*PEER TUTORING WITH OR WITHOUT HOME-BASED
REINFORCEMENT, FOR READING REMEDIATION*

JOSEPH TROVATO AND BRADLEY BUCHER

THE UNIVERSITY OF WESTERN ONTARIO

An operant-based corrective reading program was established to study effectiveness of peer tutoring in the school, for reading deficient children. Sixty-nine second to fourth grade students were matched on measures of initial reading ability and level of reading retardation, and randomly assigned to three groups: Peer Tutoring Only, Peer Tutoring with Home Based Reinforcement, and Control. SRA materials were used in training for the experimental groups, supplemented with additional reading materials. The program continued for 15 weeks, in seven public schools. Changes in oral reading accuracy and comprehension were assessed. Both reading and comprehension were significantly increased by peer tutoring, relative to the control group. The addition of home-based reinforcement doubled this increase. The measured gain in oral reading, based upon standardized testing, was 0.19 years for the controls and 1.27 years for peer tutoring with home-based reinforcement. Internal measures of gain showed similar results, and comprehension gains were also comparable. A high rate of compliance with the tutoring program was maintained by the tutors. High ratings of satisfaction were obtained for the program, from all groups of participants. The feasibility of the program for application in the school system is discussed.

DESCRIPTORS: reading, tutoring, comprehension, peers, elementary school students

Several investigators have demonstrated that contingency management procedures can effectively modify academic performance (Coleman, 1970; Harris, Finfrock, Giles, Hart, & Tsose, 1975; Kazdin & Bootzin, 1972; O'Leary & Drabman, 1971; Willis, Crowder, & Morris, 1972). These procedures are often costly in terms of teacher time. Measurement and observation of performance, and the management of teacher attention or other contingencies may be impractical in many classroom settings.

The behavior of children in school can also be influenced by reinforcement administered at

home by parents. Use of such a procedure may decrease the demands on the teacher. For example, Harris et al. (1975) employed a home-based reinforcement program and a response cost procedure to increase weekly assignment completion. Recent investigations have also demonstrated that token reinforcement procedures can be monitored effectively by the peers of underachieving children (Willis, Crowder, & Morris, 1972; Willis, Morris, & Crowder, 1972), and that the peers of underachieving children can effectively tutor children at a number of educational levels (Davis, 1972; Jenkins, Mayhall, Reschka, & Jenkins, 1974; Niedermeyer & Ellis, 1971; Richardson & Havlick, 1975; Shaver & Nuhn, 1971). Until recently most tutorial studies have been evaluated in terms of verbal reports from teachers and students rather than in terms of objective data documenting increased academic performance. Using objective measures, Willis, Morris, and Crowder (1972) and Willis, Crowder, and Morris (1972) showed improvement

Reprints and copies of the manuals may be requested from Bradley Bucher, Psychology Department, University of Western Ontario, London, Ontario, Canada; or Joseph Trovato, Department of Behavior Therapy, Oaklands Regional Centre, Oakville, Ontario, Canada. We thank the many people in the London Public Schools who helped us, especially Dr. Richard Stennett, Research Psychologist for the London Board of Education. This research was supported by a grant from the Canada Council.

in the average reading performance of over fifty children, enrolled in grades two through eight. Further support for this approach is given by Robertson, DeReus, and Drabman (1976).

The separate studies of home-based reinforcement and peer tutoring indicate that potential benefits may accrue from the union of these two methods. The two procedures were studied in combination by Wilson (Note 2), who found that when peer tutoring was combined with a home-based reinforcement procedure, gains in reading accuracy were greater than the gains achieved by peer tutoring alone. In addition, gains achieved in the combined condition were consistent across all subjects whereas gains for peer tutoring alone were found for some subjects but not for others. However, Wilson's study reports data for only five subjects, and no clear effects of contingencies were shown for reading comprehension.

In research on reading the importance of reading comprehension is often overlooked. Staats and Butterfield (1965) and Staats, Minke, Goodwin, and Landeen (1967) gave contingent tokens for accurate oral reading of words and passages and for correct answers to questions about silently read passages. Significant pre-post test changes in comprehension were found for a single subject in Staats and Butterfield (1965) but no significant changes were observed for 18 subjects in Staats et al. (1967), nor by Camp and Van Doornick (1971) using similar procedures. Lahey, McNees, and Brown (1973) studied comprehension more directly, for two children. They found comprehension increases, but there was no evidence that the training results were more general than for the particular materials involved in training.

