JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1974, 7, 223-232 NUMBER 2 (SUMMER 1974)

# *CROSS-AGE TUTORING: FIFTH GRADERS AS ARITHMETIC TUTORS FOR KINDERGARTEN CHILDREN*[1]

MARTHA JOHNSON AND JON S. BAILEY

FLORIDA STATE UNIVERSITY

Five fifth-grade students tutored five kindergarten children in basic arithmetic skills for 7.5 weeks. A control group consisted of five kindergarten children who received no tutoring and were matched with the experimental group in arithmetic ability. Pre-, mid-, and posttesting was done using a skills-based arithmetic test. Results showed that the experimental group made far greater gains than the control group on a post-test comparison (matched pairs signed ranks test $p = 0.062$). In addition, a subanalysis of specific arithmetic skills showed they were improved only when tutoring for that skill was carried out. Systematic observations made of the tutor-student interactions indicated wide tutor-to-tutor variability in the percentage of student responses praised, and very little use of negative, disapproving statements. It was concluded that trained fifth-grade students can effectively teach basic arithmetic skills to kindergarteners.

Students can be trained to modify the behavior of other students in both classroom and residential settings, thus reducing the need to hire expensive specialized personnel such as speech therapists (Bailey, Timbers, Phillips, and Wolf, 1971) and behavior therapists (Surratt, Ulrich, and Hawkins, 1969). The current enthusiasm for the use of children as behavioral engineers coincides with an interest in the field of education in what Lippett and Lohman (1965) called "cross-age" tutoring. Many projects have been initiated in which older children are used to tutor younger children in a variety of subject matter areas. The rationale for these projects is much the same. It is assumed that: (1) the older child can provide the younger child with a model for appropriate behavior, (2) the older child can profit from the tutoring relationship, in that the responsibility given to him often serves indirectly to increase his own motivation and academic performance, and (3) the trained older tutor can effectively teach many skills that require more time and attention than the classroom teacher has available.

Although it is generally agreed that cross-age tutoring can be a valuable learning experience for both the tutor and "student", there are few empirical data to support the notion. Most tutorial projects have been evaluated in terms of the verbal reports of students and teachers involved, rather than in terms of objective data documenting increases in academic performance and changes in classroom behavior (Fleming, 1969; Geiser, 1969; Groff, 1967; Harris, 1971; Lippett and Lohman, 1965).

Pre-posttesting designs have also been used in evaluating cross-age tutorial projects. Cloward (1967) found that fourth and fifth graders tutored in reading by high school students for five months showed an average gain of six months in reading achievement, in contrast to a control group that showed only 3.5 months gain during the same period. Frager and Stern (1970) used a similar design and found that kindergarten children tutored by sixth graders made significant gains on a language readiness test over a control group that received no tutoring.

[1]This paper was based on a thesis submitted by the first author to the Florida State University in partial fulfillment of the requirements for the M.S. degree in psychology. Reprints may be obtained from Jon Bailey, Psychology Department, Florida State University, Tallahassee, Florida 32306.

Niedermeyer and Ellis (1971) is the only study reported in the literature in which systematic observational procedures were used to record tutor-student interactions in the tutoring situation. Support was found for the notion that tutors trained in specific instructional skills exhibit more appropriate tutoring behaviors (*e.g.*, praising, correcting incorrect responses) than untrained tutors. However, whether or not the trained fifth-grade tutors were more successful in improving the reading skills of their kindergarten students is not clear from the data.

The present study was carried out in a public elementary school to determine if fifth-grade students could effectively tutor kindergarten children in arithmetic skills requiring a great deal of repetition and reinforcement. The basic arithmetic tasks of counting, number recognition, and number naming were chosen because: (1) the authors have observed in a variety of settings a significant number of kindergarten children who cannot perform these tasks, (2) kindergarten teachers generally expect their pupils to be able to perform these tasks, and yet are unable to give the individual instruction often needed, (3) the tasks can be easily defined in behavioral terms, and (4) the tasks are well suited to the drill and review exercises that a student tutor can provide without a great deal of initial training. The overall effect of the tutorial program was evaluated by comparing pre- and posttest scores of the tutored students with a control group matched in arithmetic ability. Individual analyses were also made of the daily acquisition of arithmetic skills for each tutored subject. Finally, systematic observations were made of the tutor-student interactions in order to identify specific tutor behaviors that contribute to the success of a tutorial program.

