differed from Experiment 1 tutors in two ways: (a) they were eighth graders, not ninth graders, and (b) they did not have a severe record of difficulty with school discipline. They were, however, performing below average in reading and had low on-task scores (Figure 6). Tutor A had an SAT reading score of 7.4; Tutor B, 6.8; and Tutor C, 7.

Tutees. Fifteen tutees were drawn randomly from a seventh-grade remedial reading program. Again, five tutees were randomly assigned to each tutor. The mean SAT for tutees was 5.4.

Response Definitions and Data Collection Procedures

The response definitions and data collection procedures were the same as reported in Experiment 1, but with an expansion in the number of dependent variables. Data were collected for tutor approvals of tutees throughout all phases of the experiment. Also, data were taken on tutee attention to tutor approvals. In summary, data were collected on (a) tutors' contingent approval of tutees in the tutoring setting in all phases, (b) the tutees' attention to their tutors' approvals in the tutoring setting, (c) reading responses of tutees in the tutoring setting, (d) on-task responses of the tutors in their own eighth-grade reading class or the nontutoring setting, (e) reading responses of tutors in the nontutoring setting, and (f) three administrations of the SAT for the tutees and tutors.

Tutee attention to tutors and tutor approval. Tutee attention to tutors was defined as verbal or nonverbal approval, acknowledgment, or eye contact given to tutors during tutor-tutee interaction in observation intervals. Examples of verbal approvals were: "Uhuh," "Thank you," "Thanks," "Okay," "I got it," "Look—I got it right." Examples of nonverbal approvals and acknowledgments were head nods, smiles, eye contact, or any combination of these actions. Tutee attention data were collected using the 20-sec observe/10-sec record procedure outlined for observing on-task behaviors. An interval either contained tutee attention behavior(s) or it did not. No distinctions between verbal and nonverbal attention were made when recording data.

Tutor approvals and tutee attention were observed in rotating intervals in the following manner. Two observers were assigned to each tutor-tutee group. During the first 20-sec interval, tutor approvals were observed followed by a 10-sec record interval. During the second interval, tutees' attention to tutors was observed. Rotation of observation intervals continued for 20 observational intervals for each behavior

category. Each tutor received a total of 10 min of observation-record time spread across a 20-min period. Tutee groups also received 10 min of observation-record time spread across 20 min. Observations began 7 min after the late bell sounded. Observe and record cues were used as described in Experiment 1.

Two observers were assigned to each of the tutor-tutee groups, a primary data observer and a reliability observer. Reliability observations were collected for all sessions.

On-task behaviors. Tutor on-task data were collected in a separate remedial reading class in the same manner as in Experiment 1. A reliability observer was present for all observations.

Reading responses. Tutor and tutee daily reading responses were collected in the same manner as in Experiment 1. However, the specific reading materials were changed. For Experiment 2, questions were drawn from the Grolier Educational Corporation's reading program, "Reading Attainment System 1 and 2" (Educational Design Inc., 1975). Reliability for scoring was again 100%. The RAS material was selected because it was more recently norm-referenced for each ability level.

Tutor and tutee reading achievement was also assessed using the Stanford Achievement Test (SAT) administered at the following times: (a) the beginning of the year, (b) at midyear prior to the onset of the experiment (as a pretest), and (c) at the year's end as a posttest.

Reliability

The interobserver reliability (interobserver agreement) for the observational procedures was calculated in the same manner as in Experiment 1. The reliability for tutor approval and tutee attention data ranged from 78% to 96% with a mean of 93%. Reliability for tutor on-task behavior in the nontutoring setting ranged from 88% to 98% with a mean of 92%. Reliability observations were made for all sessions for all dependent variables and tutor approval statements. Three new observers were trained bringing the total number of observers to six. Observers were rotated in the same manner as described in Experiment 1.

Tutor Training

Tutor training was undertaken in the same manner as in Experiment 1 with one exception. The training sessions were abbreviated from five to three sessions. During the three sessions, tutors achieved the same level of skill as tutors had in Experiment 1 during five sessions.