The present study examined reading accuracy and reading comprehension for groups of children given peer tutoring, with or without an added home-based reinforcement system. Gains were assessed using measures both internal and external to the training program. In addition, some auxiliary measures were taken. First, tutor reliability was assessed for accuracy of tutors

in following the procedure in which they were trained. In a previous study of a peer tutoring procedure, for arithmetic, Johnson and Bailey (1974) found considerable variation in the accuracy with which different tutors followed the procedure. Second, ratings of the program were solicited from its various participants, following the recent advice of Wolf (1978) that behavior analysts must begin to ask their clients and customers about the acceptability of their procedures.

METHOD

Subjects

Children were selected from seven local elementary schools, through a variety of referral sources. Children from a range of socioeconomic levels were included, but predominantly from the middle class. All were children who were identified as deficient in reading skills, on the basis of ongoing school performance, as verified by grade records and teacher interviews; and whose parents expressed tentative willingness to cooperate in a home-based reinforcement program. Ninety such children were selected from grades 2, 3, and 4.

Experimental Design

A group design was used, with three groups. Children were matched in triads. The members of each triad were then randomly assigned to the three experimental groups: control (C); peer tutoring (PT); and peer tutoring plus home-based reinforcement (HB). Assignments to triads were based on four measures: level of reading retardation, and three pretest reading measures. Reading retardation level was the discrepancy in years between the teacher's estimate of grade level in reading and a mean age-grade table for the Province of Ontario, as provided by the local board of education. Two of the three pretest reading measures were the initial test scores from the Spache Diagnostic Reading test (oral reading and comprehension subscales). The third was the pretest score from the SRA Starter Stories (Parker & Scannell, 1973).

Final selection depended on the results of a parent interview. One parent of the home-based member of a triad did not agree to participate when the program was outlined in some detail, so the triad was eliminated. Six other triads were eliminated for various reasons (non-English speaking parents, no telephone or unlisted phone) for the child in the HB group. Twenty-one subjects were thus eliminated, leaving 69 children or 23 triads. Eleven students were from grade two, 48 from grade three and 10 from grade four.

T-test values were computed as a device for comparing all pairs of groups on all of the four pretest measures. All *t*-values were non-significant.

The 23 subjects in the Peer Tutoring Group (PT) received peer tutoring for approximately 30 min per available school day. The 23 subjects in the Peer Tutoring plus Home-based Reinforcement Group (HB) received peer tutoring for approximately 30 min each available school day, and points earned for accurate oral reading and comprehension could be exchanged for rewards in the home. The tutoring sessions were arranged so as not to conflict with whatever reading instruction the children received in their classrooms. Twenty-three children served as a control group (C). This group included some children who were enrolled in in-school remedial reading programs in addition to in-class reading instruction. No attempt was made to control the instruction received in this group.

Reading Materials

Two types of reading material were used. The basic reading was the SRA Reading Kit (1b) (Parker & Scannell, 1973), which ranges from grade level 1.4 to 4.5 (based on a readability formula that analyses sentence length and average number of syllables per word). Reading selections are arranged in levels of increasing difficulty, each of which is identified by a specific color. Comprehension checks follow each reading. Supplementary readers were also made available for each student. Calibration with SRA

material was accomplished using the Fry readability formula (Fry, 1968).

The initial SRA reading level was based on the student's oral reading and comprehension performance in pretests using the SRA starter stories. This initial level was the highest difficulty level for which the student scored above 80% accuracy on both oral reading and comprehension, measured in terms of lines read and questions answered without errors.

After reading began, new SRA material was selected on the basis of daily performance, using three criteria: reading error rate (lines read with errors) on first reading, and on second reading (done only for lines incorrectly read at first reading), and comprehension error rate. To be moved to the next difficulty level in the SRA stories, a student had to obtain an average error rate of 30% or less on first reading in both SRA and supplementary reading, and 20% or less on second reading and on comprehension, for three consecutive days. If the student did not meet these criteria in any one of five consecutive SRA stories at one level, the student was moved back one level. If all the SRA stories at any one level were exhausted, and the criteria were not met, the student read from the supplementary reader until a criterion of 20% error rate on initial reading was met.

Pretests, Posttests, Dependent Measures

Overall change in reading ability was assessed using the oral reading and the comprehension subscales of the Spache Diagnostic Reading Scale (Spache, 1963) and the test of reading ability provided by the SRA Starter Stories.

Pretests were administered a maximum of one week prior to commencement of the reading program. The posttests were administered a maximum of one week following its termination. All testing performance was reinforced, on the basis of literature that suggests that extrinsic motivation can influence test performance (Ayllon & Kelly, 1972; Clingman & Fowler, 1975, 1976; Edlund, 1972; O'Connor & Weiss, 1974).

The test was presented as a game whose outcome depended on the number of red poker chips accumulated, one for each line read correctly or question answered correctly.

