AN EXAMINATION OF THE LITERATURE ON TUTORING

Carol Taylor Fitz-Gibbon

CSE Report No. 117
July 1978

Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

Following these activities, an interim report was submitted to nine reviewers selected by NIE. Committee members and reviewers subsequently met together for two days of round-table discussions.

The production of a single final report would have been an inappropriate way of presenting the body of information that project staff had at hand following these activities. For example, much information that would be of interest to teachers and parents planning or running tutoring projects would not be of interest to policy makers and vice versa. Furthermore, although the position had been reached that one particular kind of tutoring project--the Learning-Tutoring Cycle (LTC)--should be recommended for widespread implementation in Title I projects, there was much to be said about the planning of tutoring projects in general and about highly innovative projects which, although not as immediately feasible as the Learning-Tutoring Cycle, could have greater impact if implemented.

In view of this need to speak to various audiences about diverse topics such as theories, research, policy, practical planning and innovative ideas, six separately bound volumes were prepared:

- Report A. The Learning-Tutoring Cycle: Overview
- CSE Report No. 122. Setting Up and Evaluating Tutoring Projects (formerly Report #1)
- CSE Report No. 118. A Survey of Tutoring Projects (formerly Report #3)
- CSE Report No. 121. Tutoring: Some New Ideas (formerly Report #4)
- CSE Report No. 117. An Examination of the Literature on Tutoring (formerly Report #5)
- CSE Report No. 116. Tutoring and Social Psychology: A Theoretical Analysis (formerly Report #6)

These reports provide an information base and a rationale for actions at both federal and local levels.

Three reports--the Survey, the Literature report, and the Social Psychology report--bring together information from a wide range of sources to provide background knowledge concerning current practice, the perceptions and recommendations of practitioners,

ACKNOWLEDGEMENTS

This study has benefitted immeasurably from the thoughtful contributions of many individuals. Members of the committee represented diverse backgrounds and professional experiences, and their deliberations were invariably stimulating and challenging.

Members of the Committee

- Harry F. Silberman (Co-Principal Investigator)
  Professor of Education
  UCLA

- Marvin C. Alkin
  Professor of Education
  UCLA

- Ron Apperson
  Legal Adviser
  Los Angeles Unified School District

- Dorothy Bradshaw
  General Secondary Curriculum Consultant
  Compton Unified School District

- Erick Lindman
  Professor of Education
  UCLA

- Neil Malamuth
  Professor of Psychology
  UCLA

- Gerald Newmark
  Co-Director
  Educational Communications Corp.

- Floraline Stevens
  Coordinator, Research and Evaluation
  Los Angeles Unified School District

- Richards Williams
  Research Associate
  Center for the Study of Evaluation

CHAPTER I. INTRODUCTION

Since there are a number of literature reviews available for the topic of cross-age tutoring (Devin-Sheehan, Feldman and Allen, 1976; Bloom, 1976; Klaus, 1975; Rosenshine and Furst, 1969), it is important to specify the distinctive features of the present analysis. In this analysis a few studies have been selected for examination in detail. The level of detail is such that when the conclusions of the original authors are rejected or questioned (e.g., as in Hamblin and Hamblin, 1972; Rosenshine and Furst, 1969, and Niedermeyer and Ellis, 1971), readers will find sufficient information in this report to form their own judgments concerning the critique offered.

The fairly detailed examination of selected studies also serves to illustrate various points concerning both tutoring as an instructional procedure and concerning the methodology of research and evaluation. The description of a variety of projects and experiments should prove helpful to school district personnel and instructional staff planning studies of their own tutoring projects.

A Classification Scheme for Tutoring Projects

What might one expect to find in a school reported to have a "tutoring project"? A little experience reveals that a wide variety of organizational changes may have occurred under the general rubric of a "tutoring project." In some projects two or three older students may work in an elementary classroom,

(or meet tutees' other needs if the project is not one concerned with learning in the academic sense). In LbT projects the needs of tutors are diagnosed, and the tutoring task will be defined by those needs. Then tutees are located for whom the instruction (or other activities of the task) will be suitable.

Although it is not an established finding, there is a good deal of anecdotal evidence to suggest that tutors will lose interest if their tutees are not clearly benefitting from the tutors' activities. Thus in LbT projects, it is probably vital to ensure that, although the project was designed to enable tutors to learn by tutoring, tutees receive appropriate instructions and benefit from it. Thus LbT projects require that both tutors and tutees benefit. This is ensured by the selection of appropriate tutees rather than be adjustment in the tutoring task. In Appendix A, the classification of tutoring projects shown in Figure 1 is contrasted with two other classifications that have been proposed.

The Organization of This Report

In Figure 2, the studies that are discussed in most detail in this report are listed along with a number of their basic characteristics: kind of project, size and duration, relationship to the regular school program, and population characteristics. Since the studies are listed in the order in which they are considered in the report, it can be seen from Figure 2 that the report begins with a description and analysis of several experiments run for short periods of time by persons external to the schools. These studies have the virtue of data collected in the framework of a good experimental design involving control groups formed by random assignment.

The next set of studies examined ("field studies") includes only projects run by school personnel. These studies generally lack readily interpretable

so many of the following pages will be concerned with cognitive outcomes, an elegant quotation from Jerome Bruner is presented here to conclude this introduction and to redress the balance a little:

> ...when we read Urie Bronfenbrenner's perceptive analysis of the isolating tendency within American society to grade and segregate by age, one realizes that the idea of each teaching another may indeed be a revolutionary step toward maintaining community in a society where the forces of urban organization, of mass production, and indeed of mass education are all centrifugal (p. 62).
>
> The contemporary effort to reconstruct our schools is, I think, part of the general effort to reduce isolation, to re-establish mutuality and exchange. Sharing one's skills and knowledge with others, being teacher as well as learner are efforts to that end.
>
> Perhaps superintendents, principals, and school boards may have to be lured with pragmatic rationale. It is not the right reason for adopting this plan. It is a bonus. The real reason should be to get us an inch on the way toward making the helper and the helped the universal unit of exchange within a culture that continues to produce lonely crowds, lonelier than ever (p. 63).

CHAPTER II. COGNITIVE OUTCOMES IN EXPERIMENTS AND FIELD STUDIES

Experiments in School Settings

Experiments often introduce an artificiality into a school situation: "outsiders" intrude and help to run the experiment, many more tests than usual are given, questionnaires must be answered, and the entire flurry of activity is a short-term transitory phenomenon. It is difficult to assess the extent to which these exigencies affect the outcome of an experiment conducted in a school setting.

However, although experiments introduce an artificiality that makes generalizations from the results a hazardous procedure, they also allow for controls to be established so that, within their context, strong conclusions can be drawn. If experimental results and field study results point in the same direction, confidence in both is considerably increased.

Being a Tutor vs. Remaining in Class

Fitz-Gibbon, 1975. In one experiment that was conducted in a school, a single Learning-Tutoring Cycle* was implemented for 40 ninth grade students randomly selected from four low achieving "general math" classes in an inner-city junior high school, population approximately 90 percent Black and 10 percent Chicano. Tutees were 68 fourth grade students randomly selected from

*The Learning Tutoring Cycle is a Learning-by-Tutoring project which is a primary focus of the CSE Reports on Tutoring. See Preface.

Means and 95 percent confidence limits for ninth graders on the immediate posttest are displayed in Figure 3.

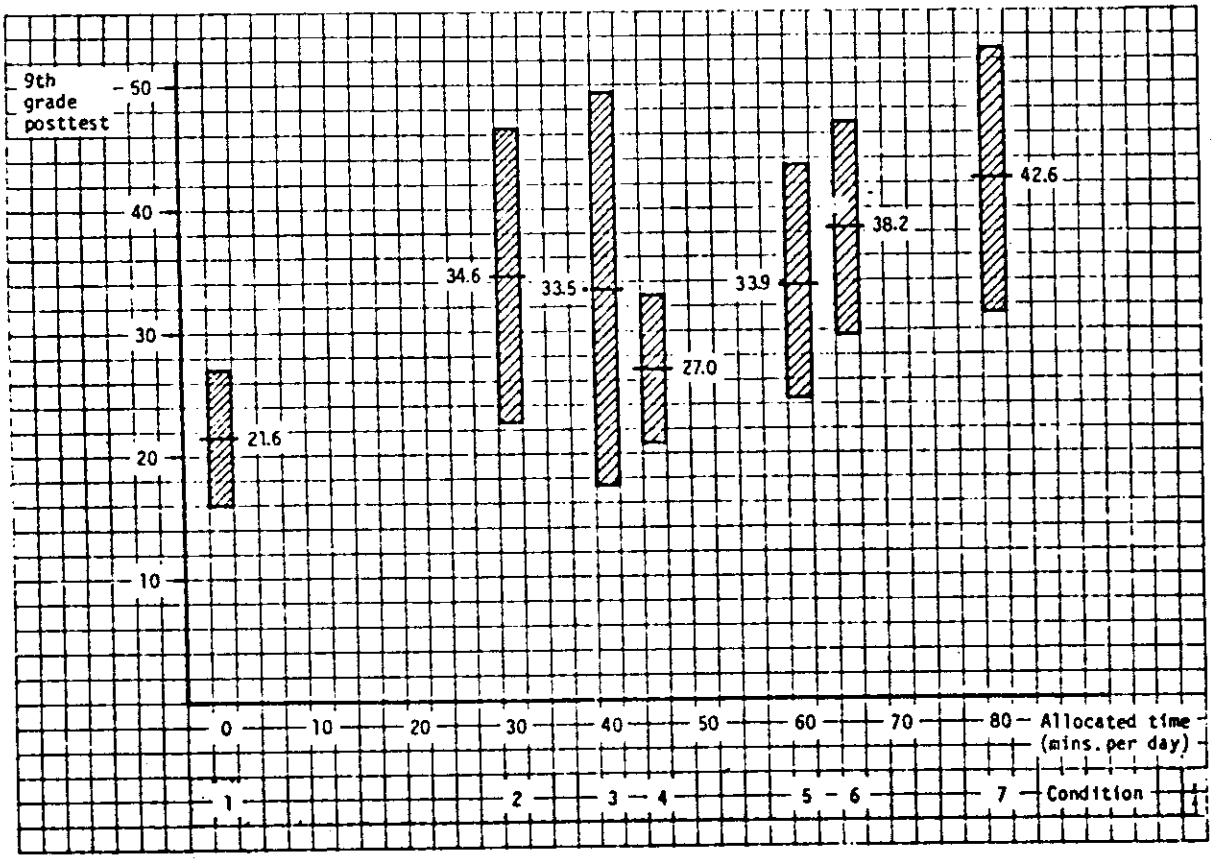

Figure 3. Ninth grade posttest results: posttest means and 95% confidence limits for ninth graders in the seven conditions.