Experimental Design

The design used for the second experiment also was a combined multiple baseline and withdrawal (ABCBC). The phases were: untrained tutoring (A), tokens for contingent approvals (B), tutoring alone (C), tokens for contingent tutor approvals (B) and a 2-wk follow-up phase (C) which was a replication of the tutoring alone phase. The independent variable was amount of tutor approvals to tutees, which in turn was controlled by the dispensation and withdrawal of tokens. The dependent variables were: (a) tutee attention to tutors, (b) tutee reading responses, (c) SAT scores, (d) tutor on-task behavior in a nonexperimental setting, and (e) tutor reading responses in the nonexperimental setting.

RESULTS

Tutor Approval to Tutees

Tutor approval data (Figure 4) indicate that the mean approvals to individual tutees were stable during the untrained tutoring baseline (A), increased during the tokens for contingent tutor approval phase (B), and returned to untrained baseline levels during the tutoring alone phase (C). During the second tokens for approvals phase (B), approvals again increased and were sustained at high levels during the follow-up phase (C).

The mean number of contingent approvals dispensed by tutors to each tutee in their tutee group during the untrained tutoring baseline (A) was 2.46 for Tutor A, 1.62 for Tutor B,

and 2.64 for Tutor C. During the first phase in which tokens were administered to tutors for their approval of tutees (B), the means increased to 4.53 for Tutor A, 4.47 for Tutor B, and 4.71 for Tutor C. The means for the tutoring alone phase (C) decreased to 2.44 for Tutor A, 1.85 for Tutor B, and 1.94 for Tutor C. For the second tokens condition (B), the means again increased to 5.40 for Tutor A, 5.50 for Tutor B, and 5.20 for Tutor C. The means for Tutors A, B, and C for the follow-up phase (C) were 5.08, 4.86, and 4.93, respectively.

Tutee Attention to Tutors

The mean tutee attention responses (Figure 4) to tutors was low and stable during the untrained baseline (A), rose during the tokens for contingent approvals phase (B), and returned to untrained tutoring baseline levels (C) during tutoring alone. During the reinstatement of tokens for tutoring (B), attention again rose and was sustained during the postcheck tutoring alone phase.

The mean number of tutee attention responses for each tutee during the untrained tutoring phase (A) was 1.45 for Tutor A's group, 1.05 for Tutor B's group, and 1.50 for Tutor C's group. For the first token for tutor approvals phase (B), the means rose to 3.42, 3.40, and 3.37 for Tutor groups A, B, and C, respectively. During the tutoring alone phase (C), the means dropped to 1.60, 1.75, and 2.25. For the rein-

