DOCUMENT RESUME

ED 293 959 UD 026 147

AUTHOR Chaikind, Stephen; Sullivan, Helen
TITLE An Analytical Review of the Evidence on Chapter 1 Cost-Effectiveness.
INSTITUTION Decision Resources Corp., Washington, DC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 15 Dec 86
CONTRACT 400-85-1008
NOTE 23p.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Compensatory Education; *Cost Effectiveness; Federal Programs; Program Budgeting; *Program Costs; Program Effectiveness; *Program Evaluation; Research Design; Research Methodology; Research Problems; *Research Reports
IDENTIFIERS *Education Consolidation Improvement Act Chapter 1; Research Results

ABSTRACT

This paper reviews the following studies on the cost-effectiveness of the Chapter 1 compensatory education program: (1) Sustaining Effects Study; (2) Tallmadge Study; (3) the Kiesling Study for Rand; (4) Instructional Dimensions Study; (5) Response to Educational Needs Project Cost-Effectiveness Study; (6) Educational Testing Service/Ragosta Analyses; (7) An EValuation of the Costs of Computer-Assisted Instruction (Levin and Woo); and (8) Recent Cost-Effectiveness Debate. The following components in each study are reviewed: (1) a consistent set of outcome measures from the instructional process; (2) a complete and detailed cost estimate; (3) a well-defined control group; (4) universal or unbiased sample data; and (5) a well-defined methodology. There is no concrete body of evidence that can be said to show that expenditures on Chapter 1 programs are more cost-effective than other instructional practices; there is much diversity among these studies in the focus of the study, the data set and methodology used, and the results obtained. A bibliography is included. (BJV)

5-17-88

ED293959

An Analytical Review of the Evidence on Chapter I Cost-Effectiveness

by

Stephen Chaikind

with the assistance of

Helen Sullivan

DRC
Washington, D. C. 20036

December 15, 1986

This research was conducted under NIE Contract No. 400-85-1008. I would like to thank Mary Moore, Richard Murname, and Mary Kennedy for reviewing this work. The research reported here does not necessarily reflect the opinions of DRC or the Office of Educational Research and Improvement (formerly NIE), U. S. Department of Education.

UD026147

An Analytical Review of the Evidence on Chapter I
Cost-Effectiveness

by
Stephen Chaikind
with the assistance of Helen Sullivan

Introduction and Conclusions

This paper provides a review of selected research pertaining to the cost-effectiveness of the Chapter 1 compensatory education program. The research reviewed here covers many diverse types of analyses, ranging from studies specifically focused on finding the cost-effectiveness of the Chapter 1 (formerly Title 1) program, to those more generally concerned with singular cost or effectiveness issues. This range of studies can be seen, for example, in our reviews of a Title 1 cost-effectiveness analysis prepared for the Sustaining Effects Study (SES), of another multiyear project to assess the effectiveness and cost of a computer assisted instructional program in Los Angeles among students in need of compensatory education, and of a study that determines the most cost-effective compensatory education instructional method among a group of methods already proven effective.

This review indicates that there is no concrete body of evidence that can be said to show that expenditures on Chapter 1 programs are more cost-effective than other instructional practices. While we do not review all of the many studies relating to the cost-effectiveness issue, many of the major studies over the past ten years are covered. Our aim is to show the diversity of focus within these studies, point out the range of (and problems in the) data sets and methodologies used, and the variations in the results obtained. Many of the earlier studies have been reviewed in Is More Better? The Effectiveness of Spending on Compensatory Education, by Stephen P. Mullin and Anita A. Summers (1983). Their conclusions are similar to those of this analysis; that is "...no significant association can be found between dollars spent and achievement gains. No approach and no program characteristic was consistently found to be effective. And those that were identified as effective in specific studies were not necessarily the costlier ones."

Each study reviewed, however, can be said to provide useful information on several of the aspects necessary for a complete cost-effectiveness study. One analysis, for example (see Tallmadge, below), provides good indications of the effectiveness of Title 1 programs in raising test results, while another (Levin and Woo) details accurate cost calculations. As noted, several of the studies show that in certain instances for a very narrow set of instructional approaches and for specific groups of students, compensatory programs will increase educational achievement to a greater degree than would have been obtained without such programs. Yet data and methodological inconsistencies within each analysis mean that few comprehensive definitive conclusions can be stated about Chapter 1's overall cost-effectiveness from these studies.

It is important to emphasize that pointing to the uneven results in the Chapter 1/Title 1 cost-effectiveness literature should not be interpreted by policymakers to mean that the Federal compensatory education program is wasteful or inefficient. Rather, these results indicate that any cost-effectiveness study is sensitive to the underlying assumptions of the researchers, and bound by the conceptual and data problems inherent in such studies; the simplifying assumptions used by the authors can play a crucial role in the results of each study. The sensitivity to the assumptions used can be seen in the debate on the cost-effectiveness of computer assisted instruction (CAI) between Levin and Meister on one hand, and by Niemiec et al. on the other, as summarized below. Differences between the authors' assumptions concerning the appropriate data sources to use in each analysis lead to different conclusions about the cost-effectiveness of CAI programs relative to other treatment methods.

Thus, the conclusions drawn from this review do not imply that there is no cost-effectiveness within the Chapter 1 program. Instead, they indicate a need for a comprehensive, well-designed cost-effectiveness evaluation.