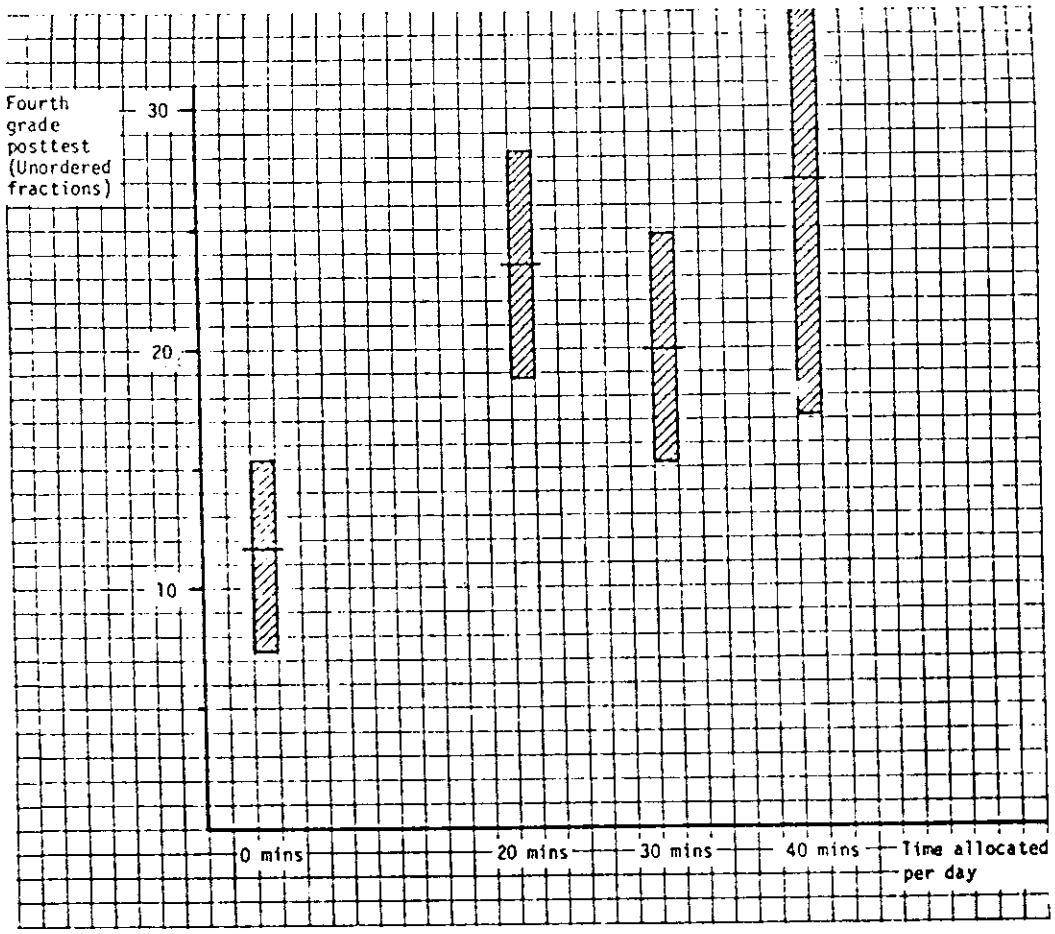

Figure 4. Fourth grade posttest results: Posttest means and 95% confidence limits for fourth graders allocated zero, 20, 30, or 40 minutes of tutoring daily.

Ninth graders were administered the Scrambled Fractions test at posttest time in addition to the ninth grade posttest. This permitted a direct comparison of ninth and fourth graders who had and had not participated in the tutoring project. The distribution of scores on the Scrambled Fractions test is shown in Figure 5. In three weeks of tutoring the ninth graders had moved the fourth grade score distribution for tutees close to that of the ninth grade control group students.

(-.14) posttest (-.15) and retention test (-.28), although only the latter correlation was significant at the .05 level. There was a significant negative correlation between average tutee residual gains and tutors' effort ratings ($r = -.30$, $p = .03$, $n = 39$). These trends indicated that low effort, low achieving, students were slightly more effective as teachers than were high effort, high achieving, students. The tendency was not a strong one but was sufficient to refute any suggestion of a tendency in the opposite direction for this group of students, a group, it must be remembered, of generally low achievement. These findings were not inconsistent with the observations of many practitioners: poor students are frequently relatively effective tutors.

How can persons less competent in the subject be as effective or more effective when acting as tutors? In considering answers it must first be recalled that tutoring is not teaching. Tutoring means explaining a limited, usually prescribed, set of concepts and very often it is performed following training. Had tutors the responsibility for selecting curriculum, analyzing and organizing it into clear objectives, then one might expect subject matter competence to become important if not crucial to success. But in the context of clearly delimited instruction, and with training, low achievers can make effective tutors.

(An indication of one way in which child tutors may be effective is obtained from some reports by Stallings and Kaskowitz (1974) and Soar (1972). They presented evidence from large-scale analyses that open-ended questions, i.e., questions high in the Bloom taxonomy (Bloom, 1956), are negatively related to achievement. Direct questions at lower levels of the Bloom taxonomy facilitate acquisition of knowledge by students. Lower achieving or less competent tutors may be likely to ask simple direct questions rather than indirect, abstract, reasoning-type questions.)

Being a Tutor vs. Studying Alone

Allen and Feldman (1973). The hypothesis that low-achieving children will learn more when placed in the role of teacher than when spending the same amount of time studying alone was tested by these researchers.

Tutors were 10 low-achieving fifth graders and tutees were 10 third graders (small numbers, but it was a repeated measures design). All children were paid to participate in the experiment during the summer vacation. Tutors and tutees were sex-matched.

The tasks were 10 different lessons dealing with science, language, and reading stories. The experiment ran a fortnight with tutoring taking place each weekday. Each day the fifth graders studied the lesson for eight minutes and then either continued to study alone for the next 20 minutes or spent the same amount of time tutoring. The "study alone" and "tutoring" conditions were assigned on alternate days to each fifth grader. Thus each day half the subjects were tutoring and half studying alone. The third graders were assigned on alternate days to the tutee role or to study alone. An objective test on each lesson was given at the end of each session.

Test scores were converted to z-scores and the results are shown in Figures 6 and 7 , reproduced from the article. (The graphs are based on only seven subjects due to experimental mortality: one tutor did not attend all sessions and two did not follow instructions.)

Analysis of variance for repeated measures indicated that for the younger children there was an interaction between condition and session significant at the .10 level but no significant differences in orthogonal comparisons.

For the older children (tutors) the interaction between condition (tutor or study alone) and session (occasion) was significant at the level of $p < .08$, and there were significant differences at the .05 level in favor of tutoring at sessions 4 and 5. Thus, as children became used to tutoring, their performance improved relative to that obtained under the study alone condition.

The authors concluded that the results supported the hypothesis that low-achieving children will learn more when assigned a tutor role than when spending the same time studying alone.

Possible Processes Involved in Promoting Learning in Tutors

The two experiments just described implied that acting as a tutor led to more learning than studying alone (Allen & Feldman, 1973) or working in class (Fitz-Gibbon, 1975). What processes may account for such effects? Allen and Feldman posed the following problem:

> Of considerable theoretical importance is the determination of the locus of the enhanced performance due to tutoring. As one possibility, perhaps the mere expectancy of teaching someone else is sufficient to produce increased learning--thus, the tutoring situation itself might be of only secondary importance. The expectation of tutoring could have resulted in the child's exerting greater effort and working harder on the material while preparing the lesson (p. 4).

In other words, does the act of tutoring, in itself, promote learning or is the beneficial impact of the tutor role accounted for solely by the initial motivation to learn the work in order to teach it?

Allen and Feldman indicated that this is a question of "theoretical" importance. Not so; it has immediate policy-related implications: if the motivation to learn rather than the process of tutoring is the major source of influence on cognitive gains, then presumably other kinds of motivation could replace tutoring and be equally effective. Motivation might be induced by attractive and meaningful instructional materials, livelier teachers, or material rewards for example. If, however, there is something about the activity of tutoring that produces cognitive gains, then one must insist that this activity occur, and reject as unpromising a search for other forms of motivation to produce equivalent gains.

that by verbalizing, one is alerted to errors in one's own reasoning or to aspects of a problem that would otherwise pass unnoticed. Many persons report the experience of struggling with a problem, verbalizing it by asking someone's help, and then immediately seeing the solution without help; posing the question sparked the solution. Conversely, one may believe one knows some concepts thoroughly. Someone asks for an explanation and, after a confident start explaining the work, one notices a small non sequitur which raises more issues and leads to a realization of a new aspect of the problem, one yet to be understood.