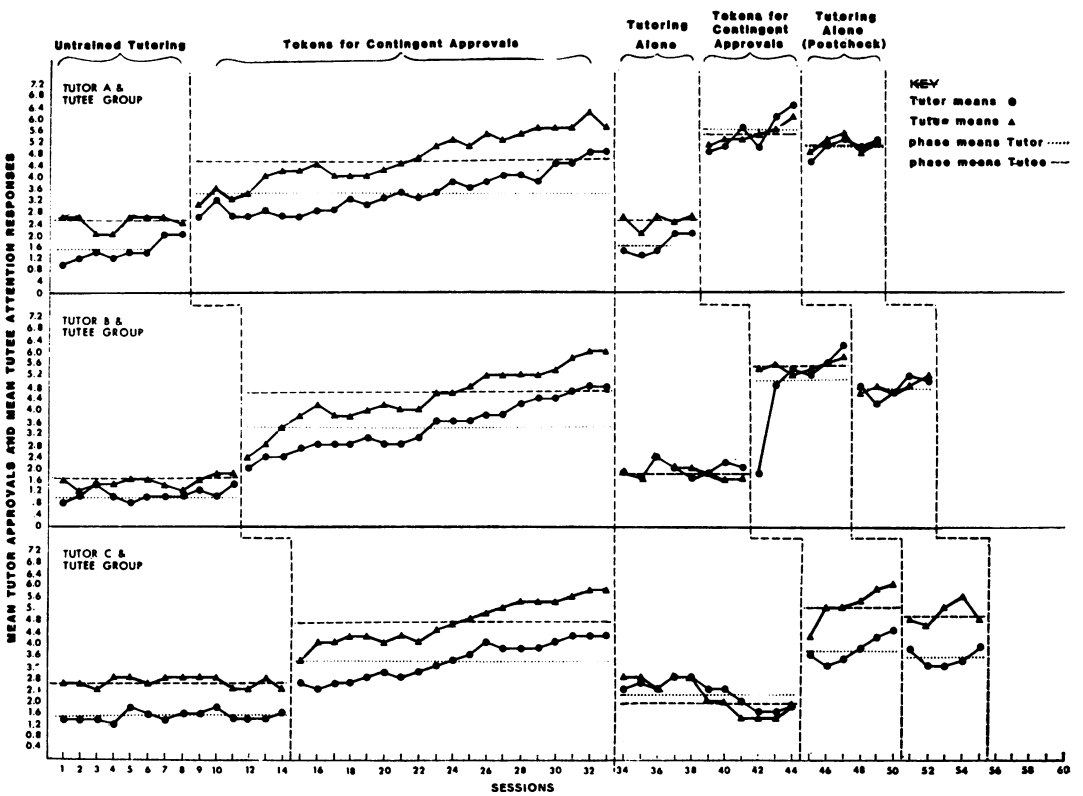

Fig. 4. Mean tutor approvals emitted to individuals in tutee groups for Experiment 2 and mean tutee attention responses. Data points are the total number of approvals emitted per session by tutors during sampled intervals divided by the number of tutees in the tutee group and total number of attention responses of tutees divided by the number of tutees in each group. (For Figures 4, 5, 6, and 7, the postcheck sessions occurred 2 wk after the last session of the preceding phase for each tutor individually. Phase 4 and the postcheck phase were set a priori at six and five sessions, respectively, to control for experimenter bias in setting the initiation and termination of treatment.)

