Ross, J. A., Barkaoui, K., & Scott, G. (2007). Evaluations that consider the cost of educational programs: The contribution of high quality studies. *American Journal of Evaluation, 28*(4), 477-492.

Evaluations that Consider the Cost of Educational Programs

The Contribution of High Quality Studies

Abstract

Cost studies are program evaluations that judge program worth by relating program costs to program benefits. There are three sets of strategies: cost-benefit, cost-effectiveness, and cost-utility analysis, although the last appears infrequently. We searched relevant databases to identify 102 cost studies in education and then reduced the set to 30 using eight criteria. We found that these studies contributed to understanding program effects by meeting the four evaluation purposes identified by Mark et al. (2000). Cost studies (1) provide evidence of the worth of educational spending at the macro and individual program levels, information that is not provided by other evaluation approaches; (2) they provide direction for program improvement that differs from recommendations based solely on effect sizes; (3) they contribute to knowledge development by constructing and testing models that link spending to student learning; and (4) they can lead to the rewriting of regulations to make programs more efficient.

Keywords: Cost studies; cost-benefit; cost-effectiveness; cost-utility

Evaluations that Consider the Cost of Educational Programs

The Contribution of High Quality Studies

Resources for education are limited and are constantly threatened by other public needs (especially health) and demands for tax relief. Although cost studies are of growing importance in health evaluations, educational evaluators have shown little interest in addressing questions about whether the programs they evaluate are worth the money these programs cost. Most evaluation textbooks (with the exception of Rossi, Lipsey, & Freeman, 2003) give cursory attention to cost analysis; the discussion of cost and fiscal propriety in the Standards (Joint Committee, 1994) refers to the costs of the evaluation not the program; and the Guiding Principles (AEA, 2000) make no reference to examining program costs. For many educators, the demonstration of a statistically significant effect by a rigorously controlled study is sufficient to recommend program continuance. For example, the standards for inclusion in the knowledge base represented by the What Works Clearinghouse (n.d.) funded by the U.S. Department of Education address only internal validity issues. Yet a rank ordering of programs by effect sizes may not match an ordering of the same programs by cost-effectiveness. The lack of attention to cost considerations impedes the development of a knowledge base of what works in education, which impedes educational policy formation.

Our primary purpose in this article is to argue that evaluators rarely examine the costs of educational programs, that their failure to do so limits our ability to develop a knowledge base of what works in education, and that examination of costs improves our knowledge of educational program effects. To support this argument we will identify the range of cost studies that have been completed, assess their quality, and describe their contribution to our understanding of the effects of educational programs. Many educators view cost studies as unhelpful because the field

is framed by the controversial meta-analyses of Hanushek (e.g., 1997) which found that student achievement is not affected by variation in educational spending. Hanushek's methods have been challenged at multiple levels and re-analyses have argued that greater spending on education leads to higher student achievement than lesser spending (Greenwald, Hedges, & Laine, 1996; Kreuger, 2003; Verstegen & King, 1998). In our review we will focus on studies that investigated the effects of educational spending on achievement that produced findings of greater utility than the gross question of whether spending matters. We will also examine other categories of cost studies that generated findings with important policy implications.

Our secondary purpose for the review is to identify criteria for determining the quality of cost studies. Although our application of these criteria will be limited to cost studies in education, we will argue that these criteria are equally useful for assessing the quality of cost studies in other domains.

Definitions

In this article we define cost studies as a program evaluation approach that judges program worth by relating program costs to program benefits. In education, two methodological strategies dominate: (1) Cost-benefit analysis is a method in which the costs and benefits of a program are converted into monetary units; the program is worthwhile if its benefits outweigh its costs (Levin & McEwan, 2001). (2) Cost-effectiveness analysis is similar, except that outcomes are not converted to monetary values. The evaluator determines cost-effectiveness by measuring program outcomes while holding costs constant or calculating costs while holding outcomes constant. A third approach, cost-utility analysis, in which the costs of a program are related to stakeholder perceptions of its value, appears rarely in educational cost studies and will be only briefly discussed.