## METHOD

### *Subjects*

Tutors were five fifth-grade students recommended by their teachers on the basis of the following criteria: (1) academic performance in the top 15% of their class, (2) "conscientiousness and dependability", as determined by teacher judgement, and (3) desire to be a tutor. The tutors ranged in age from 10 yr, four months to 11 yr, 3 months.

Ten kindergarten children participated in the study, five as experimental subjects and five as control subjects. The kindergarten teachers initially identified children from three classrooms who could not count objects up to 10 and who could not name the numerals from one to 10. Two groups of five subjects each were then matched in arithmetic ability on the basis of an arithmetic test designed by the authors. Groups were also matched as closely as possible in terms of age and sex, with the mean age of both groups being 5 yr, nine months. The groups were then randomly designated as either experimental or control.

### *Testing and Tutorial Materials*

A skills-based arithmetic test was constructed and administered by the experimenter.[2] The arithmetic test was used as a pre-post measure and was also given twice at approximately two-week intervals during the tutorial program (midtest$_1$ and midtest$_2$). Feedback as to correctness of response was not given during the testing. Only praise, noncontingent upon any particular test response was given; *e.g.*, "You're working hard today, Jimmy."

The arithmetic test was designed to measure the skills involved in "knowing" the numbers 0 to 100. It was hypothesized that a child must have at least four skills in order to conclude that he "knows" the numbers. He must be able to: (1) count aloud, (2) count objects, (3) recognize or point to a given numeral in a group of numerals when told the name, and (4) supply the name of a given numeral. It was expected that the kindergarten child would be able to perform some of the above mentioned tasks before others. For example, counting aloud would probably occur before counting objects,

[2]A copy of this test is available from the authors.

since counting aloud involves only an intraverbal chain (saying "one" is a stimulus for saying "two", *etc.*), while counting objects involves the additional skills of one-to-one correspondence. It was also expected that some numbers from 0 to 100 would be easier to learn than others.

Thus, both skills and numbers were broken down into categories, yielding a number-by-skill matrix, as shown in Figure 1. Each cell in the matrix defines one major task used in both the arithmetic test and in the tutorial program (the shaded cells were not taught). By reading across the rows of the matrix, one can see the sequence in which the tasks were taught. Thus, the student was first taught to count aloud from 0 to 10, then count objects from 0 to 10, then point to 0 to 10 when named, then name 0 to 10, and so on (tasks were always introduced in this order, although it was possible for a given subject to be working on more than one task at a time or to skip a task in the sequence, if he indicated mastery of that task on the pretest).

The arithmetic test consisted of items taken directly from the number-by-skill matrix. Students were tested sequentially over each task, and were given every item within each task, with the exception of tasks 13 and 14. For these tasks, a sampling procedure was used such that students were asked to point to a sample of eight

| NUMBERS \ SKILLS | COUNT ALOUD | COUNT OBJECTS | RECOGNIZE | NAME |
|---|---|---|---|---|
| 0 - 10 | 1 | 2 | 3 | 4 |
| 11 - 20 | 5 | 6 | 7 | 8 |
| (10, 20, 30 ... 100) | 9 | | 10 | 11 |
| 21, - 99 | 12 | | 13 | 14 |

Fig. 1. Number-by-skill matrix, with the numbered cells showing the sequence of tasks used in the tutorial program and the arithmetic test.

numbers drawn from the pool of numbers 21 to 99 and to name a sample of eight numbers drawn from that pool (with the stipulation that one number from each decade be sampled for each task). A total of 122 items were on the arithmetic test, and the first response to each item was scored either as correct or incorrect.

The tutoring materials consisted of a set of flashcards, which were made by drawing or writing each item on a 5 by 8 in. (12.5 by 20 cm) index card.