Throughout the study daily reading performance was assessed using the total number of lines read; the number of lines read correctly on first reading; the total of lines correct after the second reading; the number of comprehension questions answered correctly; and the number of comprehension questions answered incorrectly.

Program Managers and Peer Tutors

Much of the day-to-day supervision was done by Learning Resource Teachers who volunteered for the study and had been trained to manage tutor-student pairs. These Resource Teachers are typically responsible for implementing and supervising such special programs. They were responsible specifically for tutor selection and training and day-to-day supervision of reading sessions. The first author was responsible for measuring tutor reliability, monitoring Learning Resource Teachers, implementing and monitoring the home-based reinforcement systems, monitoring day-to-day progress of students, assigning new reading materials as required by the program specifications, and collecting data.

Peer tutors were obtained from the senior grades in the schools from which the students were selected (6th, 7th, and 8th grades). Tutors were recruited on the basis of high scholastic standing, willingness to tutor, and teacher agreement for their participation. Backup tutors were also obtained, for cases where other tutors dropped out, or were not available for a given session. There were 98 tutors in all.

Learning Resource Teacher training involved three sessions of approximately one and one-half hours each, following the guidelines of a manual entitled "Program for training student tutors in reading". Tutors were trained by the Learning Resource Teachers. Training involved a maximum of six sessions of approximately one-half hour each. During training the tutors used a

tutor's manual which described each step of the tutoring procedure in detail. The teacher training manual was adapted from an earlier manual by Mothersill, Hundert, Lai and Willis (Note 1).

Tutoring Procedure

The tutor picked up the student from class and went to the tutoring room at a prescribed time. The tutor went to the SRA materials box and selected the reading materials specified in the student's record book. This book was updated every two or three days by the experimenter. The student then read from the SRA story while the tutor followed, noting and marking errors, as described below. The tutor watched the student's text or a duplicate copy, which was attached to each story in the SRA box. If time permitted after an SRA story was completed and comprehension questions answered, the student read from the supplementary reader.

A vocabulary error was defined as pausing for 5 sec without trying to read a word, reading a word incorrectly, or leaving out an unfamiliar word. The term unfamiliar was left to the tutor to determine. Other errors were: adding words, omitting familiar words, or mixing word order. If the student self-corrected a mistake before reaching the end of a line, and before the tutor responded, the mistake was not counted.

For each line of text read correctly the tutor gave the student a blue chip and, occasionally, praise. For each line read with an error that was not self-corrected before the line was completed, the tutor placed an "X" at the end of the line. For vocabulary errors the tutor also penciled an "X" above the error. The tutor made these markings on the tutor's text, which was not also the student's text when duplicate texts were being used. To ensure that the student noted the error, the tutor also stated briefly what kind of error was made. When a page was finished the tutor asked the student to reread all the lines where an "X" appeared. If the line was reread correctly, the tutor put a circle around the "X" wherever it appeared and gave the student a

blue chip and praise. If the line was reread with another error or the same error repeated, the tutor made another "X" and helped the student read the line correctly.

After the student completed an SRA story the tutor folded the story out of sight and asked the student to read the comprehension questions to himself, and then asked the student to write the answers on a card. When a student indicated that he or she was finished, the tutor corrected the answers. For each correct answer the student was given an orange chip and praise. For incorrect answers the tutor marked an "X" and gave the correct answer. Correctness was decided on the basis of content, i.e., factual correctness, rather than on correct spelling. The Flash Card Procedure followed (see below), and then reading from the supplementary reader, for the remaining time available. Points and chips were also given for this reading. Finally, the student and the tutor counted the chips that were earned, and collected the chips and other materials. The students in the home-based reinforcement group were given a "points transfer card" which showed the number of points earned that day (blue chips earned one point, orange chips earned five). The PT students received no points for chips and the number of chips earned was not recorded.

Flash Card Procedure

When a vocabulary error was repeated on the second reading, the tutor recorded the word, and at the end of each SRA story or supplementary reading the tutor printed each word on a flash card. The tutor then showed the words in sequence to the student, first asking the student to say the word, and then helping if necessary. If the student said the word, the tutor concealed the word and asked the student to spell it, and then showed the word if the student did not succeed, and helped with the spelling. When an error was made at any step the card was placed at the bottom of the pile. If the student completed both saying and spelling without help, the tutor removed the word from the pile.

The tutor continued until all words were mastered. These trials were nonreinforced.