Meta-Evaluative Stance

By improving understanding of the effects of programs on stakeholders, evaluation should ameliorate social conditions, develop human capital, and reduce problems, be they chronic or acute. Following Mark, Henry and Julnes (1999; 2000) we assessed the contribution of cost studies in terms of four purposes: (1) assessment of merit (i.e., the extent to which the evaluation generates warranted judgments of the value of programs to individuals and society); (2) program improvement (i.e., the evaluation's contribution to the enhancement of the program); (3) oversight and compliance (i.e., generation of warranted evidence of the program's observance of policy requirements); and (4) knowledge development (i.e., discovery or testing of general theories).

Method

First Level of Search and Coding

We began our search for cost studies with the knowledge that previous reviewers (Hummel-Rossi & Ashdown, 2002; Levin, 2001; Levin & McEwan, 2001; Monk & King, 1993) had found few examples in education. We searched educational (ERIC, ProQuest, WilsonWeb, PsychINFO) and economic (EconLit, EconLibrary, PolicyFile) databases, using keywords such as cost-effectiveness, cost-utility, value for money, cost estimates, cost-benefit, and educational economics with screens such as elementary or secondary or post-secondary and school or education. We extended the search with manual branching and examination of the contents of selected journals (e.g., Educational Evaluation and Policy Analysis). We read 643 abstracts, selecting 102 studies for review.

Five reviewers read the full texts of the 102 studies and applied 18 codes.[1] The codes, shown in the Appendix, focused on the design of the study, its purpose and theoretical

framework (codes 1-3), the sample (4, 5), the variables and their measurement (6-8), procedures for assessment of costs, benefits and their relationship (9-11), application of the Standards for Program Evaluation (13-17) and a summative judgment of the benefits of considering costs in the evaluation. Each review consisted of 1-9 (mode=4) single-spaced pages of notes. A narrative review of the 102 studies is available from the senior author.

Second Level of Search and Coding

For this article, we further filtered the 102 cost studies using eight criteria to distinguish high from low quality. We first applied the eight criteria (described below) to the 400+ pages of notes generated from the first level. The second and third authors read each of the reviews from the first level. Differences in their coding were resolved through discussion, examination of the full text of the study, and appeal to the senior author.

Cost studies combine two sets of criteria: rigor in determining program effects, based on quantitative reasoning, and meticulous assessment of costs, based on cost-benefit and cost-effectiveness manuals (e.g., Levin & McEwan, 2001). The first three criteria appraised the credibility of the study's claim about program outcomes; the remainder addressed the credibility of the study's treatment of financial data.

Evaluation Design. We distinguished four types of designs based on their internal validity, following typical design hierarchies (e.g., Posavac & Carey, 2007) and giving special place to the true experiment as the optimal design for developing evidence-based practice (Slavin, 2002). We coded evaluations that lacked credible comparison groups as 0 and deleted them from the review. *Correlational designs* used regression, path modeling or hierarchical linear modeling to distinguish the effects of programs from the effects of other variables. *QED* designs were quasi-experimental studies that used statistical adjustment to equate nonequivalent

groups; the most common design of this type was a pre-post comparison of groups that had not been matched or randomly assigned. *Experiments* consisted of studies with random assignment of participants to experimental conditions or used matched samples. We included in this category population studies in which cost and benefit data were applied to an intact population, eliminating the need for inferential statistics.

Population. Evaluations were coded as having an *identifiable population* if the authors drew a random sample from a defined population or presented evidence that the sample represented a known population. If not, we coded the study as having an *unknown population* but retained it in the review.

Database Quality. This criterion considered the reliability and validity of instruments (e.g., student achievement) used to generate the data. However, a large proportion of the evaluations drew upon census data or other institutional databases without providing any information about the quality of that database. When such databases were the only data source and no psychometric information was provided for them, we coded the study as *high quality* if the database was national or provincial in scope and *low quality* if based on school district data. In both cases we retained the study for the review.

Adjustment of Financial Estimates for Time. In cost studies, benefits, costs or both may extend beyond a single year. Higher quality studies adjusted money values for time, using an explicit discount rate. We coded studies as *yes* or *no* on this criterion but retained both types in the review.