### *Tutor Training*

Three 30-min sessions on consecutive days were conducted by the experimenter to train the tutors. The first training session was primarily introductory, with emphasis on describing the tutoring program and modelling appropriate teaching behaviors for the trainees to imitate (*e.g.*, speaking clearly, praising appropriate academic and social behaviors, ignoring inappropriate social behavior, correcting an incorrect response, and repeating a stimulus after correcting an incorrect response). During the second session, appropriate teaching behaviors were reviewed and the tutoring materials and data recording sheets were introduced. A role-playing procedure was used, with one trainee playing the role of "tutor" and another the role of "student". Each trainee was given at least one opportunity to practise the skills involved in handling the materials, interacting with the student, and recording the data. For each task in the tutorial program, the tutors were taught specific verbal instructions to be given to the student (*e.g.*, "How many squares are there? Point to each square as you count them aloud.") and specific criterion responses indicating mastery. The third training session was entirely devoted to the role-playing procedure, with special emphasis on giving social reinforcement contingent upon appropriate academic and social behavior of the student. Throughout the modelling and role-playing sessions, the experimenter gave social approval to the trainees, contingent upon appropriate tutoring behavior.

### *Classroom Treatment*

The kindergarten teachers were instructed to include the experimental and control subjects in whatever arithmetic instruction was normally given in the classroom. Normal arithmetic instruction generally consisted of a 3-min group drill twice a week. The group instruction usually focused on naming the numbers 0 to 10 and counting aloud to 10, and did not differ significantly from class to class. The kindergarten teachers were asked to make certain that none of the subjects received any additional arithmetic instruction during the study. The control group had no contact with tutors and remained in the classroom throughout the study, except for testing.

### *Tutoring Procedures*

When tutor training was completed, each tutor was randomly assigned to work with one of the experimental subjects. Twenty-minute tutoring sessions were held daily in a school conference room, followed by a 10-min reinforcement or "activity" time. Each tutor-student pair attended daily tutoring sessions until they had participated in a total of 26 sessions (approximately 8.7 hr in actual tutoring time). Because of absences, all pairs did not attend each scheduled session and it was necessary to run 38 sessions (spaced over 7.5 weeks) to ensure that all pairs had their 26-session total.

Each tutor was expected to go to the kindergarten classroom at the prescribed time, pick up his "student", and report to the conference room for the session. The experimenter handed out materials for the session and gave each tutor a brief review of what he was to do that day. During the session, the experimenter was available to answer any questions that arose and occasionally gave social reinforcement to both the tutors and the students.

Tutors spent the major portion of the tutoring session teaching specific tasks to the student. The number of tasks that any one tutor-student pair worked on during any one session varied, depending upon how near the student was to mastery of the tasks.

Approximately 3 min at the end of each session were used for "review test"; *i.e.,* the tutor simply went over all of the tasks worked on that day and recorded the student's responses. The tutor then scored the tests and graphed the data, giving the student immediate feedback as to his daily progress. When a student received a perfect score on any task for three consecutive review tests, teaching of the task was discontinued and a new task was introduced.

Upon completion of the daily review tests, students were allowed to choose and engage in a play activity for 10 min. These activities included games such as Candyland and Bingo, or coloring, and were led by the tutors. Activity time was not made contingent upon attainment of a particular academic goal, only upon appropriate social behavior, defined simply as "cooperation and willingness to work with tutor without disruptive behavior". Social behavior was not systematically measured; however, the experimenter was prepared to return a student to his classroom should inappropriate social behavior occur. There were no such occurrences.

### *Observations of Tutors*

Observations were made by trained undergraduates using a modified version of the Tutor Observation Scale (Niedermeyer, 1970). Continuous observations, at least 3 min in length, were made daily of as many tutoring pairs as possible. The observers sat at the side of the conference room approximately 10 ft (3 m) away; they did not interact with the tutors or students during the sessions. Although the tutors were aware of observers in the room, they were not aware that they were being observed at specific times by the primary and/or reliability observer.