Merz (1969) examined different conditions for the solution of the Progressive Matrices problems. (The Standard Progressive Matrices [Raven, 1958] is a test of reasoning employing items that consist of a matrix of patterns with one cell missing. The examinee must choose the design that completes the pattern inherent in the matrix. The test has been widely used as a "culture-fair" test of reasoning.) Subjects instructed to reason out their answers aloud obtained significantly higher scores than subjects working the test silently. The whole body of literature on overt responding lends support to the idea that the verbal-activity involved in tutoring might be an important source of the gains.

On the other hand, the superior learning might be due to the sustained motivation provided by tutoring. Initial, short term motivation might not differ between tutoring and non-tutoring conditions but tutoring might pall less rapidly than other practice situations. The sustained motivation would be evidenced by more time-on-task behavior engaged in by the tutor during tutoring sessions, than by non-tutors who were in classes or studying alone. No measures of this important, observable variable have been made. The time-on-task hypothesis must remain as an untested competing hypothesis for the

Analysis of covariance showed a significant effect in favor of students who had received tutoring. On the 16-item posttest this effect amounted only to about one item. However, this small effect occurred consistently and represented an effect gained after just 30 minutes of instruction. There were no significant differences between the results obtained by the undergraduates with no teaching experience and the student teachers who had had teacher training and some practice in classroom teaching.

All experimental teachers tutored their three pupils before they taught the class in which the three other pupils were embedded. This sequence should presumably have worked against the hypothesis that tutoring was the more effective instructional method since teachers would have had practice with the unit by the time they instructed the class.

Bausell et al. commented that "classroom instruction is obviously more efficient in terms of total learning produced per given unit of instructional time." Such a statement is only correct if we view the tutor's time as a cost to the system and do not include tutor learning as a benefit to the system. This kind of assessment of costs and benefits is appropriate when tutors are paid adults and the goals are limited to tutee gains. In the Learning-by-Tutoring projects, however, the tutor's time is not a cost since the tutors themselves are benefitting.

One of the claims made in favor of the use of peer-group or cross-age tutors is that, because of language differences, tutees may understand the instruction provided by such tutors better than they understand their teachers' instruction. Tutors speak a language, it is hypothesized, that is more closely related to that of the tutees than is the language of the teacher.

The researchers investigated the independent and combined effects of two factors, token reinforcement and tutoring, on the reading progress of disadvantaged, inner city preschoolers, 26 white and 6 black. The two levels of the tutoring factors were defined by the characteristics of tutors. Tutors were either those preschoolers in the classroom who had learned quickly or one of five teenage job corps workers. The two levels of the token reinforcement factor consisted of tokens for either attendance or reading.

The 32 preschoolers were randomly divided into four groups to fit the cells of a 2 x 2 factorial design with the factors being the kind of behavior reinforced and the kind of tutor.

Results showed significant main effects for each factor in favor of preschool rather than teenage tutors and reinforcement for reading rather than for attendance. The results which were the most task-relevant are displayed in Figure 8. (As already noted, the label "adult tutoring" in this figure refers to tutoring by the job corps teenagers and "peer tutoring" refers to tutoring by preschoolers.)

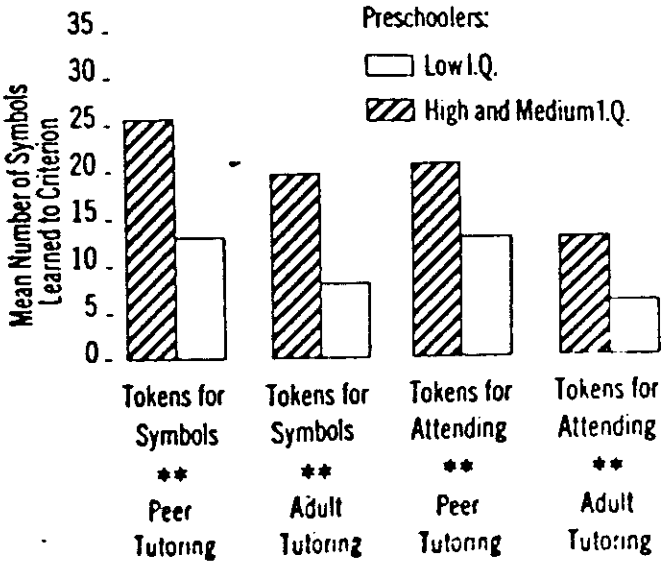

Figure 8. Hamblin and Hamblin results: The relationship of symbols learned to tokens and tutoring by IQ levels.

Caution re Standardized Tests

An experiment designed to provide data on the efficacy of tutoring employed randomly constituted control groups of tutors and tutees and was run for eight weeks (Snapp, Oakland and Williams, 1972). Fifth and sixth grade students tutored first, second and third grade students in a disadvantaged population. Tutoring took place 20 minutes per day four days per week and was conducted before school. A "Word Recognition Test" developed from the materials used in the study assessed tutee outcomes in comparison with the randomly constituted control group and indicated significant advantage to tutees ($F_{1,58} = 8.09$, $p < .01$). A standardized reading achievement test (the Metropolitan Achievement Test), however, showed no effects of the tutoring on tutees' reading scores. Tutors were assessed only on the standardized tests and showed no significant advantage over the randomly constituted control group.

These results are entirely consistent with the position that students learn the specific content of tutoring lessons but that the general content of standardized achievement tests is not likely to detect the effects of these specific learning gains with any sensitivity. One might expect eventual gains after perhaps a year or two of consistent progress, but certainly not after eight weeks.

Tutors were recruited through a publicity campaign which included flyers to local academic and vocational high schools. Most of the 240 tutors came from the academic high schools of the area in which the tutoring centers were located. Sixty-one percent of the tutors were white, and the remainder consisted of approximately equal numbers of black and Puerto Rican students with a few (2%) Oriental. Sixty-three percent of the tutors were female. There were similar numbers of tenth and eleventh grade students (52% were tenth grade), and the mean reading level was 10.0. No students were employed as tutors if they were reading more than three years below grade level. School guidance counselors had been asked to screen applicants who were "emotionally disturbed" or in "serious academic jeopardy."

Tutees, who were fourth and fifth grade pupils, were randomly selected from a pool of students who were reading below grade level and whose parents responded to an invitation to apply for the tutorial program. However, students with problems such as poor behavior, truancy, or retarded mental development were screened out by counselors before random selection was used to form an experimental group of 410 pupils and a control group of 185 pupils. Attrition, due to the failure to attend of 54 experimental pupils (13% of the experimental group) and 28 control group pupils (15%--no data were available for the 24, and 4 were inadvertently used to replace non-reporting tutees) reduced the numbers to 356 tutees and 157 non-tutees (the control group). These equivalent groups were composed roughly of 10% white, 30% black, 57% Puerto Rican, and 3% Oriental, and of even numbers of male and female.

Eight afternoons were devoted to the preservice training of tutors. "These preservice training experiences focused on (a) the goals of the tutorial program, the organization of the program, and the duties of a tutor, and (b)

TABLE 1

Cloward's Results for Tutees

Sample	N	Pretest mean	Mean gain
2 tutoring sessions a week	100	21.33	+6.20
1 tutoring session a week	73	22.89	+5.33
Controls (no tutoring)	79	21.98	+3.48

Covariance Analysis

Source	Before ss	After	df	MS	F	Required .05
Between	331.87	332.31	2	166.15	3.68	3.04
Within	11,603.84	11,022.59	244	45.17		
Total	11,935.72	11,354.90	246			

Factors used as covariates were sex, ethnicity, school grade, access to school programs in reading remediation, and pre-study reading level. Applications of the Duncan Range Statistic indicated a significant difference in means occurred only between the two sessions-a-week tutees and the no-tutoring controls. The author reported unsuccessful attempts to detect the relationships between the tutorial experience and school marks, behavior, attitudes, and aspirations.

A significant effect, however, was found for the tutor-pupil matching variable and an almost significant ethnicity-matching interaction. Examination

TABLE 2

Cloward's Results for Tutors:
Summary of Covariance Analysis
of Change from Prestudy to Poststudy,
Iowa Silent Reading Tests (Standard Scores)
for Experimental and Control Subjects

Test	Corrected Mean Change[a]		F
	Experiments	Controls	
Rate	-4.9	-6.4	0.12
Comprehension	9.0	1.4	7.76*
Directed Reading	24.8	15.0	9.20**
Poetry	12.0	7.9	2.57
Word Meaning	12.6	11.4	0.23
Sentences	9.9	10.3	0.05
Paragraphs	15.3	10.1	4.98*
Use of Index	18.7	13.8	3.67
Key Words	13.3	9.6	2.39
Total (Median Score)	13.2	7.6	14.14***

[a]Covariates used were prestudy reading level, Quick-Word score, sex, and school grade.

*Significant at .05

**Significant at .01

***Significant at .001

The attrition rate was 37.4% among tutors and 20.8% among non-tutors. Twelve percent of the tutors failed to report for the tutoring job and 23% resigned during the program. It was stated that most of those who resigned had class changes at the semester which prevented their tutoring. Whether the tutors who remained in the program to posttest time were high aspiring, determined students who would have made considerable progress anyway cannot be known.

The relevance for the present report of the program described by Cloward is that it was a large scale field implementation of a secondary-to-elementary tutoring program. More of the treatment appeared to make more of a difference for tutees, indicating that secondary school students can assist the learning process for elementary school children. Benefits in reading achievement seemed to accrue to tutors, but of course they spent extra time on reading each week and tutoring was not compared with any alternative use of the tutors' time. However, the tutors' school work did not suffer as a result of their participation as paid tutors. The treatment was not very intensive: as little as one hour of instruction per week could hardly be expected to build up much momentum. In the paragraphs quoted above, Cloward clearly stated several of the hypotheses which encourage the implementation of the Learning Tutoring Cycle.

The Youth-Tutoring-Youth project is run during out-of-school hours and tutors are paid. The YTY program is one for which successful tutors from LTC projects might be made eligible. Work as a tutor during the school year followed by employment as a tutor might be a synergistic arrangement.