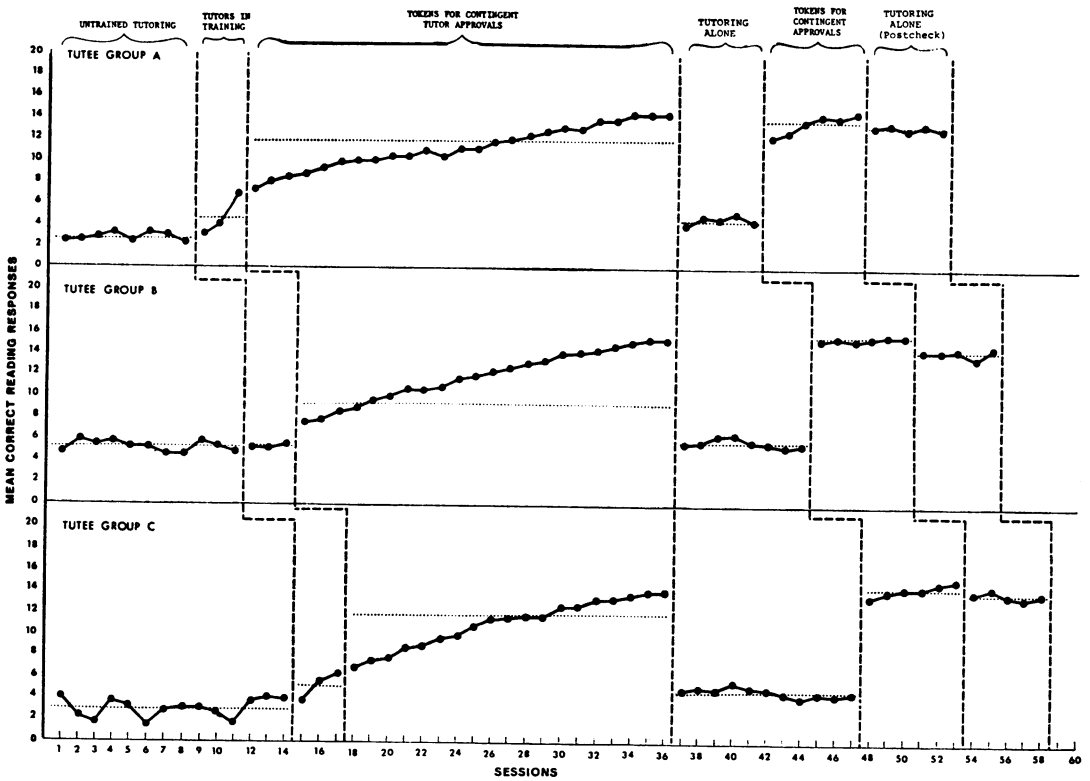

Fig. 5. Mean correct reading responses for tutee groups in Experiment 2. Data points are the total number of correct reading responses per session divided by the number of tutees in the group.

statement of tokens for contingent approvals (B), the means again increased to 5.45, 5.00, and 3.77. During the follow-up phase (C), the means were 4.93, 4.76, and 3.50.

Reading Responses of Tutees

The correct reading responses of tutees are shown in Figure 5. The scores that are graphed for each session are the total scores of all students in a given tutee group divided by five. The graph shows a low rate of correct responses during the untrained tutoring baseline, a slow but consistent acceleration of responses during the tokens for contingent approvals phase and a return to untrained tutoring baseline levels during the tutoring alone phase. During the reinstatement of tokens for contingent approvals phase, mean reading scores again increased. During the follow-up phase or second tutoring alone phase, mean reading scores were only slightly lower than during the second tokens for contingent approvals phase, and substantially higher than during the untrained tutoring baseline or the previous tutoring alone phase.

Daily reading means for each group were summed across phases and divided by the number of sessions in each phase to obtain a mean for each group for each phase. During the baseline, the means were 2.75 for Tutor A's group, 5.22 for Tutor B's group, and 2.89 for Tutor C's group. For the tokens for tutor contingent approvals phase, the means increased to 11.76, 11.09, and 11.86 for the respective tutee groups. The means for the first tutoring alone phase declined to 4.60, 5.85, and 4.62. The means for the reinstatement of tokens for tutor contingent approvals again increased to 14.70, 15.60, and 14.37. The means for the follow-up phase were 13.45, 14.46, and 13.90.

Tutor On-Task Behavior

Figure 6 shows the sampled intervals in which tutors were on task in a separate reading class in which no experimental conditions were in effect. The data show fairly stable rates during the untrained tutoring baseline and the first tutoring alone phase (C). On-task behavior increased during the tokens for approval phases and remained high during the follow-up phase (C).

The phase means were derived by dividing the total on-task intervals in a phase by the number of sessions in the phase. The means for the baseline (A) were 8.88 for Tutor A, 5.91 for Tutor B, and 7.92 for Tutor C. The respective means for the first administration of tokens to tutors for their approvals (B) increased to 15.96, 13.95, and 15.32. After tokens for approvals were removed (C), the means decreased to 8.88, 6.88, and 6.73. When tokens were reinstated (B), the means again increased to 17.83, 17.50, and 16.17. During the follow-up phase (C), the means were 16.75, 15.82, and 14.55.

Tutor Reading Responses

Figure 7 shows the effects of each experimental phase on tutors' reading responses in the separate reading class in which no experimental conditions were in effect. Reading levels were low, stable, and comparable during the untrained tutoring baseline (A) and the tutoring alone phase (C). During the first tokens for contingent approval phase, reading responses rose slowly and consistently. During the second tokens for tutoring phase (B), reading responses returned to peak levels in an abrupt manner when compared with the first trained tutoring alone phase (C). During the follow-up (C),