Home-based Reinforcement Procedure

The first parent contact was by telephone to describe the program; reconfirm the parents' interest (permission letters had been previously obtained); discuss the concept of the reinforcement system and its operation for the reading program, the types of rewards that might be used, and the use of the point-delivery form which would inform them of the child's daily achievement. The parents were asked to discuss the program with their child, particularly to determine events and items that could be used as reinforcers. Two days later another call was made to determine what reinforcers were agreed to by the parents and the child, and to set up an appointment for a home visit. In the interim the reinforcers were incorporated into a tentative contract which specified how the points earned could be exchanged, and described the accounting system used to monitor earning and spending of points. On the home visit a typewritten copy of the contract was given to the family with a brief guide to the nature of the contract procedure and a list of sample reinforcers. The exchange system was discussed in more detail, and the guide to the home-contract procedure was explained and discussed. The guide repeated what had been discussed orally, and included short sections on reasons for use of rewards; what the child did to earn points; how point-exchanges were made and recorded; how to handle problems such as hoarding, requests for credit, and extra-experimental child misbehavior; and how to make contract renegotiations. Additional reinforcers could then be written into the contract. Reinforcers included those items typically used in home-contracting procedures, including access to desirable activities, either routine or occasional; preferred foods and similar small tangible items; and usually some high-cost items as well. The most popular reinforcers were to be able to invite a friend to the home to play or to go with the family to a ham-

burger franchise establishment. The average number of reinforcer items listed was 11.5, with a range of 6 to 18.

When all parties were satisfied, the contract was signed by parents and child, and then posted conspicuously. The child was given a personal bank book to record earnings and expenditures. Throughout the program telephone contact was maintained, to encourage continued and appropriate parental involvement. Parents were asked if their child was bringing his points-record sheets home, and the current bank book balance was checked against the experimenter's record to check if points were being correctly added and removed from the bankbook. Problems with the contract, or in other areas, were dealt with briefly. Such problems were raised very infrequently. After the project had ended a final home visit was made. The bankbook and the current contract were collected, and the parents were given a questionnaire to fill out and return by mail.

Tutor Reliability

Periodic checks were made at the beginning, middle, and end of the study, by the first author and by an independent observer, to assess the reliability of the tutor's performance in carrying out program procedures. For each line read the assessors measured whether the tutor responded accurately to the student's reading, and scored the line as read correctly or incorrectly, and responded properly to errors by marking them and reassigning improperly read lines as required. For each comprehension question the assessors measured whether the tutor accurately managed the comprehension question and scored the answer. The assessors scored the tutor for accuracy of recording the session data also. Errors in tutoring were discussed with the tutors.

Program Evaluation

Questionnaires were given to Learning Resource Teachers, teachers, principals, parents, tutors, and students, asking how satisfied they were with various aspects of the program. These

questionnaires were constructed using a 7-point scale. All evaluations were returned anonymously. The parents were told that the questionnaires were completely anonymous and that honesty in answering was essential if any true evaluation was to follow. They mailed their responses in preaddressed stamped envelopes. The other questionnaires were given to the Learning Resource Teachers to distribute in their individual schools, and returned to the board of education office.

RESULTS

Twenty-three students were included in each group. Their average ages were: 9.1 (HB), 9.0 (PT), and 9.1 (control). In interviews with teachers when they recommended one of their students, the teachers were asked how far the student lagged in reading performance from the grade level appropriate for the students' age. Age-achievement discrepancies, based on these reports, were: 1.62 years (HB); 1.56 years (PT); and 1.59 years (controls).

The average time between first and last training session was 15 weeks, 2 days. Based on a school year of 43.3 weeks this is .35 academic years. Average number of sessions was 46, and the average session-days/school-days was .74.

Pre-post Test Data

Figure 1 shows results for four measures; three derived from the Spache Diagnostic Reading Scales and one for the SRA Starter stories.

Spache Diagnostic Reading Scales

Oral reading. The Instructional Level subscale of the Spache Diagnostic Reading Scales was used to measure oral reading. The average gain from pretest to posttest for the HB group was 1.27 years (range .5 to 3.2). The average gain for the PT group was .77 years (range .0 to 1.7). For the control group, the average gain was .19 years (−2.7 to .7).

A 2×3 analysis of variance was done, with Time of Testing as a within-subject factor and

Fig. 1. Average pretest and posttest scores for four measures of reading performance: Spache Instruction Level, Spache Independent Level, Spache overall reading performance, and SRA Starter Story performance. Scores are given for the home-based reinforcement group (HB), the peer tutoring only group (PT), and the control group.

Teaching Mode as a between subject factor. The main effect of Time of Testing was significant ($F(1, 132) = 24.49, p < .001$). The main effect of Teaching Mode was not ($p > .05$). The Time of Testing $\times$ Teaching Mode interaction was significant ($p < .05$), indicating that increases in oral reading depended on the type of teaching. Newman-Keuls tests revealed significant differences between all groups at posttest, at better than the .001 level.