Organization of the Reviews

This section will provide a common context for the study-by-study reviews that follow. Because of the differences in the methodologies, data, and assumptions underlying each of the studies, it is difficult to place equal weight within each review on each of the major concepts necessary to evaluate cost-effectiveness studies. For example, some studies (SES study) were conducted using sufficiently large data bases, but faltered on several methodological points. Others had well defined control groups, but were based on small or localized samples that might not be nationally representative (Tallmadge; Ragosta). Yet another group of studies were based on poor data and questionable analysis (Erracart). Thus, after providing a summary of each study, we focus on what we consider the essential strengths and weaknesses of each, within the confines of analyzing a good cost-effectiveness study. The accompanying chart provides a brief summary of the important aspects of the major studies reviewed.

Each review focuses on those aspects of a good cost-effectiveness analysis most relevant to the particular study under review. A cost-effectiveness study must contain a number of ingredients; we attempt to evaluate the appropriate treatment of these ingredients in each study. A first important component requires that a consistent set of outcome measures from the instructional process be determined--including a well-defined and common outcome metric. The outcome measures in the studies are usually defined in terms of scores on standardized tests; these standardized tests can, however, measure only some of the skills for which various Chapter 1 programs aim at improving (Murnane, 1986, personal communication). Certain benefits of the instructional process that might go unmeasured, such as improvements in students' skills and attitudes, should also be considered in defining outcomes.

Next, a complete and detailed cost estimate must be made. These cost estimates should include not only expenditures of Chapter 1 funds for classroom instruction, but also implied costs of the physical and other human capital used in the instructional process, including opportunity costs, if any. Further, if the intensity of treatments among programs differ, then attempts should be made to identify program participants and tie the costs to individual participants. The intensity of treatment, and hence the level of cost, are not random however; in many programs, children with the greatest learning needs receive more intensive and expensive treatments than those with less need. This "...creates fierce methodological problems for studies that attempt to estimate the effectiveness of spending an extra dollar to finance a more intensive treatment." (Murnane, 1986, personal communication) In addition, the costs measured across programs and among treatment methods should have a large enough variation to permit significant measurement of any impact relating dollars spent to outcomes. It is possible, for example, that Chapter 1 impacts on achievement levels or other output measures might be curvilinear--i.e. achievement results may be large and significant above a certain dollar level of resource inputs, but the required expenditures per participant has not been reached at current Chapter 1 (district) spending patterns to achieve such results.

In addition, a key component of any cost-effectiveness study is a well defined control group. Control groups are standards by which the results of Chapter 1 and non-Chapter 1 instructional programs among groups of students with otherwise similar characteristics (i.e., groups similar in race, family income, initial achievement levels, etc.) can be compared. Because Chapter 1 programs focus their efforts on those most in need of such services, empirically finding a reasonable control group frequently becomes problematic, since students not receiving Chapter 1 services are usually not comparable to those who are.

Another aspect to look for in evaluating cost-effectiveness studies includes analyses that are based on universe or unbiased sample data. Sample data should have a large enough number of participants to enable statistically significant results. One recurrent problem among efforts to determine whether Chapter 1 is cost-effective is the lack of a uniform, national data base to address the issue. Various studies among various regions of the country use data that differ in measures obtained, quality and completeness (Tallmadge; Errecart). Frequently, researchers apologize for the insufficient data (Errecart), and preface their conclusions by warning that the data are weak, so their conclusions need to be interpreted with care.

Finally, each study needs a well-defined methodology. This review indicates the diverse ways researchers chose to measure cost-effectiveness. One method for evaluating cost-effectiveness is to calculate a ratio indicating standardized test score gain relative to dollars spent (see Tallmadge; Levin and Meister; Niemiec et al.) Others look at gains in scores for various programs, and conclude that the most effective ones are also more cost-effective if the outcomes can be bought for the same amount of dollars (Kiesling; Cooley and Leinhardt). Alternatively, other studies examined the statistical relationships between cost and outcomes--frequently using regression analysis--to determine cost-effectiveness (SES).

There is also a great divergence in methodologies in the measurement of effectiveness and costs, and thus in cost-effectiveness. Most of the research analyzing Chapter 1 programs focuses on effectiveness. As noted, effectiveness may be defined as some form of gain in achievement as measured by the scores on a standardized set of tests. Yet there were a wide variety of tests used for such measures in the studies reviewed here. Costs, too, can be calculated completely, including personnel, capital, and other ingredients in the instructional process (see Levin and Woo), or based on incomplete or questionable data (Errecart). Finally, in instances where a cost-effectiveness ratio or measure for one form of instruction can be reasonably calculated (see Ragosta, for example), one may not be able to conclude that the method is cost-effective relative to other methods, since such ratios may not be obtainable for those other methods, or because the cost-effectiveness ratio calculations for other instructional methods were poorly calculated.

A deficiency often noted among studies reviewed here is the lack of focus on the Chapter 1 programs. Many studies look at the effectiveness and costs of compensatory education programs and of programs in districts with a relatively high concentration of Title 1 students (see, for example, Tallmadge; Ragosta; Levin and Woo), but the programs evaluated for cost-effectiveness may or may not be funded by Federal Chapter 1 funds. The SES study is one of the few studies designed to analyze the overall achievement gains relative to costs among students who participated in Title 1 programs compared to those who did not. The SES study's results, though, did not show any cost-effectiveness advantage to Title 1 programs.

Our reviews focus on these key components. No study reviewed here combines all these aspects into a complete cost-effectiveness analysis; the components, though, will give the reader a frame of reference when judging the studies reviewed. In addition, this framework, together with the evaluations of the problems and complexities across a range of diverse methodologies, can be useful in the design of future cost-effectiveness analyses.