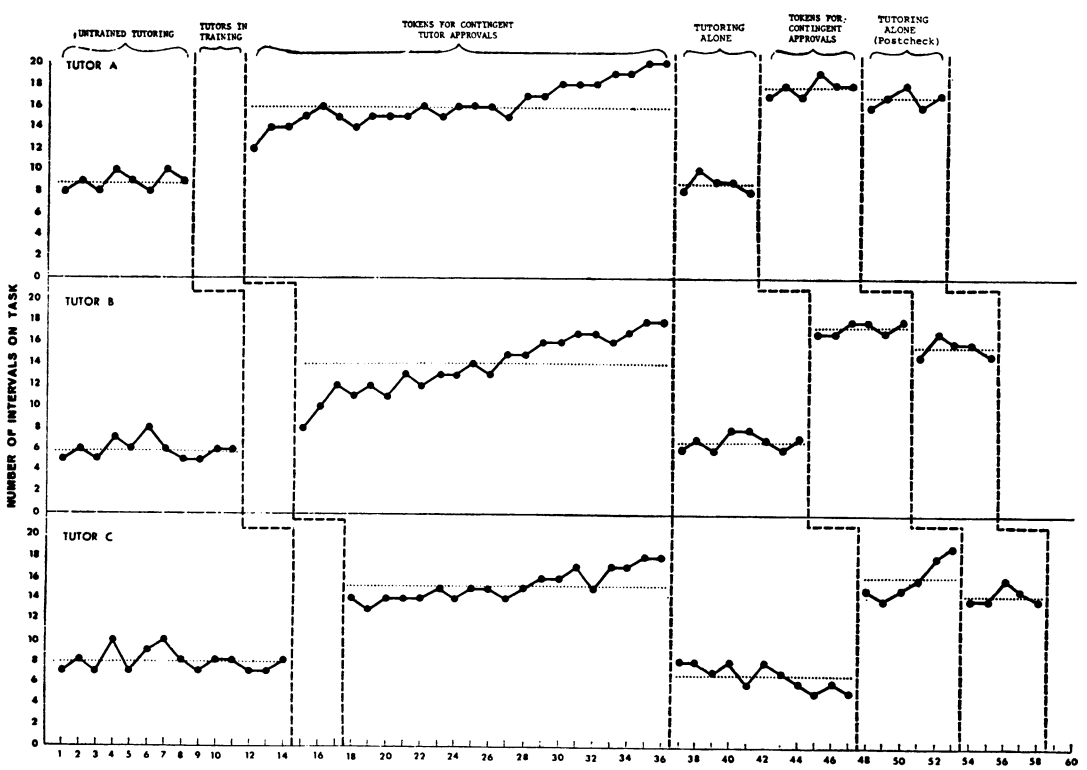

Fig. 6. Tutors' on task in the nonexperimental setting for Experiment 2.