Sensitivity Analysis. Cost studies require assumptions about cost and benefits as well as economic conditions. Higher quality studies, coded as *yes* on this criterion, made their assumptions explicit and replicated the analysis by testing alternate economic functions (e.g.,

Behrman & Birdsall, 1983) or alternate models of input variables (Card & Krueger, 1992); they vared strategies for calculating costs (e.g., real versus nominal costs in Dolan & Schmidt, 1987) or varied discount rates (Krueger, 2003). We coded studies that conducted no sensitivity analysis as *no* on this criterion but retained these studies in the review.

Cost Rationale. Evaluations that provided data on benefits but not on costs were coded as 0 and deleted from the review. Higher quality studies, coded as *with rationale*, provided an explicit rationale for cost calculations, for example, using the ingredients method (Levin & McEwan, 2001) in which a program is decomposed into specific components with the costs of each summed to produce a total cost. Studies that included detailed costs but provided no information about how the costs were compiled were coded as *no rationale* and retained for the review.

Benefits Rationale. Evaluations that provided data on costs but not benefits were coded as 0 and deleted from the review. Studies were coded as *public*, indicating the study provided information on how benefits to the community were calculated or *private*, indicating that the study described how benefits to individuals were determined. We treated increases in cohort achievement as public benefits. Higher quality studies included a rationale for the calculation of public and private benefits.

Decision Rule. Evaluations that had no explicit procedure for relating benefits to costs were coded as 0 and deleted from the review. Cost-benefit studies were coded as *CB*. In these studies the program was worth the money if: (i) the benefit-cost ratio was greater than 1.0 (i.e., for every dollar invested in the program there was at least one dollar of social benefit); (ii) the net present value of the program was greater than 0 (net present value is the result of subtracting benefits from costs after discounting both) or (iii) the internal rate of return exceeded the market

rate or some other socially acceptable rate of return (the internal rate of return is the discount rate that causes the net benefit of the program to equal 0). Cost-effectiveness studies were coded as *CE*. In these studies the program was worth the money if: (i) the program had greater effectiveness for the same cost than competing programs or the status quo; (ii) the program produced the same effectiveness as competing programs at lower cost; (iii) the program had a greater cost-effectiveness ratio than competing programs when effectiveness was divided by costs for the program and its competitors. We found only one example of a cost-utility study. It was coded as *CU*. In cost-utility studies the program was worth the money if the cost-utility ratio of the program was lower than the cost-utility of competing programs or the status quo.[2]

Third Level of Coding

Following the appraisal of the studies, we examined each of the studies that were retained in order to cluster the studies into a small set of categories based on the content addressed: (a) studies that examined the effects of educational expenditures on student achievement (e.g., which types of spending contribute positively to higher achievement and which have negligible or weaker effects?) ; (b) studies that examined the private and social benefits of educational attainments (e.g., do more years of schooling produce higher career earnings? i.e., the focus in type (b) is on financial returns as opposed to the focus on learning outcomes in type (a)); (c) studies that assessed the cost-benefit/effectiveness of single programs (e.g., did the Perry Preschool program have enduring positive effects on its participants?); (d) studies that assessed program alternatives (e.g., which components of the program had the greatest effect on outcomes?) We then assessed the strengths and weaknesses of each study cluster in terms of their contribution to our understanding of the effects of educational programs.

Results

Appraisal of Study Quality

We found that 30 of the 102 Cost studies that we reviewed met our minimal criteria for inclusion; i.e., they had a credible design for assessing program effects, they provided data on costs and benefits, and they related costs and benefits using a defensible procedure. However, only five evaluations reached the highest level on all eight criteria. Four of these five were in the same content category (cost evaluations of single programs) and three were produced by a single researcher (W. Steven Barnett). Table 1 shows the quality codes for the 30 evaluations with the five best marked with an asterisk.