Reviews

SES Study A major study of the costs and effectiveness of Title 1 compensatory education programs was conducted by Gerald C. Sumner, Leonard S. Klibanoff and Sue A. Haggart as part of the Sustaining Effects Study. The results of their study (An Analysis of the Cost and Effectiveness Of Compensatory Education, 1979) show no meaningful differences in cost-effectivneess between those groups of students who have received services provided under Title 1 compensatory education programs and comparison groups. The authors however, were not "...quite prepared to conclude that the level of resource utilization has no independent effect on outcome." (p. ix)

The method of analysis employed in this study basically compared the relationship between combinations of inputs to test score gains for different groups of students in various grades. These comparisons were made using both cross-tabulations and regression techniques. The students were grouped in a number of ways, generating many control groups with which achievement gains could be compared. For example, students were classified according to whether they were selected for compensatory education (CE) services or not, whether these services were provided in Title 1 schools, or whether they attended Title 1 schools but did not receive CE services. In addition, several other control groups were created, consisting of students who did not receive CE services but were believed to be in need of them by teachers.

The data were derived from the SES data base, a large and extensive data base for the 1976-77 school year. The sample size included over 95,000 students. Data were used from both a nationally representative survey as well as smaller sub-samples focusing on more specific student and instructional characteristics.

Resource inputs used in the instructional process were quantified by quantity and quality components, with a dollar value (dollar-metric) assigned to each input unit, adjusted for the quality of each input. A single price was assigned to each input based on a sample of districts. Thus, costs are really a resource index, measuring combinations of physical inputs going into the instructional process. Costs reported in this study, therefore, do not vary across geographic regions, nor do they differ as the result of differential wage scales across districts. While such costs are not true costs in the sense that they would indicate Federal funds to higher cost of living regions would yield the same (and presumably less efficient) outcomes, they could indicate that different combinations of inputs, which can be valued in dollar terms, can result in either similar or different outcomes.

Outcomes were calculated as the difference between fall and spring reading and mathematics test scores. These scores were used to compare the relationships between the calculated costs and gains for each of the several sample and control groups across grade levels.

The effectiveness of the program was analyzed, as mentioned, using both cross-sectional and regression analysis. The results show that both the absolute and percentage gains in pretest and posttest scores are not responsive to the level of costs spent on each of the sample groups. Students in CE programs achieved approximately the same relative score increases as attained by those who were not in such programs or by those in the low achiever control group who did not receive CE funded instruction, even though services received by those in the CE group were much more costly (intensive). What the study does show is that low achievers--those most needy in terms of the goals of the CE program--receive the services with the highest dollar-metric relative to all of the control groups. These conclusions were borne by both the cross-tabular analysis and by the regression analysis.

One of the problems with this analysis is that the authors chose not to perform the analysis using a marginal analytical technique, but rather to compare total variable costs for each combination of resources use. Thus, the study shows, for example, that a CE participant increases his reading score by about the same as a student in a control group, but at a much greater total cost for services received (or at a much higher level of resource input). No analysis is shown giving the gains achieved for each additional input of services or dollars spent.

In addition, the regression results offer the same conclusions by testing equations that only test the gain in scores against the total cost and an error term. There is no structural model postulated (which the authors readily admit) where the gains can be related to costs, holding all other social and economic factors thought to affect gains constant as a control. A rough attempt was made at a structural model by performing a stepwise regression using all the variables the authors had on hand, but this attempt was made in order to confirm the authors' conclusion that increases in costs did not explain variations in achievement, and not as a means of creating a structural model which might have proved otherwise. One can only guess whether a well-specified structural model would have shown a significant relationship between increases in costs and gains in achievement.

Tallmadge Study An earlier study--at the State level--that tried to relate the achievement gains in reading and math to Title 1 per-pupil expenditures is G. Kasten Tallmadge's March 1973 report, An Analysis of the Relationship Between Reading and Mathematics Achievement Gains and Per-Pupil Expenditures in California Title 1 Projects, Fiscal Year 1972. The study examines the relationships between either Title 1 or supplementary expenditures for each Title 1 participant and achievement gains for those in schools that are "saturated" with Title 1 students (75 percent or more eligible students in schools), and in "unsaturated" schools. The study concludes that there is some significant relationship between per-pupil expenditures and reading gains in saturated schools, but no such relationship between expenditures and math achievement in saturated schools and no relationship at all in unsaturated schools.

The study was based on data made available to the author by the California State Department of Education. These data included achievement scores by grade collected from schools for saturated schools and school districts for unsaturated schools. Expenditures were available only from schools or districts, but were not disaggregated by grade. Expenditures were applied to grade levels, though, under an assumption that the pattern of per-pupil expenditure variation by grade level is similar from school to school. That is, because grade to grade expenditures were not available across schools, an assumption was made that relative school expenditures by grade are similar across these schools, and should vary in proportion to the variation one finds in expenditures between schools. The author states that this assumption should not bias the results, although assuming expenditures are distributed across grades implies that each grade in each school was indeed served by the Title 1 program--an implication that might not hold in reality. No pupil specific expenditure data were available.

The sample size by grade ranged from 194-321 districts with 116-127 pupils per district for reading projects for grades 1-6, and 501-526 pupils per district for a similar number of districts for math projects. Observations for grades 7-12 were not sufficient for significant analysis, although they are reported in the study. Median pretest and posttest grade equivalent scores w.re used for the analysis.