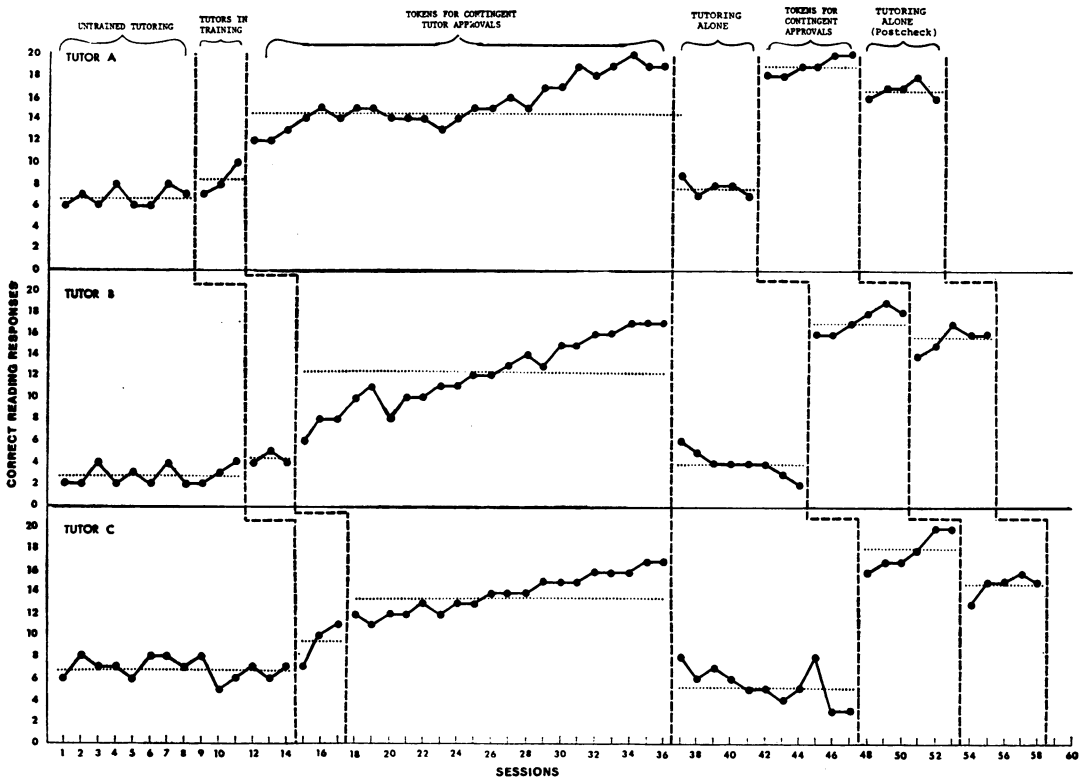

Fig. 7. Tutors' correct reading responses in the nonexperimental setting for Experiment 2.

reading scores were maintained at levels comparable to that of the second tokens for contingent approvals phase (B).

The phase means for reading scores represent the total correct responses for each tutor in a phase divided by the number of sessions in the phase. The means for the baseline (A) were 6.75 for Tutor A, 2.73 for Tutor B, and 6.86 for Tutor C. During the first tokens for approvals phase (B_1), the means were 14.54, 12.27, and 13.37 for Tutors A, B, and C, respectively. The means during tutoring alone (C) were 7.80, 3.25, and 5.10. Means for reinstatement of the tokens condition (B_2) were 18, 17, and 18; and the means for the follow-up phase were 16.80, 15.55, and 14.85.

SAT Reading Scores

The SAT means for each tutee group and tutors are shown in Table 1. Progress during the first semester prior to the experiment was slight. From midyear until after the experiment (posttest), tutees' progress was greater than from the beginning of the year to midyear. Progress for tutors was more pronounced than that for the tutees during the second half of the year.

Table 1

Tutor and tutee SAT reading scores at the beginning of the academic year, midyear (pretest), and at the end of the academic year (posttest).

Ss	*Academic Year Beg. (Sept.)*	*Midyear (Pretest)*	*Gain*	*Academic Year End (Posttest)*	*Gain*
I. Tutor Scores					
A	7.0	7.4	.4	9.2	1.8
B	6.6	6.8	.2	8.0	1.2
C	6.8	7.0	.2	8.6	1.6
II. Mean Tutee Scores					
A	5.1	5.3	.2	6.2	.9
B	5.0	5.3	.3	5.9	.6
C	5.4	5.6	.2	6.5	.9

Table 2

Correlations Between Tutor Responses and Tutee Responses: Experiment 2

Tutor	*Untrained Tutoring*	*Tokens for Tutor Approval*	*Tutoring Alone*	*Tokens for Tutor Approvals*	*Tutoring Alone (Postcheck)*
I. Contingent Tutor Approvals and Tutee Attention to Tutors					
A	.70*	.85*	.72*	.97*	.93*
	(n = 8)	(n = 25)	(n = 5)	(n = 6)	(n = 5)
B	.74*	.84*	.75*	.94*	.91*
	(n = 11)	(n = 22)	(n = 8)	(n = 6)	(n = 5)
C	.78*	.85*	.86*	.88*	.96*
	(n = 14)	(n = 19)	(n = 11)	(n = 6)	(n = 5)
II. Tutee Attention to Tutors and Tutor Reading Responses					
A	.37	.72*	.45	.86*	.82*
	(n = 8)	(n = 25)	(n = 5)	(n = 6)	(n = 5)
B	.74*	.85*	.66*	.81*	.88*
	(n = 11)	(n = 22)	(n = 8)	(n = 6)	(n = 5)
C	.38	.88*	.84*	.92*	.85*
	(n = 14)	(n = 19)	(n = 11)	(n = 6)	(n = 5)

n = number of paired observations
*significant at $p < .05$ level

Correlations between Variables

An index of the degree of covariance between dependent and independent variables was obtained by using the Pearson product-moment coefficient of correlation. The correlations and their significance from zero are shown in Tables 2 and 3. The coefficients are moderate to high and statistically significant in all but one of the phases for correlations between (a) tutor approvals and tutee attention to tutors, (b) tutor approvals and tutor on-task behavior, and (c) tutor

