American Educational Research Journal
Summer 1982, Vol. 19, No. 2, Pp. 237–248

Educational Outcomes of Tutoring: A Meta-analysis of Findings

PETER A. COHEN,
Dartmouth College
JAMES A. KULIK and CHEN-LIN C. KULIK
University of Michigan

A meta-analysis of findings from 65 independent evaluations of school tutoring programs showed that these programs have positive effects on the academic performance and attitudes of those who receive tutoring. Tutored students outperformed control students on examinations, and they also developed positive attitudes toward the subject matter covered in the tutorial programs. The meta-analysis also showed that tutoring programs have positive effects on children who serve as tutors. Like the children they helped, the tutors gained a better understanding of and developed more positive attitudes toward the subject matter covered in the tutorial program. Participation in tutoring programs had little or no effect, however, on the self-esteem of tutors and tutees.

The tutoring programs offered in many elementary and secondary schools today differ in an important way from yesterday's tutorial programs. In most modern programs, children are tutored by peers or paraprofessionals rather than by regular school teachers or professional tutors. The use of peer and paraprofessional tutors has dramatically affected the availability of tutoring programs. No longer a luxury available only to an aristocratic elite, tutoring programs today are open to boys and girls in ordinary classrooms throughout the country.

Hundreds of teachers and researchers already have written reports on the effects of such programs on children. Although some of the reports are based

The material in this report is based upon work supported by the National Science Foundation under Grant No. SED 79-20742. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation.

on subjective impressions and informal observations and thus are of limited scientific value, other reports describe sound experimental studies of tutoring. In such studies, investigators usually compare the performance of equivalent groups of students assigned to classrooms with and without tutoring programs. Comparisons often focus on learning gains in the two types of classrooms, and sometimes also cover affective growth of tutored and untutored students.

Several major reviews of such studies have appeared in the educational literature in recent years (Devin-Sheehan, Feldman, & Allen, 1976; Ellson, 1976; Fitz-Gibbon, 1977; Rosenshine & Furst, 1969). Each of the reviews concluded that tutoring programs can contribute to the academic growth of the children who receive the tutoring and probably to the growth of the children who provide the tutoring as well. However, two of the reviews (Ellson, 1976; Rosenshine & Furst, 1969) reported that these contributions had been clearly demonstrated only for well-structured and cognitively oriented programs. Since each of the four reviews used relatively informal narrative and box score techniques for summarizing findings, none provided precise statements about the size of the learning gains to be expected from tutoring or about the conditions under which positive effects are likely to occur. For more precise conclusions, more formal review methods are needed.

In 1977 Hartley introduced the use of more powerful review methods into this area. The methodology that she employed, called "meta-analysis," was first described by Glass (1976) in his presidential address to the American Educational Research Association. Meta-analysis is simply the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings. Applying this method to findings on mathematics teaching in elementary and secondary schools, Hartley showed not only that the effects of tutoring were positive, but that they were stronger than those from such other individualizing teaching methods as computer-based instruction, programmed instruction, and instruction with individual learning packages. Hartley also showed that the effects of tutoring were significantly stronger in some situations than in others.

Although Hartley's study advanced knowledge of tutoring considerably, her work was still somewhat limited in scope. First, Hartley's analysis covered only mathematics teaching. Hartley did not determine whether tutoring programs had the same remarkable effects in all subjects or whether such effects were restricted to mathematics teaching. Second, Hartley studied only achievement effects. She did not determine whether tutoring had positive or negative effects on other instructional outcomes, such as attitude toward school, attitude toward school subjects, or on self-concept.

Hartley's analysis also suffered from some possible methodological weaknesses. First, her study aggregated effects on those being tutored and on those providing tutoring. Other reviewers have preferred to look at these as

two distinct types of effects. Second, Hartley's analysis was based on far more findings than studies (73 findings from only 29 studies). The dependence among findings made it difficult for Hartley to determine the amount of error in her statistics. Third, Hartley included in her pool of studies some methodologically inadequate work. Many reviewers would not take seriously studies that lacked a control group.