The data were partitioned by grade level and according to the percentage of students eligible for Title I services Partial and multiple correlations were calculated between both reading and mathem. ics gains and expenditures per-pupil, using a marginal approach that held constant (or "partialled out") the amount of regular per-pupil expenditures and the effects of pretest score differences from the relationship between gains in scores and expenditures. These calculations led the author to conclude that "...(i)f there is a positive relationship between expenditures and gains, it is apparent from the data ...only in reading projects in saturated schools." (p. 27)

The major weakness in the study, as stated by the author himself, is in the limited nature of the data. Some of these data problems include the need to use prior year data for expenditures since current (1972) data were not available, the lack of per-pupil data, the bias in using median grade equivalents test scores, the use of various unstandardized test instruments at different points in time throughout the study, and a small sample size for many grade levels. These weaknesses limited the usefulness of the results. As the author says, "Any study of this type is seriously limited with respect to the scientific rigor which can be brought to bear on the issues." (p. 5)

Kiesling Study Herbert J. Kiesling's 1972 study, Some Esimates for the Cost Effectiveness of Educational Inputs for Reading Performance of Disadvantaged Children in California Title I Projects, took a unique approach to the question of cost effectiveness. Kiesling first found a set of resource inputs that proved effective in producing achievement gains. He then questioned which of the resultant resource inputs were the most cost effective. He concludes that reading specialists, working alone or in combination with paraprofessional assistants, seem to be most cost efficient. He estimates that an additional $300 in expenditures for these resources would bring Title I children close to the national reading gain rate. Instruction in separate facilities and by paraprofessionals aiding classroom teachers had a larger apparent cost-effectiveness, but the results of these two resources were much less statistically significant than the reading specialist resource. Kiesling does not estimate the relationship, though, between actual expenditures and achievement, only the responsiveness of additional resources to probable additional gains.

Kiesling's sample was chosen on a stratified random basis among 6 percent of California's Title 1 projects, enrolling 10 percent of the Title 1 students. The sample was limited to students who took the Stanford Reading Test. Information was collected for four elementary grades from students and teachers. For the selected group of California Title 1 students, achievement is measured only in terms of scores on the single reading exam (test dates unavailable from this paper), with gains usually reported as the additional gain per month per 10 minutes of each of the alternative types of instruction per pupil per week (e.g., 10 minutes of instruction by reading specialists, paraprofessionals, or some combination of teachers and paraprofessionals). Achievement was given in terms of the national norm of the Stanford Reading Test. In addition, cost data were estimated based on California averages, using several assumptions, and were not actual measures for this group of students. In 1971, the author assumes classroom teachers and reading specialists earned $10,000 and $12,000 respectively. Paraprofessionals were assumed to earn $5 per hour. Other assumptions were made for the cost of school construction and depreciation.

The author tested a variety of multiple regression specifications fitting pooled reading achievement data for all pupils in grades 2, 3, 4, and 5, and for grade 3 alone with variables thought to affect such achievement. After determining his "best' specifications, Kiesling estimated the gain in reading scores achieved from each additional $100 spent on each independent variable. Instruction by reading specialists was the most statistically significant input, although the model showed that instruction by reading specialists heavily assisted by paraprofessionals aiding regular teachers added to the gain, but also increased the probability of the added gain occuring by chance. The cost-benefit relationships given in this paper, according to the author, are meant only to be suggestive of actual relationships.

Instructional Dimensions Study While The Instructional Dimensions Study (1980), by William W. Cooley and Gaea Leinhardt, is not a cost-effectiveness analysis of compensatory education programs, the results of the study may be interpreted as having cost-effectiveness implications. The purpose of this study is to identify classroom procedures that are effective in teaching reading and mathematics to disadvantaged children in regular elementary grade classrooms. If superior processes are identified, they can also be assumed to be more cost-effective than less effective processes, for a given amount of expenditures.

The authors selected a sample of 400 classrooms in 100 different schools from 14 school districts in five states in order to study the classroom processes. They identified four separate sets of variables--called constructs--which were thought to explain classroom outcomes. These sets of variables were opportunity, instructional events, motivators, and structure. In addition, initial student performance was thought to affect outcomes. Outcomes were measured by the Comprehensive Test of Basic Skills of the California Test Bureau. Grades 1 and 3 were selected as representative of the elementary grades.

The constructs used in this survey were developed by the authors from a large variety of data, classified into the four categories ultimately studied. Data were collected from three primary sources: interviews with teachers; analysis of curricula; and videotapes of classroom activities. This method of data collection has a number of problems. Incomplete data were problematic for most of the variables, partly because not all classrooms could be videotaped. In addition, combining the variables into four broad constructs required subjective judgments on the part of the researchers; according to the authors, such constructs are not directly measurable among other researchers. It is possible to combine the observed measures into a variety of composites; it appears that the particular combination of measures chosen to represent a specific process can influence the observed significance or insignificance of the given process.

Once constructed, composites were then analyzed using a commonality statistical technique. which is a form of regression analysis that separately tests the impacts of the sets of variables (constructs) on outcomes, so that each process affecting the outcome is counted as influencing the outcomes individually. One problem with such commonality analysis is that it does not evaluate outcomes as potentially occurring as the result of the interaction of a variety of processes.