Comprehension. The Independent Level sub-scale of the Spache was used to measure comprehension. The average gain from pretest to posttest for the HB group was 1.04 years (range .0 to 2.7). The average gain for the PT group was .50 years (range .0 to 1.7). For the control group, there was a loss of .13 years (range -2.0 to .7).

The data were analyzed using a 2×3 analysis of variance, with Time of Testing and Teaching Mode as factors. There was a main effect for Time of Testing ($p < .001$), and a significant interaction ($p < .01$). Newman-Keuls analyses of posttest data revealed significant differences at better than the .001 level between all groups.

Overall reading. The overall reading ability score was calculated as the average of the Spache scores for oral reading and comprehension. These combined data were also subjected to a 2×3 analysis of variance. The main effect of Time of Testing was significant ($p < .001$), but Teaching Mode was not. A significant interaction ($p < .05$) showed that increases in overall reading ability were dependent on the type of teaching. Significant differences at posttest were found between all groups ($p < .001$), using the Newman-Keuls test.

SRA Starter Stories

The average gain from pretest to posttest for the HB group was 1.35 years (range .0 to 2.7). For the PT group the gain was .95 years (range .0 to 1.8). For the control group, the average gain was .23 years. Correlations between SRA gains and the Spache measure of overall read-

ing ability were significant at the .001 level, for each group.

A 2×3 analysis of variance showed an effect for Time of Testing, Teaching Mode, and their interaction, all at $p < .001$. Newman-Keuls analyses showed significant differences for all posttest comparisons between groups ($p < .001$).

In summary, measures of oral reading, comprehension, and overall reading all showed that the home-based reinforcement children significantly outperformed the peer tutoring children who were not given the home-based reinforcement component; and these children significantly outperformed the control children.

Daily Reading Records

Reading gains were also analyzed using daily records of error rates and SRA reading levels. Table 1 summarizes these results.

At the end of the program the HB group was reading at an average grade level 1.69 years more advanced than at the beginning. On the average, they progressed through 7.61 SRA reading levels, reading 6.26 stories at each level before achieving mastery of the level. Average error rates fell as reading levels increased. The average error rate for the first level was 16.48%, and for the final level the average was 12.78% (error rates on first reading). Error rates for

Table 1
Average Results from Daily Reading Records

	HB	PT
SRA grade level advance	1.69	1.05
SRA reading levels mastered	7.61	5.65
SRA stories read/reading level	6.26	8.19
Lines read	4,223	3,645
Comprehension questions answered	219	214
(Error rate/line) for first reading level	16.48%	22.18%
(Error rate/line) for last reading level	12.78%	20.06%
Error rate for comprehension first level	6.4%	9.4%
Error rate for comprehension last level	6.2%	7.3%

comprehension also fell. Children read from 1,194 to 8,376 lines of text (average = 4,223) and answered from 160 to 290 comprehension questions (average = 219).

For the PT group, the average program gain was 1.05 years. On the average they progressed through 5.65 SRA levels, reading 8.19 stories at each, before achieving mastery. As reading level increased, average error rates fell, for both reading and comprehension. Children read from 1,233 to 7,786 lines of text (average = 3,645) and answered from 165 to 290 comprehension questions (average = 214).

When paired HB and PT children were compared, the HB child showed greater SRA starter story gains than the PT child in 13 of the 23 comparisons (with one loss and 9 ties). The HB child also mastered more reading levels, in 19 comparisons (with 2 losses and 2 ties). Average error rates in both oral reading and comprehension were lower for the HB group than for the PT group (12.8% vs. 17.99% for oral reading, $p < .001$; and 5.72% vs. 8.15% for comprehension, $p < .05$).

In summary, daily reading records corroborated and strengthened the results of pre-post testing. The home-based reinforcement group showed the greater gain. Progress of a year or more was achieved in the SRA material while the error rates for reading and for answering comprehension questions decreased.

Parent Contact Data

The average initial telephone contact time was 10 min, and 3 min for the second call. The home visit required an average of 15 min. Subsequent telephone contacts were made on the average of once every two weeks for approximately 3 min per contact, and an average of seven contacts per family. Although parents had been advised that consultation was available for any problems that arose, such problems were rarely mentioned. The telephone checks were devoted to examination of points recording accuracy. Of 161 checks for proper maintenance of the bankbooks, difficulties were re-

corded in 23 cases. These were all discrepancies between the parents' and the experimenters' count of points, due to incorrect additions or subtractions, or instances where the parents or the child did not ensure that points earned or spent were recorded in the bankbook. A final home visit was conducted upon termination of the program, which required an average of 10 min. Thus the total parent contact time, including telephone contact and home visits, was approximately one hour per family. Other problems were dealt with only briefly, primarily by restating portions of the parent guide. The experimenters aimed to reduce interaction to the essentials, to avoid involvements with problems with which a nonclinically trained professional would not be prepared to deal.