This article reports results from a meta-analysis that builds on Hartley's work. It is meant to answer several major questions about tutoring. How effective does the typical study say that tutoring is? Are certain types of tutoring programs unusually effective? Is tutoring especially effective for certain types of educational outcomes? What sorts of studies demonstrate the effects of tutoring most clearly? Unlike Hartley's study, this meta-analysis covers studies of different subject areas and describes results for different kinds of school outcomes. It treats separately outcomes for student tutors and tutees and includes only studies that meet reasonable methodological standards.

METHODS

This section describes the procedures used in locating studies, coding study features, and quantifying outcomes of studies.

Sources of Data

The first step in this meta-analysis was to collect a large number of studies that examined effects of tutoring programs on school-age children. We began the collection process by computer-searching three data bases through Lockheed's DIALOG Online Information Service: *ERIC*, a data base on educational materials from the Educational Resources Information Center, consisting of the two files *Research in Education* and *Current Index to Journals in Education*; *Comprehensive Dissertation Abstracts*; and *Psychological Abstracts*. The bibliographies in articles located through the computer searches provided a second source of studies for the meta-analysis. In all, our search yielded a total of more than 500 titles.

We used three guidelines to reduce the initial pool of 500 titles to the final set of 65 studies.[1] To be included in our analysis, studies had to take place in actual elementary or secondary school classrooms. Second, they had to report on quantitatively measured outcomes in both a tutored group and a nontutored control group. And third, studies had to be free from such crippling methodological flaws as different aptitude levels in the comparison groups and unfair "teaching of the test" to one of the groups. Finally, we used guidelines established for our previous meta-analyses (e.g., Kulik, Kulik, & Cohen, 1979a, 1979b) to ensure that each study was counted only once in each analysis.

[1] A complete list of studies used in the analyses described in this article is available from Peter A. Cohen.

Characteristics of Studies

The 65 studies used in this analysis were of many different types. To describe the main features of the studies, we defined 15 variables. Four of the variables described the types of tutoring programs used in the studies: whether the tutoring was structured or nonstructured; whether the tutoring was cross-age or not; whether tutoring was a supplement to or a substitute for classroom instruction; and whether or not tutors received training. The next three variables described aspects of the experimental design of the studies: random versus nonrandom assignment of students to comparison groups; control for teacher effects by using the same teachers for both experimental and control groups; and control of author bias on tests through use of standardized examinations. Six other variables described features of the course setting, including duration of the program, class level of tutors, class level of tutees, subject matter, average ability level of tutees, and level of skills tested on examinations. Finally, two variables described publication features of the studies: the manner of publication of the study and the year of publication.

Study Outcomes

The 65 studies described effects of tutoring programs on both tutors and tutees. These effects were in three major areas: student achievement as measured on examinations, favorability of student attitudes toward the subject matter, and favorability of student self-concept. To quantify the effects of tutoring programs in each of these areas, we used the Effect Size (ES), defined as the difference between the means of two groups divided by the standard deviation of the control group (Glass, 1976). For studies that reported means and standard deviations for both experimental and control groups, we calculated ES from the measurements provided. For less fully reported studies, we calculated ES from statistics such as t and F, using procedures described by McGaw and Glass (1980).

RESULTS

In this section we first describe the effects of tutoring programs on the children who received tutoring. We then turn to the effects of tutoring programs on children who served as tutors in these programs.

Effects on Tutees

A total of 52 of the 65 studies reported results on academic achievement of tutored students; 9 studies reported on self-concept; and 8 reported on attitude toward subject matter.

Achievement. In 45 of the 52 achievement studies, the examination performance of students who were tutored was better than the examination performance of students in a conventional class; in 6 other studies, exami-

nation performance was better in the conventional class; and in 1 study, there was no difference between tutees and conventional students. A total of 20 of the comparisons reported a statistically significant difference in results from the two teaching approaches. Results of 19 of these studies favored tutees, and results of 1 study favored conventional instruction. Clearly, a distinct majority of studies favored tutees.