Definitions of the behavior categories used in observation were as follows: (1) *student gives correct response*—student responds correctly to stimulus, or corrects himself spontaneously upon first presentation, (2) *tutor gives* *praise*—tutor makes a positive approving statement to student following a correct response, (3) *student gives incorrect response*—student responds incorrectly to stimulus upon first presentation, (4) *tutor corrects student*—tutor tells student correct response after student answers incorrectly, (5) *tutor repeats stimulus*—after correcting an incorrect response, tutor presents same stimulus again, (6) *tutor prompts*—when student does not respond or gives incorrect response, tutor gives hint or clue as to the correct response, and (7) *tutor gives negative*—tutor verbally disapproves of student's behavior (the word "no" was not considered a negative comment when it was used only as feedback for an incorrect response).

Reliability checks were made periodically by having two observers make simultaneous and independent observation records of the same tutor. Each presentation of a new stimulus defined one trial. Reliability was calculated by dividing the number of stimulus presentations or trials in which there was agreement between two observers that a particular behavior occurred by the total number of trials observed. Agreements that a particular behavior did not occur were not included in the analysis.

A total of 20 reliability checks were made throughout the study. Reliability checks ranged from 3 to 17 min in length, with a mean length of 9 min. At least three reliability checks were made on each tutor at some time during the study. The mean reliability for each behavior and tutor is presented in parentheses in Table 1.

## RESULTS

### *Group Data*

Figure 2 shows the results of the pre-, mid-, and posttesting on the arithmetic test for the control and experimental groups. It can be seen that both the control and experimental groups responded correctly to 26% of the items on the pretest, and therefore can be considered equivalent. One control subject (S-10) transferred schools after his first midtest. Where applicable, his data are included in the figures, but his data and the data for his matched pair are not included in the statistical analyses; *i.e.*, $N = 4$ for the Wilcoxon test. After approximately two weeks of tutoring, the experimental group made its largest gain to 48% correct ($midtest_1$) and then steadily increased to 57% correct on $midtest_2$ and 66% correct on the posttest. In contrast, the control group increased from 26% correct on the pretest to 28% correct on $midtest_1$, 31% correct on $midtest_2$, and 38% correct on the posttest. A posttest comparison between the experimental and control groups using a Wilcoxon matched pairs signed ranks test (Kraft and van Eden, 1968; Siegel, 1956) yielded an exact probability of $p = 0.062$.

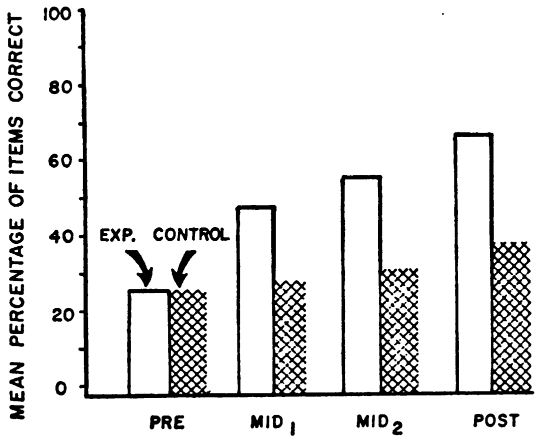


Fig. 2. Mean percentage of items correct on the arithmetic tests for the experimental group (empty bars) and the control group (shaded bars).

The pre- and posttest data on the arithmetic test for the two groups are further broken down according to specific arithmetic skills in Figure 3, so that gains made in each skill can be analyzed. It is interesting to note that both groups made gains in recognizing and naming the numbers 0 to 10 (these tasks were presumably taught in the kindergarten classroom). The experimental group made its largest gains over the control group on the counting, naming, and recognizing skills involving the numbers 11 to 20 and the decades.