Table 3

Correlations Between Tutor Approvals and Collateral Behaviors: Experiment 2

Ss	*Untrained Tutoring*	*Tokens for Tutor Approvals*	*Tutoring Alone*	*Tokens for Tutor Approvals*	*Tutoring Alone (Postcheck)*
I. Contingent Tutor Approval and Tutor On Task					
A	.75*	.95*	.78*	.92*	.87*
	(n = 8)	(n = 25)	(n = 5)	(n = 6)	(n = 5)
B	.81*	.96*	.91*	.94*	.92*
	(n = 11)	(n = 22)	(n = 8)	(n = 6)	(n = 5)
C	.82*	.94*	.88*	.92*	.88*
	(n = 14)	(n = 19)	(n = 11)	(n = 6)	(n = 5)
II. Tutor Reading Responses and Mean Tutor Approvals					
A	.82*	.97*	.45	.86*	.82*
	(n = 8)	(n = 25)	(n = 5)	(n = 6)	(n = 5)
B	.63*	.78*	.52*	.93*	.90*
	(n = 11)	(n = 22)	(n = 8)	(n = 6)	(n = 5)
C	.86*	.82*	.67*	.91*	.88*
	(n = 14)	(n = 19)	(n = 11)	(n = 6)	(n = 5)

n = number of paired observations
*significant at $p < .05$ level

approvals and tutor reading responses. Correlations in three phases were nonsignificant for tutee attention to tutors and tutor reading responses.

Discussion

When the results of Experiments 1 and 2 are compared, it is apparent that changes in all of the dependent variables were a function of tutor approval of tutee on-task behavior.

Tutor and tutee SAT reading scores increased substantially after tutors had participated in the experiment. Thus the daily reading responses are valid measures of reading. Presumably the differences from the beginning of the year to the preexperimental phase represent a comparable control to the period from the beginning of the experiment to the end of the year. The actual increases for the tutors are greater than those for the tutees.

Tutor on-task behavior and tutor reading responses were shown to be a function of the differential reinforcement that tutors gave to tutees in the tutoring setting. The tutors were not taught any of the behaviors related to those tested in the extra experimental setting. The behavior changes of tutors in the nontutoring setting were described as collateral behaviors. The effect of the treatment on collateral behavior points to the need for a systematic taxonomy of treatment effects such as the generalization network proposed by Drabman et al. (1979).

The occurrence of short-term maintenance effects was shown to be related to the reciprocal approval relationship between tutors and tutees. The relationship was not necessarily causal. It is not known why the approval reciprocity occurred during the postcheck phase and not during the first tutoring alone phase. These findings replicate similar maintenance effects in Polirstok and Greer (1977). The correlations between tutor approvals and tutee attention suggest that the reciprocity between the two variables became most pronounced in the latter phases of the study (Table 2).

Showing collateral behavior changes does not, of course, demonstrate the environmental principles that govern this phenomenon. The findings demonstrate what might be termed "procedure-generated behavior change and short-term maintenance" (Verplanck, Note 3). The reason for this may be that as the tutors used social reinforcement skills, they received increasing attention from tutees. The tutee attention became a discriminative stimulus (SD) as a function of the contrast between the first tokens for tutor contingent approval phase, the following tutoring alone phase, and the subsequent last tokens phase. At the time of the follow-up phase, tutee attention had changed from SD status to conditioned reinforcement status and thus the natural reinforcement potential in peer relationships was maintaining the use of approval by tutors.