Three Cross-School Tutoring Projects

We turn now to another project strongly influenced by the work of the Lippitts: the Ontario-Montclair project. This California project won an NIE pacesetter award and a dissemination grant under Title III ESEA. An evaluation of the project was performed by an external consulting firm, but their report was inadequate. Tutors were clearly volunteers and there was no mention made of how a control group was formed. Differences between experimental and control groups were reported by significance levels rather than by actual data. However, the method of implementation of these projects is of interest, and personnel involved are convinced that students benefit from improved self-concepts and learning skills. Many instances of dramatically improved attendance are cited in connection with these cross-age tutoring projects..

In Paducah, although the primary concern of project personnel was to enhance the social attitudes of participants, assessments of achievement were made by the Board of Education during the 1973-74 school year. For the 320 tutors involved from two junior high schools, gains during the year were half a month less "than expected" in reading skill but four months more "than expected" in mathematics. Interpretations of such data are difficult, and no correction can be made for the initial self-selection process of the volunteers. It would be interesting to know if tutors worked in math more than in reading, but data collected in the framework of a good design is needed for firm conclusions. However, the fact that tutees also gained more than expected in math but not in reading fits in with the supposition that math tutoring was occurring and producing noticeable effects. (A narrative account of the Paducah project can be found as an appendix to CSE Report on Tutoring #118.)

A Within-Secondary-School LbT Project

In contrast to the highly flexible cross-school Teacher Aide projects, in which tutors are volunteers and are assigned one or two at a time to various elementary school classrooms, a master's degree thesis prepared by a junior high school teacher illustrates a Learning-by-Tutoring project (Hoffmeister, 1973):

In an inner-city junior high school (approximately 90 percent Black), two ninth grade classes tutored two seventh grade classes in reading throughout the second half of a school year. The ninth grade tutors received pay. This was an elective course, and there were almost twice as many females as males. Tutors were, on average, two years below grade level in reading--which, however, placed them slightly above the average for the school's ninth grade.

The seventh grade tutees were selected because they were the most deficient in reading skills, reading on average at a third grade level.

> Another very important part of the peer-tutoring program was the Monday Seminar. This first day of each school week was not used for tutoring, but was set aside for a working seminar session with the tutors. They met as a group in Room 54 with the writer, while the tutees met as a group with Miss Sussman. This day was a day for continuing tutor-training, an in-service workshop so to speak. Lesson plans which had been turned in the previous Friday and checked over the weekend were discussed individually with each tutor, common problems or errors in planning were discussed with the body of tutors, new material was introduced and distributed, tutors shared problems with each other for encouragement and advice, lesson plans for the coming week could be outlined, and materials for them gathered.
>
> This seminar became a vital part of the program. It was the day when all of the tutors could see each other and share the experience of which they were a part. For the supervisor it was also a vital day, because it was the only day she could see them all together, maintain a teaching relationship with them, and listen to their needs and desires. (pages 20-21)

Two forms of the Gates-MacGinitie Reading Test, Survey E were administered, one as a pretest and one as a posttest. Tutees were evaluated on the Comprehensive Test of Basic Skills. Hoffmeister presented grade equivalent scores at pretest and posttest for all students, while maintaining strong reservation about the use of standardized tests and grade equivalent scores. The results are summarized in Table 3.

Both tutors and tutees made significant gains of about a year in grade equivalent scores. Since tutees were selected on the basis of low scores, some regression effects had to be expected,but these scores would normally be considered "successful" in a Title I program. Hoffmeister noted that ten of the fifteen poorest readers gained 1.2 years or more and all but two made at least month-for-month gains.

A Project that Nearly Failed

Projects which fail are not usually written up for publication, which is unfortunate since practitioners need to know what fails as well as what succeeds. Dreyer (1973) reported a project which nearly failed but was saved. Since it was a cross-age tutoring project using Title I funds and employing secondary students to tutor in elementary school classrooms, it is a study relevant to the LTC proposal.

Tutors were seventh graders who were having reading difficulties. They tutored first grade children who were likewise having difficulties. Tutoring lasted 15 to 30 minutes and took place in the first-grade classrooms. First-grade teachers, meanwhile, were busy with the remainder of the class. This was identified as one of the problems at the beginning of the program: tutors were ignored. They were given no feedback or reinforcements.

The training of tutors, or rather the lack of it, was reported to have been another problem that threatened the continuance of the project.

> Some [receiving teachers] had to spend so much time instructing tutors that they could have saved time by doing the tutoring themselves. Others gave assignments so vague that the tutor did not know what he was to do. In some cases the two or three tutors assigned to each first-grade classroom were definite distractors (p. 810).

A reading resource teacher and a consultant then set about developing clear, very specific procedures for tutors, based on discussions with teachers. Much of the article described these procedures for tutoring in reading.

In the proposed LTC programs these are components of the Tutor Support System which has been noted as an essential feature.

Ad Hoc Remediation vs. Full Sequence Instruction

The most usual secondary to elementary tutoring projects have been Teacher Aide projects. Secondary students are assigned one or two per room to work with selected tutees, often for remedial purposes. This pattern is evident not only in the projects described here from available literature but also from the survey, site visits, and other contacts made by the staff. Perhaps this pattern has developed because of the influence of a few projects or because it is one that fits in well with the present structure of schools. It requires only that receiving teachers accept a few aides, help for which they are often very grateful.

The LTC proposal places greater demands for change on the schools: To receive a whole class of tutors, for example, is a different proposition from receiving one or two aides. In the Learning-Tutoring Cycle, it can be argued that both more is expected of tutors and also less is expected of them. More is expected of tutors in that they are given responsibility for actual teaching, not just asked to listen to someone read or to grade papers or perform other chores that are assigned to Teacher Aides. When tutors in Teacher Aide programs are asked to teach, they must be prepared to teach whatever part of the curriculum the receiving teacher identifies as important for the tutee. Frequently, this is a part of the curriculum with which the tutee has already developed problems. Identifying misconceptions and re-educating a tutee, trying to build on what the teacher has explained but fill in gaps that are blocking tutee comprehension is probably more difficult than providing a full

upper grade teachers prepared the tutors and regularly held feedback sessions in which the group as a whole dealt with problems encountered in tutoring. As in the work of the Lippitts, major emphasis was placed upon the establishment of caring, understanding, helping relationships between tutors and tutees.

Klaus provides the following account of cognitive outcomes in the area of reading skills: (page 63)

> In addition to some specific studies of various aspects of the tutorial program, records have been maintained on the reading skills of all students since the beginning of the Tutorial Community Project. Overall, there has not been much improvement in reading at the upper-primary grades that can be attributed to the program. At the lower grades, however, both the immediate and the cumulative effects of tutorial assistance have been pronounced. Below are the year-to-year median percentile scores for grades 1, 2, and 3. The program began with first graders in the 1969-70 school year, expanded to the second grade the following year, and was made school wide in the 1971-72 school year.

	1969-70	1970-71	1971-72	1972-73	1973-74
1st grade	23	34	43	60	60
2nd grade	5	19	28	31	40
3rd grade	5	5	20	44	32

The data presented suggest that the first grade results steadily improved, and we understand that population changes did not appear to account for this improvement. If the changing scores can be attributed to the presence of the Tutorial Community Project--and the 5 percentile median scores of classes not involved in tutoring lend support to this attribution--then the major impact appears to have been at the first grade level. Median scores of cohorts typically declined the following year, then held their own. Scores did not, however, drop to the typical pre-project level of the fifth percentile.

Table 5

SECOND GRADE STANFORD READING TEST RESULTS

Nat. Stanine	PRIOR TO TUTORING May '66	Jan '67	May '67	Jan '68	May '68	Jun '69	SOME TUTORING May 1969		FULL TUTORING May 1970	
9									2.8	
8							4.2		2.8	
7		1.7					4.2		5.6	(70%)
6		2.6	(15%)					(54%)	5.6	
5		5.2					25.0		22.2	
4		6.0					20.6		30.6	
3		7.0					8.4			
2		34.4	(77%)				29.2		16.7	
1		43.0					8.4		13.9	

Column 1. In six testing periods* prior to any tutoring, 77% of the second grade scores ranked in national stanines 1 and 2, and only 15% ranked in the average range or better--in stanines 4 and above.

Column 2. The following semester, as a result of some experimental tutoring as first graders and some tutoring as second graders, 54% of the scores ranked in stanines 4 and above.

Column 3. The succeeding year with full tutoring and improved tutoring procedures, the percentage ranking in stanines 4 and above increased to 70%.

Presumably gains were most pronounced for first and second grade classes since these are the results presented, a finding that would echo the Newmark Melaragno results.

We strongly suspect that reading tests are more sensitive to gains due to instruction in the early grades than in later grades. Once a child has mastered decoding, the reading tests start measuring fluency, comprehension, general knowledge--skills not necessarily easily affected by school instruction.

Both the Newmark and Melaragno and the Ebersole programs suggest that classroom-to-classroom tutoring, aiming at consistent daily instruction rather than piecemeal remediation produced excellent results. We examine now a report of tutoring used for remediation.

their assignment, completed the Practice Exercise, and might then be assigned to "monitor" the reading of a storybook or might take the tutee out to play.

Data were presented to answer two questions: Did the tutoring benefit tutees, and did the training and/or experience of tutoring alter the behavior of tutors?

To test whether tutoring benefitted tutees, the authors randomly selected four of eight schools to use the tutorial program. In the four remaining schools, teachers managed the remediation activities without the assistance of fifth and sixth grade tutors. Four remediation-needing children were randomly selected from each of seven classes in each set of schools, and the criterion test was administered by personnel of SWRL. The authors claimed "a difference of 14 percent in favor of the classes with tutors" on the basis of the gains shown in Table 6, reproduced from Niedermeyer and Ellis.