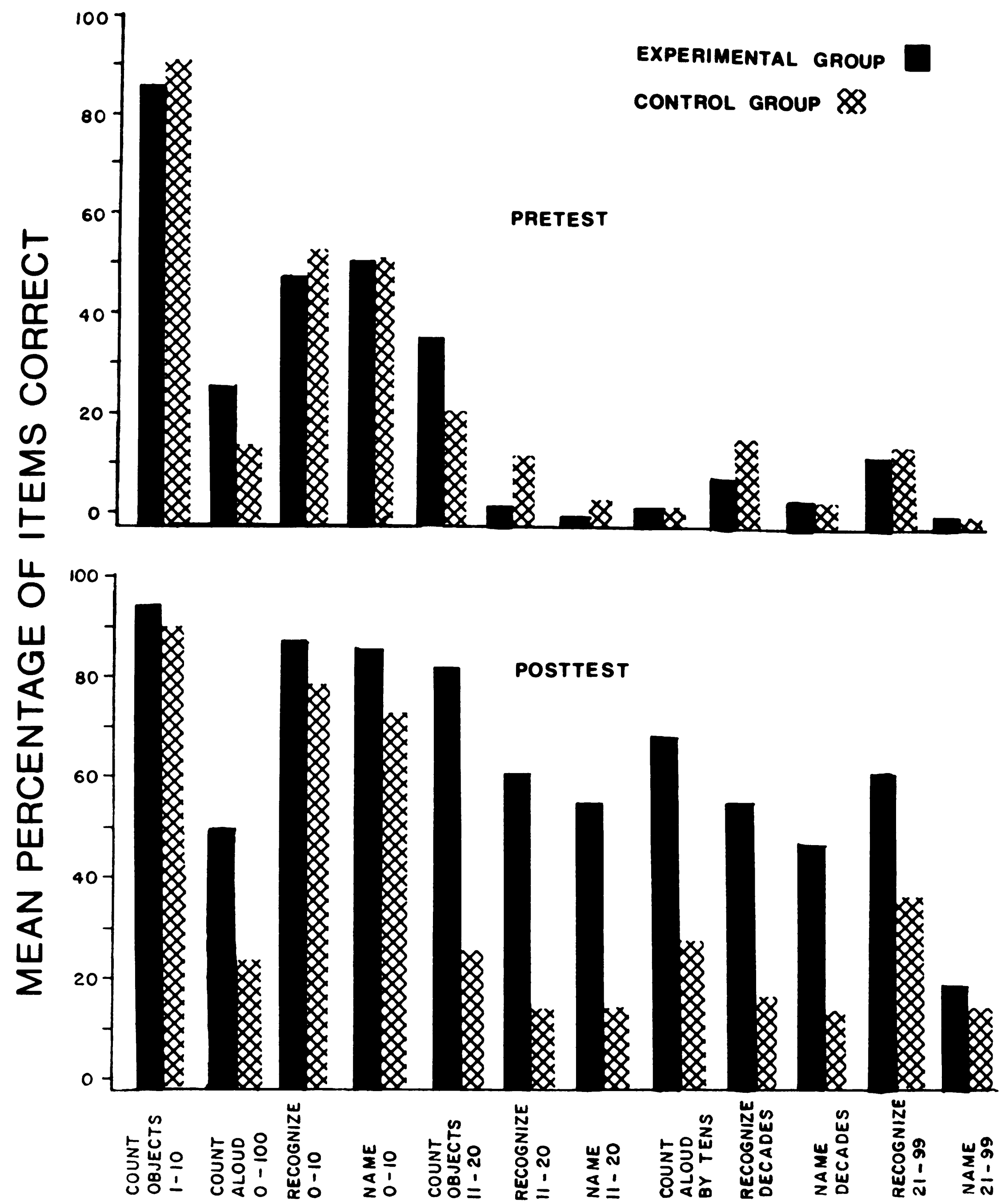


Fig. 3. A comparison of experimental (solid bar) *versus* control group (cross-hatched bar) scores by task on the arithmetic pre- and posttest.

### *Individual Data*

The specific effects of tutoring can be seen by analyzing the individual data shown in Figures 4 to 6. The data plotted across tutoring sessions are the number of items correct for a given subject on a given task on the daily review tests administered by the tutors. Open circles indicate the student scores on pre- and posttests as well as midtests (marked with the letter *m*).

As can be seen in Figure 4, Subject 1, for example, scored zero on the pretest for counting objects 11 to 20; in five tutoring sessions he mastered all 10 items and received a perfect score on the first midtest a few days later. Twenty sessions later, he got nine of 10 correct on the second midtest and eight of 10 on the posttest. This same pattern can be seen for each of the other skills that were tutored, recognizing and naming 11 to 20, counting by tens, and recognizing and naming decades.