Tutor Performance, Tutor Reliability, and Interrater Agreement

Ninety-eight tutors served in the study (2.1 tutors per student). These included backup tutors who were available when regular tutors were not. These also included six tutors who dropped out of the program: two were discontinued because of consistent low reliability and four dropped voluntarily.

Reliability estimates were made of the tutor's accuracy in scoring a line read (first and second reading), a comprehension question, and the recording of the session data. Agreement was measured between the tutor and an independent rater. Agreement was defined as the identical judgment by the two persons. A percent level of agreement was determined as the number of agreements divided by the number of agreements plus disagreements, times 100. Tutors were observed periodically in sessions scattered throughout the procedure, in rotation across tutors, although unreliable tutors were observed somewhat more frequently. Reliability for first reading was 91.3% (1,180 lines), 95% for second reading (101 lines), 92.3% for comprehension (130 questions), and 91.5% for recording the data (118 sessions). These results in-

clude two tutors whose persistent low accuracy led to their dismissal.

Interrater agreement was also assessed. Two independent raters scored the tutor's performance as correct or incorrect for the four performance components described above. For 1,645 observations, agreement about the accuracy of the tutor's performance averaged 96.7%, and the averages were very similar for each of the four components.

Questionnaire Data

All Learning Resource Teachers responded to the questionnaire. The LRTs were asked if the program was effective and whether they would recommend it. Answers ranged from an average of 1.33 to 2.00 on a scale of 1 to 7, with "completely satisfied" as 1, and "completely dissatisfied" as 7. Eighteen of twenty-two teachers responded to the questionnaires. The teachers rated five questions about the effectiveness and feasibility of the program, and whether they would recommend it. Average scores ranged from 1.28 to 2.06. All principals replied to their questionnaires. Principals rated three questions about program effectiveness, feasibility, and value. Average scores ranged from 1.50 to 1.83. Twenty-two of the twenty-three parents in the home-based group returned their questionnaires. Parents rated some aspects of program effectiveness, and whether they would recommend the program. Average scores for the five questions ranged from 1.05 to 1.86. All tutors completed their questionnaires. Average scores for the four questions ranged from 1.16 to 1.62. They rated that they were happy they were chosen as tutors, felt they were effective, and had learned from the experience.

Twenty-two of the twenty-three HB students responded. They rated that they liked sessions (average = 1.09), liked learning this way (average = 1.09), liked not making a mistake for a line or a question (average = 1.23), getting a chip (average = 1.09), and points (average = 1.00). They rated that they did not like making mistakes (average = 6.82). Twenty-

two of the twenty-three PT students answered. They rated that they liked sessions (average = 1.59), liked learning this way (average = 2.04), and liked not making mistakes for lines or questions (average = 1.64). They were neutral about receiving a chip (average = 3.78), and didn't like making a mistake (average = 5.91). The question about point values was omitted, since the PT group received no points for chips.

Post hoc *t*-test analyses for differences for the two groups showed significance for four of the five questions. The HB group liked the sessions more ($p < .05$), liked learning this way more ($p < .05$), liked receiving a chip more ($p < .001$), and disliked making mistakes more ($p < .05$).

DISCUSSION

The results of the present study clearly demonstrated that a corrective reading program using peer tutors was effective in increasing the oral reading and comprehension performance of under-achieving children, as compared to the more usual procedure in effect in the schools. The results also indicate that the addition of home-based reinforcement contributed even further to the gains. The gains for the home-based reinforcement group in both oral reading and comprehension were almost double the gains reported for tutoring alone.

The overall reading change measures used to support these conclusions were calculated as the average of scores obtained on the oral reading and comprehension subtests of the Spache Diagnostic Reading Scales. The peer tutoring with home-based reinforcement group showed an overall reading ability gain of 1.16 years whereas the gain for the control group was .03 years. These values are calculated on the basis of a 43.3-week year. These rates, if taken as realistic, reflect gains well in advance of those expected on the basis of the passage of time alone, and are far in excess of those registered for the control group.

The SRA overall reading ability test supported the data from the standardized test. Fur-

ther support came from daily reading records and daily records of errors. By the end of the program, the home-based reinforcement group had advanced in reading by an amount that was twice that of the peer tutoring group, and yet was making significantly fewer errors in both oral reading and comprehension.