Table 1 about here

Content Analysis

Effects of Educational Expenditures on Student Achievement. All but one of the eleven cost studies in this set used cost-effectiveness techniques. The more useful studies in this category demonstrated that it is how money is spent that matters, not the overall level of expenditures.

One of the strongest studies was conducted by Elliott (1998) who asked: what do school districts buy with their funding that provides a productive impact? Elliott conducted a longitudinal study using a nationally representative database that tracked student achievement from grade 8 to grade 10. A national database provided data on district spending. Elliott adjusted district costs by the special education needs for each school and by geographic variations in what money can buy. Elliott used sophisticated statistical tools to link spending to achievement. His path model included student level controls (socio-economic status, racial background, and school track) and school level controls (school composition, size, and urbanicity). Elliot examined the

overall relationship between spending and achievement (science and mathematics learning) as well as how the money was spent.

Elliott found (i) the higher the per pupil expenditures in the district, the higher the district scores on teaching effectiveness and the greater the classroom resources; and (ii) the higher the scores on teaching effectiveness and classroom resources, the higher the learning of mathematics and science. This study provided convincing evidence for the teaching-effectiveness theory of resource expenditures: per-pupil expenditures increase students' achievement when the funds are used to hire the most qualified teachers and to train them in the most effective teaching methods. The analysis also supported the classroom-resources theory: per-pupil expenditures increase student achievement when funds are used to make the use of equipment, such as computers and microscopes, regularly available. Elliott's study moved the resources-achievement much farther along than the debate over the Hanushek reviews to focus attention on what resources can buy that make a difference to student learning.

Other cost studies in this category demonstrated that educational spending impacts achievement through complex paths, rather than simple one to one relationships. Wenglinsky (1997) concluded that there was a strong relationship between money and achievement. For example, every $1,000 per pupil increase in spending on instruction was associated with nearly a 1-point increase in mathematics achievement. But the relationship became visible only when appropriate controls (such as student SES) were included in the model. Paths were tracked from (i) instructional expenditures per pupil to (ii) teacher-student ratio to (iii) school environment to (iv) higher achievement. Flanigan, Marion, and Richardson (1997) also used path analysis to demonstrate that spending directed to attracting and retaining teachers with graduate degrees had the strongest effect on reading achievement.

Cost studies also demonstrated that the relationship between educational spending and achievement is moderated by a variety of factors. For example, variations in educational spending had a greater effect in schools serving middle/upper rather than lower SES students (Dolan & Schmidt, 1997; Harter, 1999), in small rather than large schools (Figlio, 1999), and in developing rather than developed countries (Ilon, 2004). Other studies in this category replicated the mixed findings of the Hanushek meta-analyses, reporting that educational spending, when appropriate student and district controls were in place, had a positive (Geski & Zuelke, 1982), negative (Coleman, 1986) or null (Huang & Yu, 2003) effect on student achievement.

However, all the cost studies in this category used some form of correlational design. Correlational designs are threatened by the possibility that the relationship between two variables is spurious; for example, that a third mediating variable not in the equation accounts for the relationship. Although statistical controls were put in place by the better studies to guard against this threat, correlational designs are threatened by reverse correlation (i.e., the inability to determine the direction of the relationship between educational spending and achievement). For example, Harter (1999) reported that "spending for student guidance services shows a negative association with math achievement. These discouraging results suggest that high-poverty schools are particularly susceptible to the ineffective use of resources" [p. 287]. A more plausible explanation for the finding might be that school districts provide compensatory funding to high-poverty schools in the form of special services to reduce the gap between the achievement of high-poverty schools and the district mean. In addition, these evaluations were limited by available data sets. Although learning is individual, these evaluations relied on the district as the unit of analysis, frequently without adjusting for contextual factors (such as district differences in student needs and regional variations in the cost of educational services) that might render

comparisons meaningless. In addition, these evaluations assumed a linear relationship between each of the predictors and the outcome variables, an assumption that was not tested.