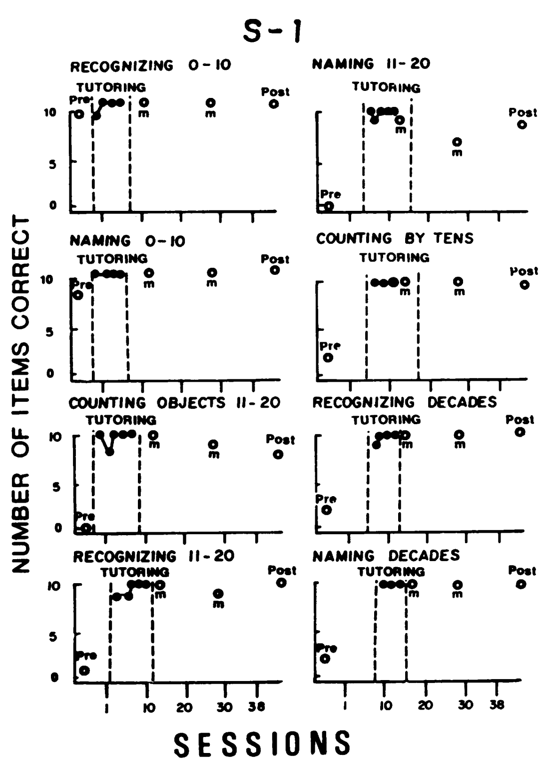


Fig. 4. Number of items correct on daily review tests (solid dots) on each of eight arithmetic skills as a function of tutoring for Subject 1. Open circles indicate pre- and posttest scores, as well as midtest scores (marked with the letter *m*).

Subject 2 (Figure 5) was tutored on only three tasks, and slow but steady improvements can be seen on the review tests given at the end of each tutoring session. Of the five skills not tutored, almost no gains are observed, as shown by mid- and posttest scores (counting objects 11 to 20 was an exception).

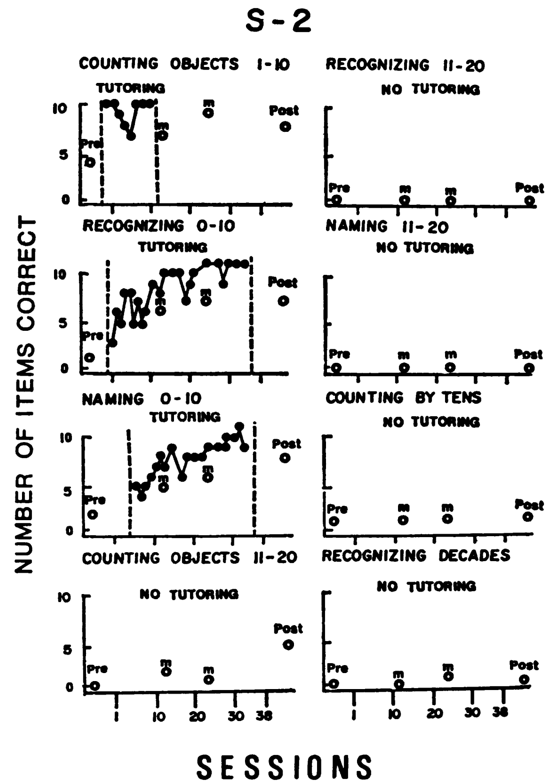


Fig. 5. Number of items correct on daily review tests (solid dots) on each of eight arithmetic skills as a function of tutoring for Subject 2. Open circles indicate pre- and posttest scores, as well as midtest scores (marked with the letter *m*). No tutoring was carried out on counting, recognizing, and naming 11 to 20, counting by 10s, and recognizing decades.

In Figure 6, the data for Subject 3 show a similar effect that resembles a multiple baseline design; when tutoring was applied across skills at different points in time, those skills improved. No gains were observed before tutoring on a specific skill.

### *Tutor Observations*

In order to give an accurate account of the ways in which the tutors interacted with the students, observations were made of aspects of the tutoring situation that may be critical to learning. A summary of observed behaviors for each tutor is presented in Table 1.

It can be seen that Tutor 2 and Tutor 4 praised a noticeably smaller percentage of correct student responses than did the other tutors.