The results of this study show that opportunity to learn and pretest scores were most significant in explaining reading and math test score gains. No other processes were shown to be statistically significant. Thus, three of the four major processes deduced by the authors could not be shown to affect outcomes. Indeed, as the authors state, "Probably the most important finding of the Instructional Dimensions Study is the absence of clear evidence of the superiority of individualized instruction over other methods of compensatory education." (p. 21) The most direct impact on achievement is shown to be increased reading and math instruction time, based on the author's interpretation of the significance of the opportunity construct. The implication is that increasing reading and math instruction could be the most cost-effective means of improving outcomes in compensatory education programs

The RENP Cost-Effectiveness Study Michael T. Errecart (Is RENP a Cost Effective Supplement to the Regular DCPS Program?, 1978) investigated whether the Response to Educational Needs Project (RENP) in the District of Columbia is a cost-effective means of improving the reading and mathematics achievement scores of students. The purpose of the study was to design a system that identifies the most cost-effective program to improve such scores. Hence, the objective was to determine whether RENP improves student achievement and, if so, whether this improvement is significant enough to justify the additional costs of RENP. Because of substantial problems with the data, indications of both the influence of the RENP program on outcomes or cost-effectiveness cannot be stated with any degree of statistical reliability.

Student performance measures focused on changes in scores on the Comprehensive Test of Basic Skills (CTBS) for students in reading and math programs at the 4th, 5th, 6th, and 8th grade levels in the District of Columbia Public School system. The RENP program was provided in 10 laboratories for each subject area; the control groups were presumably all the other students who did not take part in RENP programs.

The CTBS was administered three times during the years covered by the study. Changes in student performance could be measured by several methods. A raw gain score (RGS) measure, the difference in test scores for students between the test administrations, was the measure used by the author because of the lack of complete data for other measures. The adjusted gain score (AGS), a measure of gain in CTBS scores between the fall of 1976 and the spring of 1977 in adjusted units, would be a better measure, but it could not be calculated for several of the analyses. Additionaly, the school's percentile score change (PC) did not permit the calculation of measures of dispersion, and hence the author was unable to determine if changes in percentile ranks were statistically significant.

Analysis of variance techniques were used to test the significance of the changes in the test scores. The only reliable results where statistical significance can be determined are those using the RGS measure. In an analysis of score changes between the fall 1976 and 1977 test administrations, the differences in RGS scores were only significant at the 8th grade level in reading, and at the 5th grade level in math. The 8th grade RENP students did better than non-RENP students in reading, while the 5th grade non-RENP students attained higher average gains in math than the RENP students. In a fall to spring analysis, the difference in average gains was significant only at the 5th grade level. Here, RENP students did better in both reading and mathematics than the non-RENP students.

Several results for analyses using the AGS were reported separately due to the differences between this measure and the RGS measure. Analyses of variance were conducted using the AGS as the dependent variable. The AGS analyses found a few significant RENP effects in reading comprehension and mathematics, and interaction effects in reading vocabulary and mathematics computation.

Cost calculations were poorly documented in the available paper. The author estimates that $328 and $301 in total resources, "more or less," were targeted at students in reading and math in 1977, although he also indicates that the District of Columbia Public Schools provided approximately $1,237 per class for RENP participants, and $1,187 to each non-RENP student. He does not reconcile these differences.

Two statistics were then analyzed in estimating cost effectiveness: the average cost per student divided by average percentile gain ($/percentile); and the average cost per student divided by the average change in raw score ($/point). Low ratios indicate more cost-efficient gains than those associated with higher ratios. Cost-effectiveness results were frequently not statistically significant, and in many cases were inconsistent across the fall to fall, and fall to spring testing periods. Because many of the measures in both the numerator and denominator in these cost-effectiveness ratios are weak, little confidence can be placed in these results.

The ETS/Ragosta Analyses The final report of Computer Assisted Instruction and Compensatory Education: The ETS/LAUSD Study (1983) by Marjorie Ragosta et al., which was conducted in conjunction with the Los Angeles Unified School District (LAUSD), details an experimental design that specifically set out to determine the cost and effectiveness of one form of instructional program to increase achievement among compensatory students. The project placed computers and appropriate software in four schools in one Los Angeles school district. Selected groups of students were chosen to receive one of three types of drill and practice computer assisted instruction (CAI) for up to 20 minutes per day. The authors show that the computer assisted curricula were largely effective in raising students' standardized test scores in mathematics, reading, and language arts, as well as in raising scores on tests derived from the CAI curriculum compared to their control group. The two control groups in this experiment were well defined. They consisted of students in alternate grades who did and did not receive CAI, and students randomly assigned to one or two of the CAI curricula, but not to the others. These results were replicated over a four year period. Each 10 minute session in 1977 was calculated to cost approximately $130 per student.

Although one of the goals of this study was to determine the cost-effectiveness of a compensatory education curriculum, the researchers could not draw any conclusions about the cost-effectiveness of CAI because there were no equivalent data concerning the per unit cost-effectiveness of other intructional approaches for compensatory students, such as reducing class size or peer tutoring. Costs for CAI were calculated based on the total cost of providing such instruction. These costs included personnel costs, building and maintainence costs, and software and hardware costs. The original cost estimates used for the evaluation of CAI (approximately $130 per session per child) were developed by Henry Levin and Louis Woo in An Evaluation of the Costs of Computer-Assisted Instruction (see below). (Levin and Gail Meister, in Is CAI Cost-Effective, have since brought these estimated costs up to date and have been able to find comparison costs of other instructional approaches.) We review these findings below.

The effectiveness of CAI in the LAUSD study on increasing achievement of students was measured by an estimated "treatment effect." This treatment effect was based on a regression analysis that adjusted several achievement outcomes (the dependent variables based on the outcome test measure used) for pretest scores, sex, ethnicity and classroom differences (the independent variables). The treatment effect was standardized, so that it could be interpreted as the difference in achievement growth that CAI induces over norms within control groups. These treatment effects are expressed in standard deviation units, and were measured over a one, two and three year period. For example, one of the study's results show that the mean standardized treatment effect on the Comprehensive Test of Basic Skills in Computation was .56 two years after the study began. This can be interpreted to mean that students receiving mathematics CAI, on average, were .56 of a standard deviation higher than other students in mathematics computation at the end of two years.