Table 6

Niedermeyer and Ellis Results Regarding Effectiveness of Tutors: Percent of Pupils Scoring Above 90 Percent on Unit Tests and on Retests Following Remediation

Remediation by Teacher Only				Remediation by Teacher Plus Tutor			
Teacher	Test	Retest	Percent Gain	Teacher	Test	Retest	Percent Gain
1	81	93	+12	8	69	86	+17
2	57	86	+29	9	57	67	+10
3	85	91	+ 6	10	68	92	+24
4	51	16	-35	11	55	62	+ 7
5	55	80	+25	12	70	93	+23
6	78	81	+ 3	13	53	76	+23
7	91	91	0	14	56	68	+12
Mean	66	69	+ 3	Mean	62	79	+17

(ii) Tutors may learn the work more thoroughly when going through the process of full-sequence instruction rather than handling only specific remediation. Full-sequence instruction demands explanations, illustrations, logical development and comprehensive coverage. This processing, as was suggested earlier in this review, might be highly important in tutor gains.

(iii) Full-sequence instruction avoids an association of tutoring with failure. If everyone in the class is tutored, there is no encouragement to the tutee to resist tutoring or dislike it because it is a stigma, indicating one has failed in some way.

(iv) Tutors may prefer full-sequence instruction. Being assigned different tutees every few days and told to teach bits and pieces of the curriculum may lead to disgruntled tutors. (Much will depend on the manner of the tutee's teacher, of course.)

The studies just cited do not provide a basis for drawing final conclusions regarding the advisability of using tutors for remediation as opposed to full-sequence instruction. Suffice it to say that the issue is alive and should be the subject of further investigation: remedial tutoring seems to be the first instinct of practitioners but it may not be the best.

The next study to be described illustrates full sequence instruction with fifth and sixth graders teaching second and third graders.

A Learning-by-Tutoring, Full Sequence Instruction Project

Mohan (1972). Mohan reports a study of the effects of tutoring on unmotivated students--both tutors and tutees. Students who, on the basis of assessment by themselves and their teachers, were in the lowest 10% for motivation were randomly divided to form experimental and control groups. The control group students remained in regular classes whereas experimental group students participated in two hour-long tutoring sessions per week. Tutors were fifth

Table 8

Means and F-test Between the Post-test Scores of Control and Experimental Tutees

Variable	Control Group Means	Experimental Group Means	F	P
Mathematics	7.79	24.44	85.7570	.0001*
Motivation (T)[1]	38.86	41.38	1.4619	.2333
Motivation (S)[2]	43.79	47.69	5.2312	.0322*
Attitude	20.07	18.50	.1057	.7471
Self-Concept	23.85	24.06	.0377	.8490

*Significant at .05 level.
1 stands for motivation as assessed by the teacher
2 stands for motivation as assessed by the student

(Tables 7 and 8 reproduced from Mohan, 1972)

The mathematics test was locally developed to assess the objectives in set theory. Clearly, tutors and tutees knew the work substantially better than the randomly constituted control groups. It is not clear from Mohan's account whether tutoring replaced some normal classwork or was an add-on activity; i.e., the influence of time-allocated was not necessarily controlled for in this study. However, tutors clearly learned from teaching, and tutees learned from being tutored.

needed before the existence of either of these processes can be accepted, but the evidence for the more modest hypothesis that tutors learn what they teach is now available from many studies.

The study which provided an excellent illustration of the principle that tutors show cognitive benefits primarily when the tutoring content matches their needs was conducted by Olds (1976), a student of Urie Bronfenbrenner. Olds was able to contrast being a tutor with receiving tutoring from an adult. Sixth graders who tutored second graders for three months made greater vocabulary gains than sixth graders given their own adult tutors during the same time. Olds attributed this to the "initial enthusiasm and preference for tutoring as opposed to being tutored." At the end of a further three months, disordinal interaction was located on the assessment of gains in reading comprehension: better readers made greater gains when they received tutoring from adults, whereas poor readers made greater gains when they acted as tutors to younger children. Olds commented, "Poor readers needed basic skill review, and therefore made gains as a result of teaching basic skills to second graders. Better readers, however, needed work in more advanced areas of reading: teaching elementary skills added nothing to their repertoire of reading abilities."

As emphasized above, cognitive gains for tutors should not be expected unless they are teaching work which they themselves need to learn or practice.

More on the Issue of Standardized Tests

With the exception of the Cloward (1967) study, the tutoring projects examined in this literature review have been those reported in 1969 or later. Rosenshine and Furst (1969) undertook a survey of objective data available

> Grannick's Philadelphia group had a mean of .4 years below grade level on the pre-test. They gained one year during a seven week period. These gains, however, were not statistically significant. One might, however, argue for educational significance of a finding which shows grade level attainment in a short period (p. 6).

Rather than urge the translation of raw scores into scores with misleading labels (grade equivalents), better advice to researchers and evaluators would be to implement good research designs, provide competitive control groups that receive the same allocated instructional time and use sensitive measures of student achievement, such as tests referenced to instruction. This advice should inform the work of evaluators of tutoring projects.

It is of interest to note that the list of "unsuccessful" projects presented by Rosenshine and Furst consisted almost entirely of reading projects. (The single exception was one of their own studies which was reading and math. They undoubtedly used standardized tests as the measures.) The difficulty of measuring reading achievement is considerable. Before reading is established, results will be strongly affected by method of teaching employed (e.g., systematic or intrinsic phonics); and once reading is established, reading tests become thinking and reasoning tests that may be insensitive to instructional effects. Projects in which math achievement is examined might yield more consistent results because math achievement may be more influenced by school instruction. (See, however, p. 77 below.)

In discussing the "unsuccessful" projects, Rosenshine and Furst located a possible reason for lack of significant findings as residing in the task characteristics:

> ...the unsuccessful programs were more diffuse in their objectives, attempting to provide a variety of enrichment services. In the successful programs, more time was spent on cultural activities, and less time on academic activities" (p. 4).

(Ellson, Harris and Barber, 1968, advanced a similar explanation for the superiority of programmed tutoring over traditional adult tutoring.)

Summary of Cognitive Outcomes

Several controlled experiments have examined the cognitive impact of tutoring with positive results. Fitz-Gibbon (1975) showed that low achieving, inner-city, ninth grade tutors learned more during three weeks spent tutoring fourth graders than an equivalent control group learned by spending the same three weeks working on the same objectives in their regular classes. The positive effects of the tutoring experience were still significant three months after the project. Allen and Feldman (1973), in a summer school project with paid participants, found fifth graders eventually achieved higher scores when assigned tutoring than when assigned the same time to study alone.

Is it the expectation of tutoring or the process of tutoring that produces the cognitive benefits? Present limited evidence, it was argued, suggests that the process of tutoring is the more influential variable, which implies that the *act of tutoring makes a special contribution to learning, one that might not be effectively replaced by other means of motivation.*

The effect of tutoring on tutees was examined by Bausell, Moody and Walzl (1972) contrasting tutoring with equivalent time spent as a member of a class. Results favored being tutored over receiving classroom instruction from the same teacher. An experiment by Marks, Doctorow and Wittrock (1975) was noted because it lends credence to the idea that the closer language-match between tutor and tutee might be one of the advantages that peer-tutoring has over adult tutoring.

Results obtained by Hamblin & Hamblin (1972) which indicated that teenage tutors were not as effective with kindergartners as were kindergarten tutors were rejected as unsubstantiated. An experiment (Snapp, Oakland and Williams, 1972) in which upper grade students tutored lower grade students before school

A project reported by Dreyer (1973) emphasized the need for an effective Tutor Support System.

The review of those field studies in which tutors were secondary school students ended with the question as to whether assigning tutors to conduct remedial work was less desirable than assigning them to provide "full-sequence" instruction. For projects in elementary schools, strong results were found in two projects involving regular, full-sequence, classroom-to-total-classroom instruction [Newmark and Melaragno (1969) and Ebersole (1971)]; but unconvincing results were noted for the benefits of tutees of selective remedial work by upper grade tutors [Niedermeyer and Ellis (1971)].

Arguments advanced in favor of full-sequence instruction were that it makes teaching success for tutors more likely, promotes better tutor learning, avoids the stigma associated with individual remediation, and that tutors are likely to prefer responsibility for full sequence instruction with a single child to piecemeal remediation activities with a succession of children.

Mohan (1972) reported a field experiment in which upper grade elementary students tutored lower grade elementary students in set theory. Gains were strong and significant for both tutors and tutees in comparison with randomly constituted control groups remaining in regular classes. A test closely keyed to instruction, it should be noted, was the achievement measure employed by Mohan. Results of a recent study by Olds (1976) supported the principle that to derive cognitive benefits, tutors must teach work they need to learn or practice. A review of pre-1970 projects by Rosenshine and Furst was discussed with particular emphasis on the misleading nature of grade equivalent scores. The relevance of assessment to instruction was seen as an important feature determining the results of studies.

Myers, Travers, and Sanford (1965). These researchers examined the performance of fourth, fifth, and sixth grade pupils on a learning task when the pupils were randomly assigned to one of four learning conditions. The four learning conditions, three of which involved students working in pairs on the learning task, were: Condition I: tutor role; Condition II: tutee role; Condition III: tutor-tutee switching; and Condition IV: self-instruction.

The learning task was rote memory for the English translation of German nouns. The German noun and two choices for its meaning were printed on cards. In the tutor role the student showed the card to the other pupil, and on the back of these cards, as they were held up, the three words were displayed with the correct English choice indicated. The tutor role students could therefore see the correct answer as soon as they posed the item to their partners. The students in the tutee roles were to guess the correct answer, after which the tutor stated the correct answer without comment. For the tutor-tutee switching, students reversed roles in mid task. In the self-instruction role, students studied alone from cards. In all conditions, students were allowed whatever time they needed to proceed through a set of cards. Amount of time was generally 15 to 20 minutes. The procedure was repeated with new word lists, i.e., new tasks, on three consecutive days. Results for the posttest data indicated significant effects for conditions and tasks and a significant interaction between tasks and conditions. Figure 9 is reproduced from the article (p. 70).