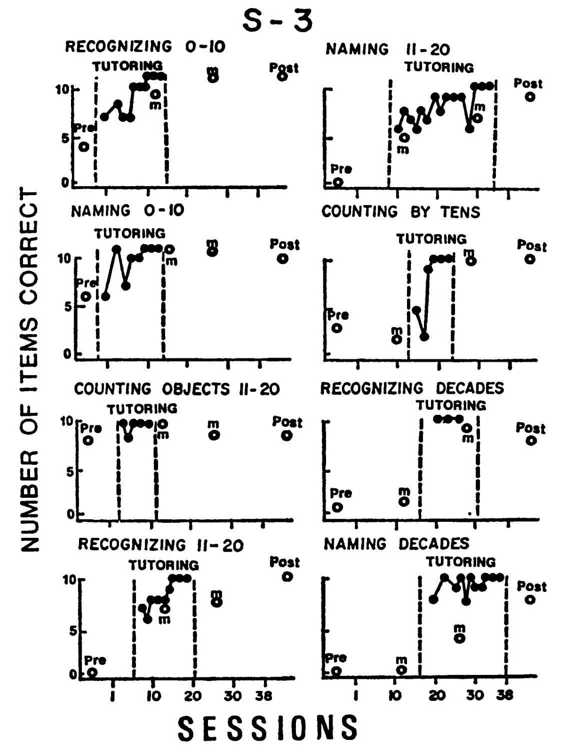


Fig. 6. Number of items correct on daily review tests (solid dots) on each of eight arithmetic skills as a function of tutoring for Subject 3. Open circles indicate pre- and posttest scores, as well as midtest scores (marked with the letter *m*).

The relationship between praise given by the tutor and student success can be seen by examining the change scores of the student. Thus, Subject 2 and Subject 4, who made the least progress on the arithmetic test, also received the least praise. In addition, the tutor of Subject 4 prompted a noticeably larger number of responses than did the other tutors.

The categories of tutor-corrects-pupil and tutor-repeats-stimulus were included in the observations on the assumption that a good tutor would correct an incorrect response and repeat the stimulus in an attempt to elicit the correct response from the pupil. Table 1 shows that Tutors 2, 3, and 5 corrected a relatively large percentage of incorrect responses, but often failed to repeat the stimulus before going on to the next item.

## DISCUSSION

The present study clearly illustrates that elementary school students can be trained with relative ease to perform routine teaching duties normally assumed by a professional classroom teacher and that younger children can profit academically from the tutoring of older children. The tutored kindergarteners, as a group, responded correctly to 26% of the items on the arithmetic pretest and 66% of the items on the posttest, an increase of 40 percentage points; in contrast, the control group increased only 12 percentage points, from 26% to 38%. The data indicated that the students did, in fact, master many basic arithmetic skills after only 26 twenty-minute tutoring sessions, spaced over a period of 7.5 weeks.

The research design did not employ a contact control group; *i.e.,* kindergarten children who

Table 1

Relationship of observed tutor behavior to student improvement on the arithmetic test (reliabilities for each behavior and tutor are shown in parentheses).

| *Tutor* | *% Correct Responses Praised* | *% Responses Prompted* | *% Incorrect Responses Corrected* | *% Incorrect Responses Corrected and Repeated* | *Student Change Score*[a] |
|---|---|---|---|---|---|
| 1 | 77 (r = 0.85) | 2 (r = 1.00) | 65 (r = 0.51) | 30 (r = 1.00) | 73 |
| 2 | 17 (r = 0.82) | 3 (r = 1.00) | 78 (r = 0.88) | 61 (r = 0.82) | 25 |
| 3 | 54 (r = 0.86) | 2 (r = 1.00) | 96 (r = 0.90) | 69 (r = 0.81) | 62 |
| 4 | 15 (r = 0.92) | 10 (r = 0.83) | 44 (r = 0.61) | 24 (r = 0.79) | 35 |
| 5 | 83 (r = 0.87) | 3 (r = 1.00) | 79 (r = 0.78) | 39 (r = 0.54) | 47 |

[a]The student change score represents an absolute increase in points from pre- to posttest on the arithmetic test.

were taken out of class and given some individual attention. It appears highly unlikely that the *response-specific* changes seen in this study could possibly have resulted from individual attention alone, however. That is, one might expect some generalized change in behavior (*e.g.*, more verbal expression, greater cooperation, *etc.*) to result from the individual attention alone; but the immediate and specific improvements in arithmetic skills that were tutored would not be expected to result from such an unstructured interaction.