The average ES in the 52 studies was .40; the standard error of ES was .069. An average ES of .40 means that in a typical class tutoring raised the performance of tutored students by approximately two-fifths of a standard-deviation-unit. Put in another way, the average child in the tutored group scored at the 66th percentile of the students in the untutored or control group. Cohen (1977) has referred to effects of this magnitude as modest in size.

Although the effect of tutoring was modest in the typical study, the size of effect varied from study to study (Figure 1). The largest ES (2.3) came from a study from Mohan (1972). Four other studies also reported large effects

FIGURE 1. Distribution showing the effects of tutoring on tutee achievement in 52 studies.

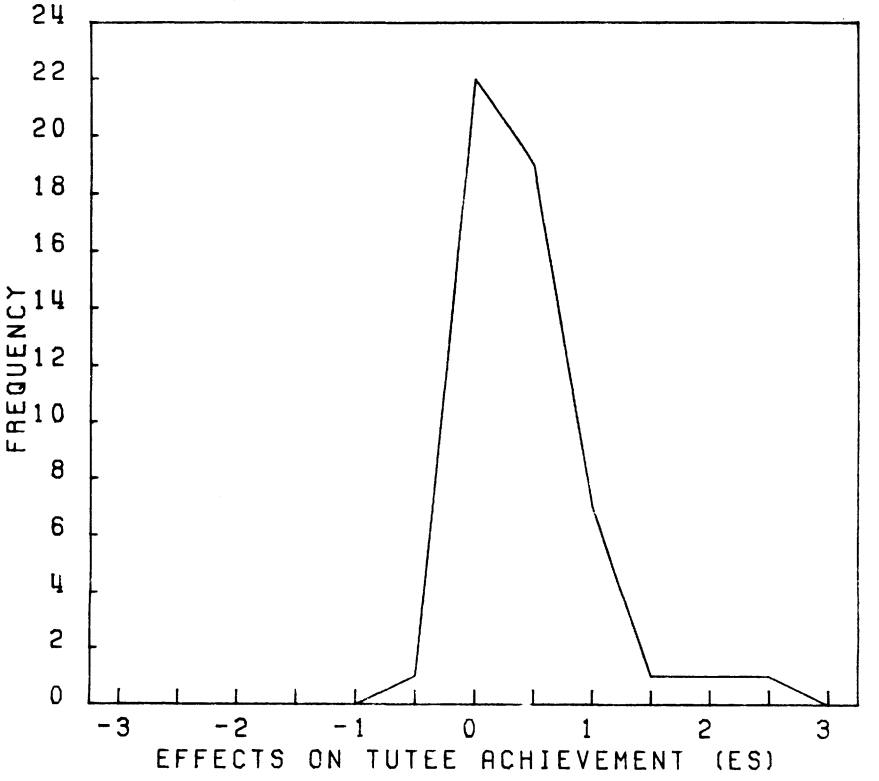

TABLE I

Mean and Standard Error of Effect Size for Tutee Achievement when Studies are Classified in Various Ways

Category of Study	Number of Studies	Effect Size	
		Mean	Standard Error
Implementation			
Substitute	29	.47	.10
Supplement	23	.31	.10
Tutor Training			
No	13	.36	.09
Yes	39	.41	.09
Cross-age Tutoring			
No	24	.29	.07
Yes	28	.49	.11
Structured Tutoring*			
No	23	.26	.06
Yes	29	.51	.11
Random Assignment to Groups			
No	22	.32	.07
Yes	30	.46	.11
Control for Instruction Effect			
Different Instructors	10	.36	.10
Same Instructor	42	.41	.08
Control for Author Bias***			
Instructor Developed Test	12	.84	.21
Commercial Standardized Test	40	.27	.05
Duration of Treatment***			
0–4 Weeks	6	.95	.28
5–18 Weeks	30	.42	.08
19–36 Weeks	14	.16	.09
Class Level of Tutees			
1–3	36	.45	.09
4–6	10	.25	.12
7–9	6	.33	.17
Class Level of Tutors			
1–3	15	.39	.13
4–6	10	.64	.27
7–9	14	.24	.07
10–12	11	.36	.11
Subject Matter**			
Math	18	.60	.16
Reading	30	.29	.06
Other	4	.30	.13
Average Ability of Tutee			
Low	40	.42	.08
Middle	12	.33	.14
Level of Achievement Measure***			
Low-order	16	.76	.17
Mixed	36	.24	.05