A problem with the study is that it did not appear to be well focused specifically on Title 1 or any other well defined group of students needing compensatory education. Two of the sample schools receiving computer services were Title 1 schools, two appeared not to be based on the descriptions in the report. The geographic area of Los Angeles (area 4) in which all of the schools were located, however, appeared to have a relatively high concentration of Title 1 students. The number of students sampled over the life of the study could not be determined from the available documentation. Thus, the effectiveness measures may not be considered as those that would result if compensatory education students were the only students involved in the study. In addition, cost calculations were generic to the equipment and software, not to the students affected.

The software involved in the study was of the practice and drill type provided by an independent vendor. The mathematics module appeared to have greater depth and breadth than the reading or language arts modules. The lack of depth in the latter two modules created some problems for the study, in that certain students who had excellent English skills finished the modules before the end of the program. In addition, all three programs required the ability to read well. Thus, non-English speaking, non-reading, and limited English speaking students were excluded from the study. (The basis for this exclusion is not established in the report; presumably these excluded students would be eligible for compensatory education.) Thus, the instructional software appeared to be either too easy or too hard for many of the potential users, based on their English reading ability.

This study, then, has shown that one form of instruction (CAI) is effective in raising various student achievement scores for a population comprising a large percentage of students in need of compensatory education, that this effectiveness is maintained over a continuous period of use (over the three year period reported in the study), that it is effective across various curricula, and that the instruction is cost feasible (e.g., affordable to Title 1 schools based on their allocated funds). It has not, however, shown CAI to be cost-effective relative to other instructional programs, nor as being more or less effective for compensatory students as opposed to non-compensatory students.

Levin and Woo The study commissioned by the Ragosta/LAUSD project to evaluate the costs of CAI was authored by Henry M. Levin and Louis Woo (An Evaluation of the Costs of Computer-Assisted Instruction, 1981). As mentioned above, this study cannot draw conclusions about the cost-effectiveness of computer-assisted instruction under Title 1, but attempts instead to estimate the costs for replicating the ETS/LAUSD system of computer-assisted instruction (CAI) in other educational settings, and to evaluate such costs under different organizational arrangements. The study estimates both the costs and the cost feasibility of implementing a particular CAI approach for compensatory education purposes. It is a good example of a comprehensive way of measuring costs.

Levin and Woo use an "ingredients approach" to estimating the cost of the LAUSD CAI program. The first step in this approach lists all ingredients necessary to implement such instruction. Second, they estimate the costs for each ingredient using actual costs or market values. Finally, they convert costs into categories appropriate for analysis (annualized, average, or marginal costs).

The authors divided the CAI program into six categories: facilities and equipment, training, personnel, curriculum rental, maintenance, and miscellaneous. Facilities and equipment include computers, terminals, and printers, as well as the cost of the facility and costs to renovate classrooms. Training is divided into direct and indirect costs. Direct costs include salaries and the costs of resources. Indirect costs are equal to the value of trainees' time. Personnel inputs include administrative personnel, CAI coordinators, teaching aides, and any substitutes needed. The curricula is rented from the Computer Curriculum Corporation (CCC). System maintenance costs only apply to the maintenance of the equipment. Finally, miscellaneous costs include insurance, supplies, utilities, and facility maintenance. All ingredient costs were converted into annualized costs; in 1977-78, the annualized cost of providing a 32-terminal classroom with the CCC A-16 system was approximately $100,000.

Once these annualized costs were determined, the authors calculated the average cost per session of computer-assisted instruction. The calculation of the average cost per session is the key to determining the cost feasibility of the CAI approach. The average cost per session depends on the number of sessions per day. This, in turn, depends upon the length of the sessions, as well as time spent between sessions, preparation for sessions, and equipment maintenance. The total number of daily sessions in the LAUSD experiment were between 21 and 25 per day, with an average of 23. The authors calculated an annual cost per daily session for each variation: 21, 23, and 25 sessions. With a median of 23 sessions a day per terminal, the total number of sessions per year equaled 736. By dividing this number by the estimated annual total cost for a CAI program, the authors found that a 10 minute daily session offered 23 times a day costs about $136 per session. Estimates of costs were also calculated for a model where two schools share an A-16 system. Under this model, with the same arrangements of a 32-terminal A-16 approach, the costs per session increased by 40 percent to $192 per 10 minute session. These costs, it should be noted, are those calculated per hypothetical session, not the actual costs per student participating in this specific experiment.

After they calculated the per session costs, Levin and Woo investigated two questions. The first concerns whether the costs of the program are feasible. Since funds for special education services for disadvantaged students are normally limited to special categorical aid, such as Chapter 1 funds, it must be determined if CAI can be provided within these budgets' constraints. If it can, according to the authors, the program is cost feasible. Levin and Woo show that, in fiscal year 1977, approximately $400 was provided per student for the Title I program. While these funds were not allocated solely for classroom instruction, the authors assumed that the $400 per student represented the maximum amount potentially spendable for compensatory education in the classroom setting. Based on this assumption, they stated that there would be enough funds to provide three daily CAI sessions at $136 per session with a 32-terminal classroom. Hence, the authors could conclude from these figures that the A-16 CAI system is cost feasible within the present allocations for compensatory education.