The steep decline across the three tasks (representing three consecutive occasions) was taken as indicative of boredom with the materials. The students in Condition I, the tutor role, achieved the lowest scores both at immediate posttest and for retention tests given from three to nine weeks

Bull and Wittrock (1975). Further support for the contention that, in order for tutors to benefit, they must be actively engaged in teaching--that is, in processing, presenting, and organizing the material for the tutee--comes from the work of Wittrock and associates. They are interested in learning as a "generative process," one in which the learner must be actively involved.

Ninety fifth graders were randomly divided into three groups, each of which was instructed to learn word definitions in one of the following ways: (1) generate (draw) an image of the word and its definition, (2) trace a picture (image given) representing the word and its definition, or (3) learn the verbal definition by copying it. The group that remembered the definitions best one week later was the group that generated their own images. Bull and Wittrock commented that, considering this and other studies, what seems to be important is that some type of generation activity is called for.

It seems likely that tutors who know what they have to teach (like those pupils in the experiment who saw the word definitions that they had to learn) will retain the work better if they have to generate their own way of teaching it (as the subjects in Bull and Wittrock's experiment had to generate their own images) than if they simply follow prescriptions for teaching (as did subjects in the experiment who simply followed instructions for tracing provided images).

Ronshausen, developer of programmed tutoring materials in mathematics, emphasized the minimal demands on tutors:

> The need for professional judgments by the tutor is reduced to a minimum--the tutor judges only the correctness of the child's responses. While the operational programs describe correct and incorrect responses and tell the tutor how to proceed in either case, the content programs give specific correct responses. Thus, persons with limited knowledge of mathematics can perform successfully as tutors (Ronshausen, no date).

"Programmed" tutoring referred to a set of "operational programs" and "content programs" which tightly controlled the tutors' procedures. These programs had been developed over several years of experimental work. "Directed" tutoring consisted of procedures recommended by a reading specialist and reading consultants; it was described as an adaptation of classroom methods to an individual situation. As such, Directed tutoring employed a wide variety of support materials such as readers, pictures, games, crayons, etc.

Selection and characteristics of tutors and tutees. Tutors were recruited by notices in school and Parent-Teacher Association bulletins. All were white and female and had had at least a high school education. They were paid on an hourly basis.

Since an important question to be answered by this field test was the relative effectiveness of Programmed versus Directed tutoring, it was a serious error that tutors were not randomly assigned to these methods of tutoring:

> The selection of tutors was subjective, based largely on the judgment that the candidate was capable of working well with children and school personnel and competent to carry out the required tutoring and recording procedures. Assignment of tutors to programmed or directed tutoring was presumably influenced by judgments of compatibility between personality traits and tutoring tasks, but the factors which entered into these judgments or their actual relevance is not known. Nine of the programmed tutors were carry overs from the previous experiments, i.e., they had one semester experience in programmed tutoring (p. 310).

Thus, not only was the assignment of tutors non-random, but the Programmed tutoring treatment was confounded with experience. A newly developed, first trial, Directed tutoring program, employing persons with no prior tutoring experience was compared with tutoring which had had several years of revisions and which was implemented in part by tutors with some prior experience. (The total number of tutors was not reported so the proportion of experienced tutors cannot be ascertained.)

correct responses, the loops or branches in the program, the use of prompts, and the fact that the material to be learned was divided and taught in small units. An experienced teacher would see many examples of her art systematically employed and also that certain common practices of teachers were systematically avoided. For example, in its present form programmed tutoring of reading does not include and is not preceded by a readiness program and no special devices beyond the learning situation itself are used to motivate and stimulate the children's interest. These appear to be far less necessary in programmed tutoring than in group teaching. It should be remembered, however, that activities designed to develop readiness and interest-oriented motivational techniques are likely to be included in the work of the classroom that is concurrent with the tutoring.

An educator would perhaps recognize the programs as simple forms of the discovery method. The first step in each program presents a problem or task in its most difficult form with a minimum of context. If the child cannot solve this form of the problem, it is progressively simplified by changing its form or by providing hints or additional information until the child discovers the solution for himself. (p. 314)

Regarding Directed tutoring the authors wrote:

...those who observed it in operation during the course of the experiment had every expectation that the participating children would benefit. The sequence of directed tutoring activities was planned in detail by an experienced reading specialist with the guidance and support of several experts.

The training of the tutors appeared to be adequate for the limited task involved and was well supported by supervision which provided additional on-the-job training when necessary.

Both tutors and those who observed them had no doubts that a real contribution was being made to the children's reading skills. (p. 343)

Table 10

Ellson, Harris and Barber Analyses of Covariance:
F-Ratios and Significances of Experiment vs. Control Effects
on Posttests based on Analyses of Covariance,
Each Tutoring Condition is Analyzed Separately[a]

Posttest Measure	Condition: Programmed tutoring 1-session F	p	Programmed tutoring 2-sessions F	p	Directed tutoring 1-session F	p	Directed tutoring 2-sessions F	p
PPVT	1.83	n.s.	.03	n.s.	6.69	<.05	.01	n.s.
Ginn Total (A + B)	.01	n.s.	12.03	<.01	1.67	n.s.	.70	n.s.
Stanford Reading Total	.00	n.s.	1.09	n.s.	2.56	n.s.	1.03	n.s.

[a]The degrees of freedom were 1 and 83. The F-ratio needed at the .05 level of confidence was 3.95 and 6.94 for the .01 level.

The authors note that the lack of significant differences in the Stanford Reading tests "may reflect one of the difficulties of obtaining a satisfactory evaluation of reading achievement at the beginning level. The Ginn tests are aimed directly at the content specifically taught through the Ginn Basal Reader Series. The Stanford test samples a broader content which only partially overlaps that of the Ginn Series" (p. 334).

The results indicated that significant differences were the result of high scores of tutees who had received two sessions of programmed tutoring per day. Only this treatment appeared to be effective and this only on the Ginn test. On this test there was a very substantial effect, presumably due to tutoring. The authors noted that "significant differences were limited almost entirely to measures highly weighted with vocabulary content" (p. 333). The significant effects for Directed tutoring in the Peabody Picture Vocabulary Test (PPVT) was not discussed.

> recognition, but can lead to lower scores on silent, timed reading tests, especially in grade 1, when the average phonics-trained child has not yet acquired sufficient skill to "sound out" words that he cannot recognize immediately. This child concentrates on working out the words and has less time and energy left to devote to getting the thoughts. By the second grade, however, when the average phonics-trained child has mastered a sufficient amount of phonics knowledge and skill, he can devote himself more to meaning. (Chall, 1967, p. 107)

There is also the question of the fairness of the tests to the two kinds of tutoring, Programmed vs. Directed. The Ginn test on which a significant result was obtained was apparently more closely tied to the instruction which had been received than was the Stanford. The problem of test-fairness arises because if the outcome measure which showed Programmed tutoring superior to Directed tutoring were a test tied to the vocabulary built into the Programmed tutoring, but not into Directed tutoring, this would hardly be a fair test. Ellson developed the Programmed tutoring over several years, with many changes (p. 346), and has written at some length on the Ginn tests. A link between program and test cannot be ruled out. One must wonder if some tutoring activities were highly test-related, such as by the presentation of similar vocabulary or item formats. Even if we set aside the suspicion that the Ginn test might have favored the specific content found in Programmed tutoring, there is still the problem already pointed out that tutors were not randomly assigned, nor equally experienced and the Directed tutoring was receiving its first tryout.

However, the authors offered an explanation for the better results from the Programmed tutoring which appears both highly plausible and interesting. They point out that the Directed tutoring involved many "support" activities, whereas in the Programmed tutoring the time was totally devoted to reading.

Her tutors, as Ellson's, are trained to avoid negative comments completely and to reward all correct responses with positive reinforcement.*

As with Ellson et al., Ronshausen's tutors were paid adults. Each tutor received about 30 hours of training in the two component programs of the "Programmed Math Tutorial" (PMT). The two programs were the operational program, which defines general tutoring methods (applicable to any subject area like teacher training "methods" courses), and the content program (the actual mathematics to be taught), organized for discovery learning and carefully structured, sequential mastery learning.

A large field test of PMT, in its third cycle of development, was conducted by Ronshausen in the 1972 to 1973 school year. This field test had two objectives: to test first graders on a revision of first grade PMT materials, and to see if these same materials could be effectively used for kindergartners as well as for the first graders for whom they had been originally developed.

The results are shown in Tables 11 and 12, reproduced from Ronshausen (1974). Note that tutored students (drawn from grades K and 1 in one school system and grade 1 in another school system) achieved significantly higher scores on all tests of mathematical concepts, but not on all tests of computational skills. Ronshausen concluded:

---

*The avoidance of negative comments for incorrect answers is incorrectly termed by Ronshausen (as frequently elsewhere) as the avoidance of "negative reinforcement." Technically a negative reinforcement occurs when, contingent on the correct response, an aversive stimulus is discontinued. This discontinuance serves to reinforce (reward) the correct response. What is avoided in these programmed tutoring prescriptions, is not "negative reinforcement" but "punishment" such as negative comments.