TABLE I
(*Continued*)

Category of Study	Number of Studies	Effect Size	
		Mean	Standard Error
Source of Study**			
Unpublished	6	.85	.42
Dissertation	30	.27	.06
Published	16	.47	.12
Study Year			
1961–1965	2	.06	.34
1966–1970	17	.32	.08
1971–1975	26	.45	.11
1976–1980	7	.44	.26

* Significant difference among effect sizes for categories of this variable, $p < .10$.
** $p < .05$.
*** $p < .01$.

(with ES equal to .8 or higher) and 11 other studies reported effects in the medium range (with ES equal to .5 or more but less than .8). The other studies reported small or trivial effects of tutoring on tutees.

Further examination of the data showed that studies with certain features consistently produced strong effects (Table I). In all, six features were significantly related to size of effect. Tutoring effects were larger in more structured programs, and in tutoring programs of shorter duration. The effects were also larger when lower level skills were taught and tested on examinations, and when mathematics rather than reading was the subject of tutoring. Effects were larger on locally developed tests and smaller on nationally standardized tests. Finally, studies described in dissertations reported smaller effects than did studies described in journal articles or in unpublished documents.

Attitude toward subject matter. Eight studies reported results on student attitudes toward the subject matter that they were being taught. In all eight of these studies, student attitudes were more positive in classrooms with tutoring programs. Only one of these eight studies, however, reported an effect large enough to be considered statistically reliable. The average ES was .29; the standard error was .08. Even though the number of studies available was small, results were consistent enough for us to conclude with statistical confidence that tutoring programs had a positive effect on the tutored students' attitudes toward the subject being taught.

Self-concept. Nine studies reported on effects of tutoring programs on tutee self-concept. In seven of these studies, self-concepts were more favorable for students in classrooms with tutoring programs; in the other two studies, self-concepts were more favorable in the classroom without tutoring

programs. The average ES in the nine studies was .09; the standard error was .042. Clearly, this effect was very small and was not large enough to be considered statistically reliable.

Effects on Student Tutors

Of the 65 studies we located, 38 examined achievement effects on tutors; 16 investigated changes in self-concept of tutors; and 5 examined changes in tutor attitudes toward the subject matter being taught.

Achievement. In 33 of the 38 studies investigating effects in this area, students who served as tutors performed better than did control students on examinations in the subject being taught. In the remaining 5 studies, examination scores were better for students not serving as tutors. Of the 38 comparisons, 10 reported statistically significant results, and in each case the difference favored students serving as tutors. The average ES in the 38 studies was .33; the standard error was .09. Only 1 of 11 study features included in this analysis was significantly related to effect size—the kind of subject matter taught in the tutoring program (Table II).

Attitude toward subject matter. In four of the five studies investigating effects in this area, attitudes were more positive among those serving as tutors; in the other study, the students who did not serve as tutors held the more positive attitudes. Only one study showed a statistically significant difference in subject matter attitudes of tutors and conventional students, and this study favored tutors. The average ES for attitude toward the subject was .42; the standard error was .46. Cohen (1977) referred to effects of this magnitude as moderate in size.