The second question concerns the issue of whether CAI is cost-effective. A program is relatively cost-effective if the benefits (expressed in terms of a common metric) derived from its methods are greater than those of other alternatives per unit of cost. They ask the question: can CAI benefit these students at costs that do not exceed the costs of other instructional alternatives? Because the authors could not obtain both cost and outcome measures for other alternatives to the CAI program, they could not state whether the CAI approach is cost-effective.

Recent Cost-Effectiveness Debate In a recent article, Is CAI Cost-Effective? (Phi Delta Kappan, June 1986), Henry Levin and Gail Meister summarize the most current information concerning the cost-effectiveness of CAI. While their evidence does not specifically refer to compensatory education programs, it does build upon the LAUSD study's results, which were predicated on CAI provided to compensatory education students. They conclude that CAI is relatively cost-effective, but may not be the most cost-effective instructional approach. These conclusions were questioned by Richard P. Niemiec et al., in CAI Can Be Doubly Effective (Phi Delta Kappan, June 1986).

Levin and Meister's study updates the Levin and Woo paper reviewed above. The earlier study investigated only the total (per student) costs and cost feasibility of providing a CAI program. That study had problems with analyses, data, and the availability of information on other alternative compensat · education instructional methods by which to compare CAI's cost-effectiveness. The . .rent analysis does compare the costs and cost-effectiveness of CAI to those of three other interventions.

These three interventions are: cross-age tutoring, reduced class size, and longer school days.

The CAI system for which costs were calculated was the same system used in the previous study of the LAUSD experiment. In the more recent Levin and Meister study, however, the prices used for computers, software, and maintenance were updated to 1984 prices. The prices for school personnel, facilities, and other resources were, however, estimated at 1980 prices due to data limitations, resulting in slightly understated costs due to the merging of the latest hardware costs with the lower, 1980 personnel costs.

Levin and Meister then calculated the costs for each of the various alternatives, all in 1980 dollars. For each intervention, the cost per student per subject includes the total value of the ingredients necessary to reproduce each intervention for either reading or mathematics, divided by the number of students. Cross-age tutoring, the first of the alternative intervention methods, uses adults or older students as tutors. Data for this method were derived from the Cross-Age Structured Tutoring Program for Reading and Mathematics used in the public schools in Boise, Idaho. Daily tutoring sessions using this method lasted for 20 minutes. Each school had approximately 60 students tutoring 60 younger students and about 26 older students being tutored by adult tutors. Adult tutoring was found to have the highest unit cost among all alternatives. This option was more expensive than both peer tutoring and reducing the class size from 35 to 20 students per teacher. The cost for CAI was only half that of peer tutoring.

The second option, increasing instructional time, added one hour of instruction per day, equally divided for reading and mathematics instruction. Data for analyzing this approach were derived from the Beginning Teacher Evaluation Study (BTES) sponsored by NIE. Finally, data on reducing class size were based on a meta-analysis by Gene Glass and Mary Smith of 80 evaluations concerning the effects of class size on student achievement in reading and mathematics at the elementary school level. Increasing instructional time or reducing the class size from 35 to 30 students per teacher had the lowest unit costs of the options studied here.

Each option's effectiveness was estimated by measuring the effect of each in standard deviation units. One standard deviation unit is approximately equal to one academic year or 10 months of achievement. The effectiveness results are reported in terms of months of additional student gain in each subject area. The CAI approach produced more than one month of student gain in mathematics and over two months in reading. Tutoring produced a gain of one full year in mathematics and approximately one-half year in reading. Reducing class size was not as effective, producing less than a month's gain in both subject areas for each five student per teacher decrease in class size. Finally, adding one half hour of instruction in each subject area resulted in only very small gains.

The authors then compared these unit costs to unit effectiveness to determine the relative cost-effectiveness of each interventioı strategy. A cost-effectiveness ratic was calculated; this ratio can measure (estimate) the expected gain in achievement against its cost. The cost-effectiveness ratio shows the educational effectiveness of each intervention in months of additional achievement gain per year of instruction for each $100 spent per student. CAI, for example, will produce two months in reading and one month in math for each $100 spent per student. Peer tutoring results in about one-half year of gain in math and one-quarter year in reading for each $100. These are the two most cost-effective intervention strategies. Other interventions show lower cost-effectiveness ratios than these, ranging from less than a month to one and one-half morths of gain per $100. As noted above, adult tutoring results in one of the largest educational effects, but produces a cost-effectiveness ratio that is among the lowest of the four options because of its high per unit cost.

The CAI intervention was found to be more cost-effective than adult tutoring, reducing class size, or increasing instructional time. It was, however, less cost-effective than peer tutoring in both math and reading. Hence, the authors concluded that CAI is a relatively cost-effective intervention, but it is not necessarily the most cost-effective approach to improving student achievement in reading and mathematics.

The methods on which these conclusion were based have been criticized by Richard P. Niemiec, Madeline C. Blackwell, and Herbert J. Walberg, in CAI Can Be Doubly Effective (1986). They indicate that there may be two main data problems in the Levin and Meister analysis: the cost data used were not as up-to-date as they might be, and the outcome data were derived from studies that may not be nationally respresentative.