A primary source of evidence of gains in this study was derived from the Spache Diagnostic Test. This choice of a pre-post test measure dictated the use of a group design. The superiority of standardized test data to within-session data is debatable, so that both were used here. Use of standardized test data to measure training gains may be compromised by research showing that results are susceptible to influences from feedback and reinforcement (O'Connor & Weiss, 1974). Thus test results may vary across subjects not only as a function of prior skill acquisition, but as a function of motivational factors in the testing conditions. It is unclear how any standardized set of test conditions could eliminate differential effects of motivation. In our study we chose to test under incentive conditions that included feedback and reinforcement, on the suspicion that a standardized test procedure might fail to motivate students to show the skills they had acquired during training. Use of a reinforcement procedure might shift scores upward, but this effect is unimportant unless the three groups were differentially affected in their difference scores. This might occur as a result of the differential exposure of the groups to reinforcement during training. That this is not likely is suggested by Ayllon and Kelly (1972), who found that previous exposure to a token reinforcement procedure did not affect the degree to which this procedure raised test scores.

No detailed analysis was undertaken here of the effective components of either the peer tutoring or the home-based reinforcement procedures. In particular, no direct observations were made of the accuracy with which parents managed their reinforcing function in the home. Even if properly managed, this function could

have included a variety of socially mediated contingencies for reading performance. The ratings of satisfaction from tutored students contribute some information on the incremental value of the home-based procedures. Ratings for the home-based students were much higher for the question whether the student liked to receive chips, which they could exchange for rewards in the home. This questionnaire result thus indicates (tentatively) that the chips actually possessed reinforcement value, as they were designed to do. However, the value of the chips may have derived from several sources: the actual goods and services they obtained, increased positive interaction with parents, or avoidance of negative interactions. One other questionnaire result may serve as a useful guide in subsequent use of the home-based procedures. Both treatment groups indicated strongly that they disliked making an error, and the home-based group rated this effect almost at the extreme end of the scale (6.82), even more strongly than did the peer tutoring group (5.91). This result implies that it is important to keep error rates low, and may be even more important if errors deprive students of exchangeable tokens. High error rates may produce aversive reactions.

Several investigators have recently indicated that the practical results of an applied program should be measured not only by the objective data that are obtained, but also in terms of consumer response, or social validity. Social validity as described by Wolf (1978) refers to three aspects: the social significance of the goals, i.e., whether society values what is being done; the social appropriateness of the procedures, i.e., whether the participants or consumers consider the treatment procedures acceptable; and the social importance of the effects, i.e., whether the consumers are satisfied with the results. Social validity may be assessed in many ways, and it seems likely that different methods of assessment will show different degrees of agreement. Most applied researchers would probably be highly reinforced by social validity measures that

reflect actual implementation of the program or procedures they have developed. Such results tend to come slowly. Questionnaire data provide more immediate measures which may relate to the eventual acceptance of a program. Among these are ratings of satisfaction with a program, from the people involved or concerned with it. Such data may give a useful method for judging whether a program is likely to meet with the community acceptance and cooperation it must have if it is to find a permanent place in the network of social services.

Another factor in the social validity of a program is the degree to which it can be implemented with the appropriate client group. Of 30 students initially selected for inclusion in the home-based reinforcement group, seven were subsequently found unsuitable for a variety of reasons. The children came from the entire socioeconomic range served by the local school board. In a community with a different socioeconomic structure, this rate of participation could be quite different.

The present program was devised after a similar local area program had been in effect for three years, to train and monitor parents as tutors. This program had not operated within the school system, but generated considerable interaction with school personnel. The success of this program (Bucher, Reaume, Mueller, & Dennis, 1978) helped make it possible to get support for the modification reported here. Experience with this program also guided the experimenters in their interactions with school personnel and with parents in implementing the home-based reinforcement programs.

Adoption of a proposed program is also facilitated if it can be shown that the program can be effectively implemented by existing social service agencies. The present program presents several defects in this respect. Researchers and research assistants fulfilled many important functions: rating reliability and intervening to retrain tutors when necessary; managing all home-based procedures; collecting daily results and pre- and posttest data; evaluating progress

through SRA materials and scheduling changes in the reading levels; and procuring and assigning the non-SRA reading materials. The framework in which the research was conducted provided contingencies to control the adequacy with which all these steps were carried out. For the program to have usefulness in the school system, it must be shown that agents in the schools can be trained to carry out the functions carried out here by the researchers, and that the school system can arrange contingencies to ensure that the work is properly done. Such a feasibility study is currently in progress.

REFERENCE NOTES

1. Mothersill, K., Hundert, J., Lai, L., & Willis, J. *A manual for training tutors in reading*. Unpublished manuscript, 1976.
2. Wilson, B. J. Component analysis of a home-based token reinforcement program for remedial reading. Unpublished doctoral thesis, University of Western Ontario, 1976.