Table 12

Ronshausen's Summary of Math Computational Skills:
Mean Scores for Treatments

School System	Grade	Subtest	Treatments: Tutored	N	Untutored Control	N	Diff. in Means	t	p
A	1	Metropolitan Achievement Tests	9.8 (10.2)**	140	10.8 (10.1)	85	.1	.80*	NS
A	1	Criterion Items Subtest, Computational Skills Section	4.3 (4.4)	140	4.5 (4.2)	85	.2	.74	NS
B	1	Metropolitan Achievement Tests	13.4	136	13.1	78	.3	.36	NS
B	1	Primary School Mathematics Criterion Test, Computational Skills Section	10.5	136	9.3	78	1.2	2.75	<.01
B	K	Metropolitan Achievement Tests, Primer, Computational Skills Section	9.7	32	4.9	33	4.8	3.81	<.001
B	K	Primary School Mathematics Criterion Test, Computational Skills Section	7.6	32	3.8	33	3.8	4.49	<.001

*t was calculated as $\sqrt{F}$.

**Adjusted mean scores obtained in ANCOVA are shown in parentheses; pretest total score was the covariate.

students two days per week for six months, fourth graders at an inner city school showed gains significantly higher than those of a control group in arithmetic concepts but no significant differences in arithmetic computation (both concepts and computation were measured by the Stanford Achievement Test).

Ronshausen (1975). Ronshausen followed her 1974 study with one using juveniles as tutors in place of paraprofessionals. It should be remembered, in assessing this report, that the focus was still on benefits to tutees rather than tutors. Tutor behavior was determined by the programmed materials so that no knowledge of mathematics was required.

Ronshausen's 1975 report described programs in several schools employing children from fourth grade through high school as tutors. Usually the cross-age tutoring program was initiated because there were no more funds for paraprofessionals, but some programs were initiated in the hope of positive attitude gains for the tutors.

Ronshausen pointed out several characteristics common to all the projects:

- The tutoring materials and projects were used as written, regardless of ages or assumed abilities of tutors. Tutors were expected to follow the programs word-for-word. [Comment: This might limit the assignment of a tutor's role to students with adequate reading achievement.]
- Tutor training, essential to the proper implementation of programmed tutoring, was done by local, professional personnel, usually the same people who later supervised the tutors' work. [Comment: The use of student-tutors thus reduced the cost of the program, but not to zero. Professional personnel were still required to train and supervise.]
- All project directors agreed that student tutors required more supervision than adult tutors.

Summary of Programmed Tutoring

Programmed tutoring has been developed primarily for effective instruction in basic skills for primary grade children. Tutors have usually been paid paraprofessionals. The effectiveness of these empirically developed materials was not questioned but several points were emphasized. First, tutors may not benefit cognitively from following highly prescribed materials and may become bored. A couple of experiments indicating the lack of cognitive gains in the situation or a lack of active cognitive processing were cited ("Thinking," said Dewey, "is the method of an educative experience.").

Secondly, contentions that "directed" tutoring is effective were rejected because of a confounding of tutor experience with the programmed or directed tutoring treatment, the possibility of insensitive tests and the possibility of instruction having been matched to the test in the programmed treatment. The question of the differential sensitivity to instruction of various parts of tests was raised by examination of Ronshausen's results. Ronshausen (1974), Fitz-Gibbon (1975) and Zack, Horner and Kaufman (1969) have all found computational skills to be apparently less easily influenced by instruction than math concepts.

Ronshausen reported a study in which high school students rather than paraprofessionals were tutors and pointed out several consequences of this change for project management.

tutors improved their attendance record but without further information, this figure cannot be interpreted as good or poor. Perhaps 80% of other students improved their attandance, for example.

What of student satisfaction? We shall deal first with tutee satisfaction. Most practitioners report enthusiasm from students although this enthusiasm can wear thin in prolonged programs. Tutees have generally been reported keen to participate (Mohan, 1972; Snapp, Oakland & Williams, 1972; Paducah, Kentucky program--Klaus, 1975, p. 57). Fitz-Gibbon found 67 percent of the randomly selected fourth grade tutees would have liked to continue being tutored, 12% were unsure and 21% did not wish to continue. Tutees who reported being given considerable amounts of drill on multiplication tables were significantly less likely to want to continue with tutoring than those reporting less drill work. Blank, Koltuv and Wood (1972) reported a marked resistance of kindergarten tutees to tutorial sessions with paraprofessional adults. As Sherertz has noted, "tutoring" is often associated with a need for special assistance and may be resisted if it is perceived as remedial. An important influence on acceptance of tutoring by tutees will be a consideration of what they are missing by receiving tutoring. This consideration apparently applies to many innovative projects; although the projects are well-liked, students do not like to be pulled out of class to participate because they might miss something (Stearns, 1977).

Turning now to tutor satisfaction, it must be noted first that tutors have been volunteers in the majority of projects reported, and might therefore be expected to like tutoring. As with the attendance question, it is often the extreme cases that are noticed. For example, a report from the Salembier project (Site #87 in CSE Report on Tutoring #118, p. 20) of a student who wanted

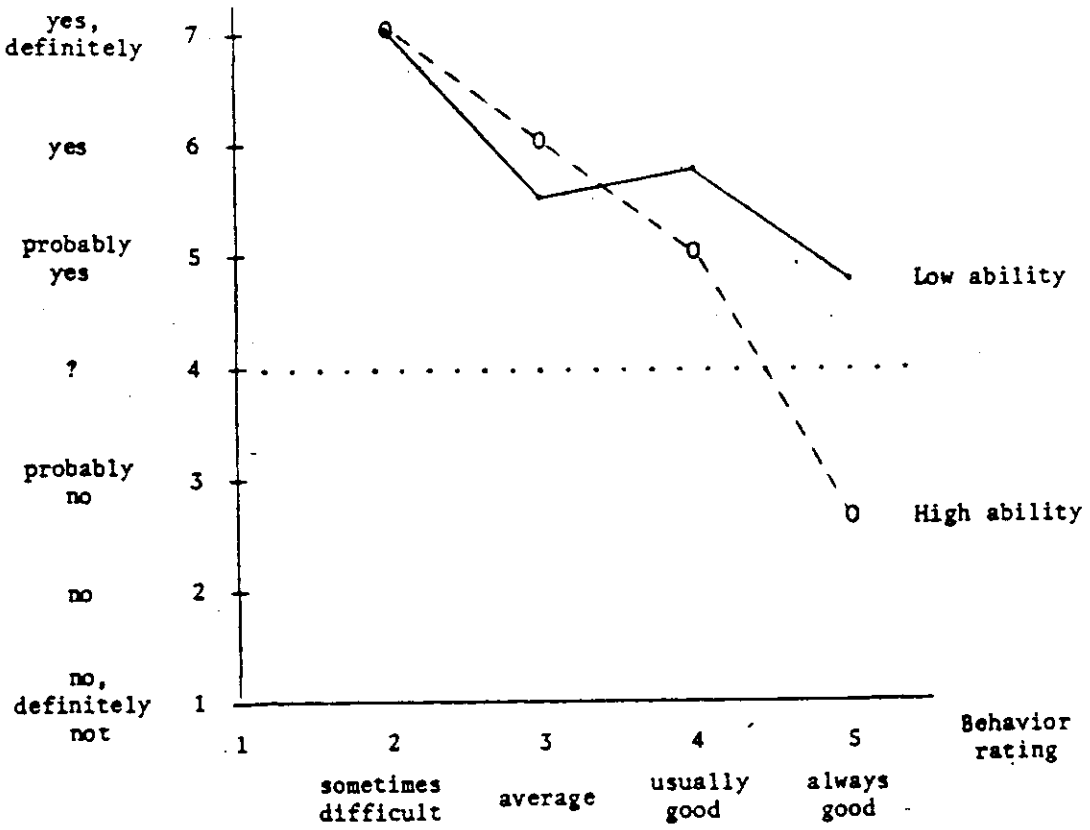

Figure 10. Average responses to the question "Would you like to have tutoring as an elective one quarter?"

The tutors in Fitz-Gibbon's study were not volunteers. Among the 40 randomly selected students who had tutored, 66 percent responded positively to the idea of having tutoring as an elective in the future, 10 percent were uncertain and 24 percent were negative. An important finding three months after the project was that students who had been required to tutor were significantly more likely to wish to tutor in the future than those who had not been required to tutor: non-voluntary exposure to tutoring had increased students' wish to tutor (see Figure 11). (Analysis of variance of the data in Figure 11 showed significant main effects for tutoring and sex.)

cooperative in school and tutors did not like trying to persuade or coerce reluctant tutees into working. Ronshausen (1974) reported that high school tutors preferred to work with the primary grades: "...high school tutors prefer to work with first and second graders. The tutors may accept a few fourth graders but they will resign rather than tutor only older children" (Ronshausen, 1975, p. 5).

Gallup polls have consistently reported "discipline" as the major perceived problems in schools, and junior high schools usually represent a high point in adversary relationships between teachers and students. One of the strongest immediate benefits to teachers from LTC programs might be a relief from the role of martinet. These are reports from two tutoring projects implemented in inner-city junior high schools:

> The supervisor of the (seventh grade) tutees...remarked that in September the twenty tutees in each class were very "rambunctious and playful." It was difficult to get them to work in this large group because their need for help was so great. When each student was assigned a tutor in December, 1972, the whole atmosphere of the class changed, and it became a classroom "alive with work." Discipline problems, she noted, became almost non-existent" (Hoffmeister, 1973, p. 26).

> For three weeks low achieving ninth graders (tutors) worked hard and presented no discipline hassles. They were colleagues, seeking advice and assistance, not students playing to a peer group and enjoying such diversions as the "she's got my pencil" disruption and the "Don't mess with me" routine (Fitz-Gibbon, 1975(b), p. 3).

When tutors themselves face the problem of trying to motivate students to work, their empathy for teachers might increase, an effect which is predicated from work in social psychology (see CSE Report on Tutoring #116) but which, it appears, no one has attempted to measure. Many practitioners comment on the effect, however, and tutors themselves often mention it spontaneously.