Self-concept. A total of 16 studies reported on effects of tutoring programs on self-concepts of students who served as tutors. In 12 studies, self-concept was higher for tutors than for those who did not serve as tutors; in the remaining 4 studies, self-concept was higher for those who did not serve as tutors. Four studies showed statistically significant differences, and in each case the difference favored students who served as tutors. The average effect on tutor self-concept, however, was small. The mean ES was .18; the standard error was .12.

DISCUSSION

The message from the educational literature on tutoring programs seems clear enough. These programs have definite and positive effects on the academic performance and attitudes of those who receive tutoring. Tutored students outperformed their peers on examinations, and they expressed more positive attitudes toward the subjects in which they were tutored. Tutoring programs also had positive effects on children who served as tutors. These tutors not only developed more positive attitudes toward the subjects that

TABLE II

Mean and Standard Error of Effect Size for Tutor Achievement when Studies are Classified in Various Ways

Category of Study	Number of Studies	Effect Size	
		Mean	Standard Error
Implementation			
Substitute	25	.40	.12
Supplement	13	.20	.11
Tutor Training			
No	11	.32	.10
Yes	27	.34	.12
Cross-age Tutoring			
No	11	.28	.08
Yes	27	.35	.12
Structured Tutoring			
No	16	.32	.08
Yes	22	.34	.14
Duration of Treatment			
0–4 Weeks	3	.56	.25
5–18 Weeks	25	.38	.13
19–36 Weeks	9	.10	.09
Class Level of Tutees			
1–3	29	.35	.11
4–6	7	.16	.10
7–9	2	.62	.22
Class Level of Tutors			
1–3	16	.25	.07
4–6	12	.48	.26
7–9	10	.28	.10
Subject Matter*			
Math	11	.62	.22
Reading	24	.21	.05
Average Ability of Tutor			
Very Low	19	.42	.17
Low	11	.23	.07
Middle	8	.25	.12
Source of Study			
Dissertation	29	.25	.04
Published	7	.22	.16
Study Year			
1966–1970	10	.35	.12
1971–1975	23	.33	.13
1976–1980	5	.28	.19

* Significant difference among effect sizes for categories of this variable, $p < 0.05$.

they were teaching, but they also gained a better understanding of these areas.

Tutoring programs apparently have much smaller effects on the self-concepts of children. Neither tutors nor tutees changed in self-esteem as a result of tutoring programs. The literature contains anecdotal reports of dramatic changes in self-concept brought about by tutoring programs, but quantitative studies do not support these reports. Dramatic changes in self-esteem appear to be atypical.

Generally, our results were consistent with findings of other reviewers. A number of reviewers, for example, have reported that structured tutoring programs produce especially strong effects (Ellson, 1976; Rosenshine & Furst, 1969). Although we found that both structured and unstructured programs produced measurable effects, the effects from the structured programs were indeed stronger. Like other reviewers (e.g., Fitz-Gibbon, 1977), we also found that the degree of effectiveness of tutoring programs depended on whether standardized or locally developed tests were used in the evaluations. Effects were stronger when measured with locally developed tests.

Results from our analysis agreed especially closely with results reported by Hartley (1977). She reported, for example, an ES of .6 for tutoring studies in mathematics. Our results were identical for studies of mathematics teaching (i.e., our ES also equaled .6), but we found somewhat smaller tutoring effects in the area of reading instruction. Hartley also reported that some additional factors were related to the size of program effects, including the type of report (dissertation versus public school report) and the type of examination (local versus standardized). Our meta-analysis also showed that these two factors were related to the size of effects.

Reviewers and meta-analysts therefore agreed to a large extent on the facts about tutoring programs. But achieving consensus on the interpretation of these facts may turn out to be a more difficult matter. Relationships that on the surface look straightforward often turn out to be complex on closer examination. This may be the case with some of the relationships between study features and tutoring outcomes noted by the reviewers.