Niemiec et al. were concerned with the accuracy of Levin and Meister's cost estimates, which showed per pupil costs of $119 for CAI and $212 for peer tutoring per year. These estimates were based on 1984 prices for computer costs and 1980 prices for all other ingredients. Niemiec et al. criticized this approach because it assumes that 1980 costs would be appropriate for estimating costs of a program in 1986. Other assumptions would produce different cost estimates. In addition, more recent data might allow the costs for the alternative instructional methods to be based on the use of rapidly improving and more efficient software and hardware. For example, better computer software might enable teachers to spend less time with students; in addition, teacher aides, who are less costly, could be used as substitutes for teachers in providing CAI. This could reduce the labor costs of CAI, which comprise one of CAI's highest component costs, indicating a better cost effectiveness.

Levin and Meister's estimates of the effects of CAI and peer tutoring were also criticized for using data for other instructional programs that may not be nationally representative. As with the cost estimates, Niemiec et al. noted that if the underlying assumptions relating to the data were changed, the cost-effectiveness ratios could also change. For example, the estimates of peer tutoring used by Levin and Meister were based on an undated, unpublished study of a Boise, Idaho tutoring program. They found approximately one-half year's gain in reading and a full year's gain in math, for a combined effect equal to a seven month gain in achievement. Based on these numbers, Levin and Meister concluded that peer tutoring was effective, and based on its costs, found it to be the most cost-effective option. Niemiec et al. argue that this conclusion is based on only one localized study. An alternative tutoring study, by Cohen, Kulik, and Kulik (<u>Educational Outcomes of Tutoring: A Meta-Analysis of Findings</u>, 1982), found lower outcome gains based on a meta-analysis of 65 independent evaluations. Using these data, Niemiec et al. estimate gains of six months in math and two months in reading. The combined achievement gain was four months; substantially lower than the seven month gain predicted by Levin and Meister. Similar criticisms of Levin and Meister's use of the LAUSD CAI data were made; they were in one school district and used software supplied by one vendor. Niemeic et al. computed CAI effectiveness from a quantitative synthesis of many CAI programs.

As a result of these different estimates of effectiveness, Niemeic et al.'s cost-effectiveness ratios differ from those of Levin and Meister's. They concluded that peer-tutoring is twice as cost-effective as CAI; in fact, they found it was the most cost-effective program of all the interventions. Niemiec et al. found the opposite of Levin and Meister, their estimates result in the conclusion that CAI, not peer-tutoring, is the most cost-effective intervention.

Bibliography

Cooley, William and Gaea Leinhardt. The Instructional Dimensions Study: Search for Effective Classroom Processes, Final Report, Pittsburgh, Pa., The University of Pittsburgh Learning Research and Development Center, August, 1978.

Educational Testing Service and RMC Research Corporation, A Study of Compensatory Reading Programs, Mountain View, California: OPBE, ED, September, 1976.

Errecart, Michael T., Is RENP a Cost Effective Supplement to the Regular DCPS Program?, Mountain View, RMC Research Corporation, February, 1978.

Kiesling, Herbert J., Some Estimates For the Cost-Effectiveness of Educational Inputs for Reading Performance of Disadvantaged Children in California Title I Projects, Santa Monica, California, Rand Corporation, March, 1972.

Levin, Henry M., Cost Effectiveness, A Primer, Sage Publications, Beverly Hills, 1983.

Levin, Henry M. and Gail Meister. 'Is CAI Cost-Effective?", Phi Delta Kappan, June, 1986.

Levin, Henry M. and Louis Woo, An Evaluation of the Costs of Computer-Assisted Instruction, Economics of Education Review, Winter, 1981.

Michigan State Department of Education, Report of the 1974-75 Michigan Cost-Effectiveness Study, Michigan Department of Education, Research, Evaluation, and Assessment Service, 1976.

Murnane, Richard J., personal communication commenting on draft of "An Analytical Review of the Evidence on Chapter I Cost-Effectiveness." August, 1986.

Mullin, Stephen P. and Anita A. Summers, Is More Better? The Effectiveness of Spending on Compensatory Education, Phi Delta Kappan, January, 1983.

Niemiec, Richard P., Madeline C. Blackwell and Herbert J. Walberg, "CAI Can Be Doubly Effective," Phi Delta Kappan, June, 1986.

Pogrow, Stanley, Shifting Policy Analysis and Formation From An Effectiveness Emphasis to a Cost Perspective, Educational Evaluation and Policy Analysis, Spring, 1983.

Ragosta, Marjorie, Computer-Assisted Instruction and Compensatory Education: A Longitudinal Analysis, Machine-Mediated Learning, Vol. 1, No. 1, 1983.

Rossi, Robert J. et al., Summaries of Major Title I Evaluations, 1966-1976, Palo Alto, California, American Institutes for Research, July, 1977.

Schweinhart, Lawrence J. and David P. Weikart, Evidence That Good Early Childhood Programs Work, Phi Delta Kappan, April, 1985.

Summers, Anita A and Barbara L. Wolff, Improving the Use of Empirical Research as a Policy Tool: An Application to Education, Universities of Pennsylvannia and Wisconsin, Institute of Poverty, Madison, Wis., Discussion Paper, 1979.

Sumner, Gerald C., Leonard S. Klibanoff and Sue A. Haggart, Report #7, An Analysis of the Cost and Effectiveness of Compensatory Education, System development Corporation Santa Monica, California, August, 1979.

Tallmadge, G. Kasten, Analysis of the Relationship Between Reading and Mathematics Achievement Gains and Per-Pupil Expenditures in California Title I Projects, Fiscal Year 1972: Final Report, Palo Alto, California, American Institute for Research in behavioral Sciences, 1972.

Wang, Ming Mei, Evaluating the Effectiveness of Compensatory Education, Santa Monica, California, System Development Corporation, April, 1980.