A change in tutors that several projects have tried to assess is in the area of self-concept. Damico and Watson (1974) examining tutoring in a self-

Summary of Non-Cognitive Outcomes

Quantitative findings regarding the effects of tutoring on self-concept, attendance, attitudes to school, and attitudes to school subjects are not strong, but it is widely believed that adequate measurement devices would find them to be positive. It may be that researchers assessing the non-cognitive effects of tutoring interventions will need to focus on subjects in extreme categories to locate significant changes in attitudes and obtain results that are congruent with the rather consistent reports of hundreds of practitioners. Practitioners observe salient cases, but most statistical tests in use examine central tendencies. It is also likely that, as with cognitive effects, the affective impact of tutoring will be closely related to the tutoring process, not a generalized effect on such a complex attitude as "self-concept."

It is vital that, difficult though it may be, research into the non-cognitive effects of tutoring projects continues. The effects observed by practitioners (Report #118) and predicted by theory (Report #116) are effects which are important for schools and for society in general.

should be developed to most effectively reach these goals, and this might entail neglecting cognitive goals for tutors. It might be found, for example, that desirable, responsible behavior and a caring relationship was best evoked in tutors when they were given kindergarten children to tutor rather than, say, fourth graders. It is not likely that secondary students would learn much in the way of cognitive skills in a project in which they worked with kindergarten children, but such a project might be the most effective in achieving non-cognitive goals.

The last category adopted by Sherertz seems to imply that all four goals are of equal importance or to presuppose that they occur together. The example just presented indicates that choices must sometimes be made on the basis of goal priorities. Can it be assumed, nevertheless, that positive outcomes are generally correlated? For example, are projects that students like usually effective academically? While it is often assumed that a project that promotes students' enjoyment of school will be cognitively successful, such a link between affective and cognitive outcomes has not been established.

Attempts to measure liking and disliking of school and to relate these attitudes to school success have shown that students may like school but do poorly, as well as like school and do well. Similarly, from studies of instructional television (ITV), we find that "liking ITV is not always correlated with learning from it" (Chu and Schramm, 1967, p. 123). Studies of Follow Through classrooms found no significant correlations between children's negative affect and math scores, with entering ability controlled for statistically (Stallings and Kaskowitz, 1975, p. 281).

These few examples serve as a warning that there may be an independence between cognitive and affective outcomes, thus implying that cognitive and

classification. Is it possible to plan for both or must one goal take precedence? Do positive affective outcomes come linked together for tutors and tutees? And would a project aiming at positive affective outcomes be non-cognitive? If it were found that in tutors the best affective outcomes derived from successfully implemented projects with cognitive goals, the classification would cease to distinguish.

The main problem with a classification by goals, however, rests with the problem of how one knows a project's goals. Does one find out by asking project personnel? Since innovators are usually, of necessity, hopeful and optimistic, a claim to be working towards all goals is not unusual, but it does not provide any information about the nature of the project. Furthermore, there's "many a slip 'twixt cup and lip." A project might have a stated goal of tutor-growth-in-responsibility, but if no attention was paid in project implementation to promoting such growth (such as by increasing tutors' significant responsibilities), the stated goal would be misleading and could scarcely be used to classify the project in a way that is informative regarding the outcomes that might be expected.

The omission by Klaus of Teacher Aide projects could be justified by arguing either that such projects are not really tutoring projects or that they are designed to help the tutee. However, Teacher Aide projects are frequently called tutoring projects, and yet are quite distinguishable from a Tutorial Service project in which tutors are consistently involved in working with individual tutees. Again, the fact that both the children in the classrooms that have an aide, and the older students acting as the aides, benefit in various ways, does not or should not preclude recognition that in a Teacher Aide project the activities of the "tutors" (aides) are chosen to best assist

In this report, programmed tutoring was dealt with separately from less structured tutoring. In general, applying structure to a project means removing options and choices and specifying procedures. The extent to which such structuring is effective for various goals will have to be a topic for formative evaluation in many projects. As a preliminary step in such studies, it should be noted that prescriptions can be made regarding the CONTENT to be taught and/or regarding the METHODS to be employed.

Examples of projects classified by this kind of structuring of the tutoring task are suggested in Figure 13, a "content/method matrix."

METHOD	CONTENT	
	prescribed	open
prescribed	Ellson et al. "Programmed"	Ebersole's reading project
open	Fitz-Gibbon (1975)	Sherertz (Ontario-Montclair project)

Figure 13. Examples of projects classified by the amount of structuring of the tutoring task

In programmed tutoring, both instructional method and content are prescribed. In Ebersole's reading project a method is taught to tutors which they can then employ in working with any reading material. In Fitz-Gibbon's study, objectives for instruction were specified, but tutors could decide daily what activities to undertake with tutees in order to reach the prescribed objectives. In the Ontario-Montclair projects developed by Sherertz and others, tutors could frequently decide what to teach and how to teach it, i.e., both method and content were "open."

REFERENCES

Allen, V. L., & Feldman, R. S. Learning through tutoring: Low achieving children as tutors. Journal of Experimental Education, 1973, 42(1), 1-5.

Bausell, R. B., Moody, W. B., & Walzl, P. N. A factorial study of tutoring versus classroom instruction. American Educational Research Journal, 1972, 10(4), 591-598.

Blank, M., Koltuv, M., & Wood, M. Individualized teaching for disadvantaged kindergarten children: A comparison of two methods. The Journal of Special Education, 1972, 6, 207-219.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). Taxonomy of educational objectives. Handbook 1: Cognitive domain. New York: David McKay Co., Inc., 1956.

Bloom, S. Peer and cross-age tutoring in the schools: An individualized supplement to group instruction. Chicago: Chicago Board of Education, 1976.

Boyd, G. S. Reading achievement and personal adjustment: A study of the effects of participation as a tutor and as a pupil in an elementary school tutorial reading program. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa, 1969.

Bremmer, B. L. Students helping students program. Seattle: Seattle Public Schools, Planning and Evaluation Department, August, 1972.

Bruner, J. S. Review of A. Gartner, M. Kohler, & F. Riessman, Children teach children. Saturday Review, 1972, 62-62.

Bull, B. L., & Wittrock, M. C. Imagery in the learning of verbal definitions. The British Journal of Educational Psychology, 1973, 43(3), 289-293.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Chall, J. S. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Chu, G. C., & Schramm, W. Learning from television: What the research says. Stanford, CA: Institute for Communication Research, 1967.

Cloward, R. D. Studies in tutoring. Journal of Experimental Education, 1967, 36(1), 14-25.

Crandall, V. C., Katkovsky, W., & Crandall, V. J. Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. Child Development, 1965, 36, 91-109.

Lyman, H. B. Test scores and what they mean. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Marks, C., Doctorow, M., & Wittrock, M. C. Reading as a generative process. Journal of Educational Psychology, 1975, 67(4), 484-489.

Mehrens, W. A., & Lehmann, I. J. Standardized tests in education. New York: Holt, Rinehart, and Winston, 1969.

Merz, F. Der Einfluss der Verbalisierens auf die Leistung bei Intelligenzaufgaben [The effect of verbalization upon performance on intelligence tests]. Zeitschrift fur Experimentelle and Angewandte Psychologie, 1969, 16(1), 114-137.

Mohan, M. Peer tutoring as a technique for teaching the unmotivated. Fredonia, NY: State University of New York, Teacher Research Center, January, 1972.

Myers, K. E., Travers, R. M. W., & Sanford, M. E. Learning and reinforcement in student pairs. Journal of Educational Psychology, 1965, 56(2), 67-72.

Newmark, G., & Melaragno, R. J. Tutorial community project: Report of the first year. Santa Monica: Systems Development Corporation, 1969.

Niedermeyer, F. C., & Ellis, P. Remedial reading instruction by trained pupil tutors. Elementary School Journal, 1971, 7, 400-405.

Olds, D. Cross-age tutoring and parent involvement. Draft of dissertation study, Cornell University, January, 1976.

Paoni, F. J. Reciprocal effects of sixth graders tutoring third graders in reading. Unpublished doctoral dissertation, Oregon State University. Corvallis, 1971.

Raven, J. C. Standard progressive matrices. London: H. K. Lewis, 1958.

Robertson, D. J. The effects of an inter-grade tutoring experience on tutor self-concept. Paper presented at the Annual Meeting of the California Educational Research Association, San Diego, April, 1971.

Ronshausen, N. L. The effects of mathematics achievement of programmed tutoring as a method of individualized one-to-one instruction. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 1974.

Ronshausen, N. L. The programmed math tutorial--paraprofessionals provide one-to-one instruction in primary school mathematics. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, 1975.

Rosenshine, B., & Furst, N. Summary of the effects of tutoring upon pupil achievement: A review of research. Paper presented at the Annual Meeting of the American Educational Research Association, 1969.

Sherertz, D. Ontario-Montclair School District evaluation summary: Cross-age teaching. Ontario, CA: Ontario-Montclair School District, 1970.

Sherertz, D. Personal communication. La Verne College, California, September 9, 1975.

Snapp, M., Oakland, T., & Williams, F. C. A study of individualizing instruction by using elementary school children as tutors. Journal of Social Psychology, 1972, 10(1), 1-8.

Soar, R., & Soar, R. An empirical analysis of selected follow through programs: An example of a process approach to evaluation. In Seventy-First Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1972.

Stallings, J. A., & Kaskowitz, D. H. Follow Through classroom observation evaluation 1972-1973. Menlo Park, CA: Stanford Research Institute, 1974.

Stearns, M. S. Evaluation of the field test of project information packages, Vol. 1, Summary Report. Menlo Park, CA: Stanford Research Institute, April, 1977.

Werth, T. G. An assessment of the reciprocal effect of high school senior low achievers tutoring freshman low achievers in English classes. Unpublished doctoral dissertation, Oregon State University, Corvallis, 1968.

Zack, L., Horner, V., & Kaufman, J. Tutoring in a slum school. Elementary School Journal, 1969, 70, 20-27.