A good example is the relationship between the source of a study and the size of the reported effects. We found in our meta-analysis that journal articles reported stronger findings than dissertations, and this result has been reported frequently in other meta-analyses (Smith, 1980). Selective publication of research results certainly can account for the difference in findings from dissertations and journal articles. But which results—those from journal articles or those from dissertations—should we accept as the more accurate? If the selection process that eventually leads to publication of research is based on the strength of findings, then the least selected results (i.e., those from dissertations) would provide the best basis for estimating the size of tutoring effects. If the selection process is based instead on the quality of

research design, then the most selected research results (i.e., those from journal articles) would provide the best basis for estimates of effect size.

This meta-analysis confirms some things that have long been suspected about tutoring. It shows, as many commentators have suggested, that tutoring benefits both tutors and tutees on both the cognitive and affective levels. In addition, it specifies the average strength of tutoring effects, and it identifies the settings and conditions where effects are strongest. Finally, the meta-analysis raises some new questions about tutoring. It challenges other investigators and reviewers to identify the key variables underlying variation in tutoring outcomes.

REFERENCES

COHEN, J. *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press, 1977.

DEVIN-SHEEHAN, L., FELDMAN, R. S., & ALLEN, V. L. Research on children tutoring children: A critical review. *Review of Educational Research*, 1976, *46*, 355–385.

ELLSON, D. G. Tutoring. In N. Gage (Ed.), *The psychology of teaching methods.* Chicago: University of Chicago Press, 1976.

FITZ-GIBBON, C. T. *An analysis of the literature of cross-age tutoring.* Washington, D.C.: National Institute of Education, 1977. (ERIC Document Reproduction Service No. ED 148 807).

GLASS, G. V Primary, secondary, and meta-analysis of research. *Educational Researcher*, 1976, *5*, 3–8.

GLASS, G. V, & SMITH, M. L. Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1979, *1*, 2–16.

HARTLEY, S. S. Meta-analysis of the effects of individually paced instruction in mathematics (Doctoral dissertation, University of Colorado, 1977). *Dissertation Abstracts International*, 1977, *38*(7-A), 4003. (University Microfilms No. 77-29,926).

KULIK, J. A., KULIK, C.-L. C., & COHEN, P. A. A meta-analysis of outcome studies of Keller's Personalized System of Instruction. *American Psychologist*, 1979, *34*, 307–318. (a)

KULIK, J. A., KULIK, C.-L. C., & COHEN, P. A. Research on audiotutorial instruction: A meta-analysis of comparative studies. *Research in Higher Education*, 1979, *11*, 321–341. (b)

MCGAW, B., & GLASS, G. V Choice of the metric for effect size in meta-analysis. *American Educational Research Journal*, 1980, *17*, 325–337.

MOHAN, M. *Peer tutoring as a technique for teaching the unmotivated.* Fredonia, N.Y.: State University of New York, Teacher Education Research Center, 1972. (ERIC Document Reproduction Service No. ED 061 154).

ROSENSHINE, B., & FURST, N. *The effects of tutoring upon pupil achievement: A research review.* Washington, D.C.: Office of Education, 1969. (ERIC Document Reproduction Service No. ED 064 462).

SMITH, M. L. Publication bias and meta-analysis. *Evaluation in Education: An International Review Series*, 1980, *4*(1), 22–24.

AUTHORS

PETER A. COHEN, Assistant Director, Office of Instructional Services and Educational Research, and Adjunct Assistant Professor of Psychology, 3 Webster Hall, Dartmouth College, Hanover, NH 03755. *Specializations:* Instructional evaluation, research on college teaching, research synthesis.

JAMES A. KULIK, Research Scientist and Associate Director, Center for Research on Learning and Teaching, 109 E. Madison, University of Michigan, Ann Arbor, MI 48109. *Specializations:* Research synthesis, educational evaluation.

CHEN-LIN KULIK, Assistant Research Scientist, Center for Research on Learning and Teaching 109 E. Madison, University of Michigan, Ann Arbor, MI 48109. *Specializations:* Research synthesis, individualized instruction, psychological measurement.