The above explanation is very speculative. Future research should locate reinforcers for high-incidence behaviors in the tutors' classroom. Subsequently, the relationship between the tutors' natural reinforcers and collateral behaviors and off-task behaviors could be assessed under experimental and baseline conditions similar to these reported in this paper. Such an analysis might be of considerable interest to the broad and long-term effects of interventions.

REFERENCE NOTES

1. Greer, R. D., & Polirstok, S. R. *Effects of serving as a high reinforcement tutor on the reading achievement and classroom rule compliance of low achievement tutors.* Paper presented at the meeting of the Association for Behavior Analysis, Dearborn, June 1979.
2. The daily scores of individual tutees may be obtained by writing to the first author.
3. Verplanck, W. S. Personal communication, May 1980.

REFERENCES

Cloward, R. Studies in tutoring. *The Journal of Experimental Education,* 1967, **36,** 14-25.

Davis, M. Effects of having one remedial student tutor another remedial student. In G. Semb (Ed.),

Behavior analysis and education. Lawrence: University of Kansas, 1972.

Devin-Sheehan, L., Feldman, R. S., & Allen, V. L. Research on children tutoring children: A critical review. *Review of Educational Research,* 1976, **46** (3), 355-385.

Drabman, R. S., Hammer, D., and Rosenbaum, M. S. Assessing generalization in behavior modification with children: the generalization map. *Behavioral Assessment,* 1979, **1**, 203-219.

Dineen, J. P., Clark, H. B., and Risley, T. R. Peer tutoring among elementary students: Educational benefits to the tutor. *Journal of Applied Behavior Analysis,* 1977, **10**, 231-238.

Harris, V. W., & Sherman, J. A. Effects of peer tutoring and consequences on the math performance of elementary classroom students. *Journal of Applied Behavior Analysis,* 1973, **6**, 587-97.

Hersen, M., & Barlow, D. H. *Single case experimental designs.* New York: Pergamon, 1976.

Kazdin, A. E. The effect of vicarious reinforcement on attentive behavior in the classroom. *Journal of Applied Behavior Analysis,* 1973, **6**, 71-78.

Keller, F. S. "Good-bye teacher . . ." *Journal of Applied Behavior Analysis,* 1968, **1**, 79-90.

Johnston, J. M. On the relation between generalization and generality. *The Behavior Analyst,* 1979, **2** (2), 1-6.

Lovaas, O. I., & Simmons, J. Q. Manipulation of self-destruction in three retarded children. *Journal of Applied Behavior Analysis,* 1969, **2**, 143-158.

McGee, C. S., Kaufman, J. M., & Nussen, J. L. Children as therapeutic change agents: reinforcement intervention paradigms. *Review of Educational Research,* 1977, **47**, 451-477.

Polirstok, S. R., & Greer, R. D. Remediation of mutually aversive interactions between a problem student and four teachers by training the student in reinforcement techniques. *Journal of Applied Behavior Analysis,* 1977, **10**, 707-716.

Risley, T. R. The effects and side effects of punishing autistic behaviors of a deviant child. *Journal of Applied Behavior Analysis,* 1968, **1**, 21-34.

Robertson, S. J., DeReus, D. M., & Drabman, R. S. Peer and college-student tutoring as reinforcement in a token economy. *Journal of Applied Behavior Analysis,* 1976, **9**, 169-178.

Sajwaj, T., Twardosz, S., and Burke, M. Side effects of extinction procedures in a remedial preschool. *Journal of Applied Behavior Analysis,* 1972, **5**, 163-176.

Staats, A. W., & Butterfield, W. H. Treatment of non-reading in a culturally deprived juvenile delinquent: an application of reinforcement principles. *Child Development,* 1965, **4**, 925-942.

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

Stokes, T. F., Fowler, S. A., & Baer, D. M. Training preschool children to recruit natural communities of reinforcement. *Journal of Applied Behavior Analysis,* 1979, **11**, 285-303.

Thurstone, T. G. *Teacher's handbook—SRA reading for understanding: junior.* Chicago: Science Research Associates, 1963.

Received June 12, 1980
Final acceptance August 7, 1981