Within the experimental group, the research design resembles a multiple baseline across behavior design, since tutoring on the different skills began at different points in time. This permits a more detailed analysis of the functional relationship between tutoring than the overall pre-post design to compare experimental *versus* control group effects.

This analysis also permits a close examination of the acquisition of arithmetic skills by students, and a wide range of differences was seen even for the small sample of students involved. One subject mastered each skill almost immediately with the onset of tutoring (S1) and was tutored on all eight tasks, while another (S2) required many more repetitions to gain mastery and ended up being tutored on only three of the eight skills. The observational data indicated early in the tutorial program that the tutor of Subject 2 was giving very little praise. It might have been appropriate at that time to give the tutor additional formal training in the use of praise or to switch the tutor-student pairs, so that a functional analysis could have been made of the effect of praise on the progress of Subject 2.

The present study explored an area relatively new to the field of applied behavior analysis, and several procedural inadequacies were noted. The observational data indicated that Tutor 2 and Tutor 4 praised less than 20% of their students' correct responses and that Tutor 4 prompted a relatively high percentage of responses (10%). These observations suggest that the tutor-training procedures used were too unstructured for some tutors. One improvement might be to require that the "trainee" participate in the role-playing activities until he can exhibit a desired percentage of appropriate tutoring behaviors in a criterion situation.

One general conclusion can be drawn from the observations. Although the amount of praise given by the tutors varied from day to day and tutor to tutor, it was very high in comparison to the number of negatives given by the tutors. Only 16 negative comments were recorded during the 61 observations made of the tutor-student interactions. As a whole, participation in the tutorial program appeared to be a positive and rewarding experience for both the tutors and students.

The present study, although small in scope, established cross-age tutoring as a viable approach to increasing the academic performance of young children. The number of academic failures and behavior problems in our schools today attest to the need for supervised tutors in all aspects of education. At a time when many school systems are training paraprofessionals and aides to assist teachers in various classroom duties, cross-age tutoring offers an opportunity to utilize untapped resources already in the schools to assist in the teaching process.

## REFERENCES


Bailey, J. S., Timbers, G. D., Phillips, E. L., and Wolf, M. M. Modification of articulation errors of pre-delinquents by their peers. *Journal of Applied Behavior Analysis,* 1971, **4,** 265-281.

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

Fleming, J. C. Pupil tutors and tutees learn together. *Today's Education,* 1969, **58,** 22-24.

Frager, S. and Stern, C. Learning by teaching. *The Reading Teacher,* 1970, **23,** 403.

Geiser, R. L. Some of our worst students teach! *Catholic School Journal,* 1969, **69,** 18-20.

Groff, P. Should you use pupil tutors? *Instructor,* 1967, **77,** 35.

Harris, M. M. Learning by tutoring others. *Today's Education,* 1971, **60,** 48-49.

Kraft, C. H. and van Eden, C. *A nonparametric introduction to statistics.* New York: Macmillan Company, 1968.

Lippett, P. and Lohman, J. E. Cross-age relationships: an educational resource. *Children,* 1965, **12,** 113-117.

Niedermeyer, F. C. Effects of training on the instructional behaviors of student tutors. *Journal of Educational Research,* 1970, **64,** 119-123.

Niedermeyer, F. C. and Ellis, P. A. Remedial reading instruction by trained pupil tutors. *The Elementary School Journal,* 1971, **71,** 400-405.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Surratt, P. R., Ulrich, R. E., and Hawkins, R. P. An elementary student as a behavioral engineer. *Journal of Applied Behavior Analysis,* 1969, **2,** 85-92.

*Received 9 August 1973.*
*(Revision requested 3 December 1973.)*
*(Final acceptance 30 January 1974.)*