REFERENCES

- Ayllon, T., & Kelly, K. Effects of reinforcement on standardized test performance. *Journal of Applied Behavior Analysis*, 1972, 5, 477-484.
- Bucher, B., Reaume, J., Mueller, H., & Dennis, S. A home-based tutoring programme for academic problem children: Implementation and evaluation. *Ontario Psychologist*, 1978, 10, 23-31.
- Camp, B. W., & Doornick, W. J. Assessment of motivated reading therapy with elementary school children. *Behavior Therapy*, 1971, 2, 214-222.
- Clingman, J., & Fowler, R. L. The effects of contingent and noncontingent rewards on the I.Q. scores of children of above average intelligence. *Journal of Applied Behavior Analysis*, 1975, 8, 90-91.
- Clingman, J., & Fowler, R. L. The effects of primary reward on the I.Q. performance of grade-school children as a function of I.Q. level. *Journal of Applied Behavior Analysis*, 1976, 9, 19-23.
- Coleman, R. A conditioning technique applicable to elementary school systems. *Journal of Applied Behavior Analysis*, 1970, 3, 293-297.
- Davis, M. Effects of having one remedial student tutor another remedial student. In G. Semb (Ed.), *Behavior analysis and education*, Lawrence: University of Kansas Press, 1972.
- Eldlund, C. V. The effect on the test behavior of children as reflected in the I.Q. scores when re-

- inforced for each correct response. *Journal of Applied Behavior Analysis*, 1972, **5**, 317-319.
- Fry, E. A readability formula that saves time. *Journal of Reading*, 1968, **11**, 513-516 and 577.
- Harris, V. M., Finfrock, S. R., Giles, D. K., Hart, B. M., & Tsose, P. C. The effects of performance contingencies on the assignment completion behavior of severely delinquent youths. In E. A. Ramp & G. Semb (Eds.) *Behavior Analysis: Areas of research and application*, Englewood Cliffs, N.J.: Prentice-Hall, 1975.
- Jenkins, J. R., Mayhall, U. R., Reschka, C. M., & Jenkins, L. M. Comparing small group and tutorial instruction in resource rooms. *Exceptional Children*, 1974, **40**, 245-250.
- Johnson, M., & Bailey, J. S. Cross-age tutoring: Fifth graders as arithmetic tutors for kindergarten children. *Journal of Applied Behavior Analysis*, 1974, **7**, 223-232.
- Kazdin, A. E., & Bootzin, R. R. The token economy: An evaluative review. *Journal of Applied Behavior Analysis*, 1972, **5**, 343-372.
- Lahey, B. B., McNees, M. P., & Brown, C. C. Modification of deficits in reading for comprehension. *Journal of Applied Behavior Analysis*, 1973, **6**, 475-480.
- Niedermeyer, F. C., & Ellis, P. Remedial reading instruction by trained pupil tutors. *Elementary School Journal*, 1971, **71**, 400-405.
- O'Connor, J. J., & Weiss, F. L. A brief discussion of the efficacy of raising standardized test scores by contingent reinforcement. *Journal of Applied Behavior Analysis*, 1974, **7**, 351-352.
- O'Leary, K. D., & Drabman, R. Token reinforcement programs in the classroom: A review. *Psychological Bulletin*, 1971, **75**, 379-398.
- Parker, D. H., & Scannell, G. *The SRA Reading Laboratory: Teacher's handbook* (Revised Edition). Chicago: Science Research Associates, 1973.
- Richardson, D. C., & Havlick, L. L. High school students as reading instructors. *Elementary School Journal*, 1975, **75**, 389-393.
- 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**, 169-177.
- Shaver, J. P., & Nuhn, D. The effectiveness of tutoring underachievers in reading and writing. *Journal of Educational Research*, 1971, **65**, 107-112.
- Spache, G. D. *Diagnostic Reading Scales*. Monterey, Calif.: McGraw-Hill, 1963.
- Staats, A. W., & Butterfield, W. H. Treatment of non-reading in a culturally deprived juvenile delinquent: An application of reinforcement principles. *Child Development*, 1965, **36**, 925-942.
- Staats, W. W., Minke, K. A., Goodwin, W., & Landeen, J. Cognitive behavior modification-motivated learning reading treatment with subprofessional therapy technicians. *Behaviour Research and Therapy*, 1967, **5**, 283-299.
- Willis, J., Crowder, J., & Morris, B. A. A behavioral approach to remedial reading using students as behavioral engineers. In G. Semb (Ed.), *Behavior analysis and education*. Lawrence: University of Kansas Press, 1972.
- Willis, J., Morris, B. A., & Crowder, J. A remedial reading technique for disabled readers that employs students as behavioral engineers. *Psychology in the Schools*, 1972, **9**, 67-70.
- 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.

Received November 2, 1978

Final acceptance June 18, 1979