Despite these limitations, cost studies that examined the relationship between macro spending categories and educational achievement helped refocus the debate from "whether money matters" to "what kind of spending makes a difference". In doing so, it moved the politically important debate about the cost of education into the realm of school improvement by linking educational spending to moveable variables like teacher quality and school size. In this sense the evaluations had a conceptual rather than instrumental use; i.e., the findings contributed to deeper understanding of the relationship between funding and outcomes rather than to the implementation of particular study recommendations (Patton, 1997). Cost studies also challenged the assumption that educational spending has a direct, linear effect on student achievement by identifying specific mediators and modifiers that affect the relationship. Recognizing the complexity of the paths between spending and outcomes also increased the proportion of achievement variance explained by financial resources, demonstrating that wisely allocated expenditures have more than the small effects reported by Greenwald et al. (1996).

Studies of Private and Social Benefits of Educational Attainments. Only four of the studies in this category met our minimal criteria. All were cost-benefit analyses examining rates of return on public and private investment in education. Two were conducted in developed nations and the other two in developing nations. The key issues addressed by these studies was whether countries should increase spending to reduce drop outs and whether individuals should invest their funds to increase the number of years of education completed.

In the oldest study we examined, Hansen (1963) calculated internal rates of return for public and private investments in education in the United States, using a variety of databases

including the 1950 census. Hansen calculated public costs (teacher salaries, supplies, interest and depreciation on capital) and private costs (earnings foregone while attending school and school related costs for tuition, books, and travel). Hansen computed the internal rate of return of these investments for society and for individuals with benefits defined in terms of lifetime earnings. The analysis demonstrated that the marginal rates of return for society increase with completion of elementary school and then gradually decrease through high school and college. However, completion of secondary and tertiary education had a greater rate of return to society than alternative investments. The rates of return were considerably higher for individuals at every level of schooling, mainly because of the public subsidization of schools.

Toh and Wong (1999) found that the social and private returns on educational investments in Singapore were greatest for the highest levels of schooling, particularly technical college. Toh and Wong attributed these findings to Singapore's economic development, arguing that its economy required and rewarded highly trained workers. Data from other Asian nations at similar levels of economic development produced similar results.

Fiszbein and Psachropoulos (1993) examined the costs and benefits of education in Venezuela by constructing regression equations to explain life time earnings in order to isolate the effect of number of years of schooling. They found that the greatest social return was for spending on primary education. The return on investment for individuals and society diminished with each educational increment but was positive at every step. The private returns on investment of education were greater than the social returns. However, in this study there was no attempt to include in the equation either the quality of the education of the quality of student performance--all graduates were treated alike.

Berhman and Birdsall (1983) considered educational quality using Brazilian data. They found that investments in education had a positive social return, as reported by other researchers, especially for primary education. But Berhman and Birdsall demonstrated that when the quality of teaching was considered, the social rate of return of primary education was greatly reduced[3], although still positive. They argued that the social rate of return was more likely to improve if governments focused on improving teacher quality than by expanding the number of years children attend school. This persuasive analysis was diminished by the measure of educational quality used in the study, average years of education of teachers, which is a weak proxy for a complex set of variables.

These cost studies had a direct effect on the policies of the World Bank, resulting in greater support for loans for primary than tertiary education (Levin & McEwan, 2001). Other uses of these studies were conceptual, focusing attention on the differential rates of return on educational investments of various groups. For example, both public and private rates of return were higher for less able students (Toh & Wong, 1999) and for females than males (Fiszbein & Psachropoulos, 1993). In addition these studies demonstrated that the private returns on educational investments exceed social returns, mainly because private costs are relatively low (consisting of foregone earnings and modest amounts for educational expenses) while the bulk of the costs of education are borne by the public.

The studies in this set are limited in several ways. First, these studies were cross-sectional rather than longitudinal. Uncontrolled variation between cohorts might distort the achievement-earnings relationship. For example, snapshot evidence does not take into account employment gaps, delayed starts and early retirements which can dramatically alter lifetime earnings. Second, most of these studies ignored factors such as educational quality and student performance that

might affect future earnings. Third, these studies report relationships between education and current or projected earnings but the accuracy of future earnings predictions is unknown (Tsang, 1997). Fourth, the translation of private earnings into social benefits assumes that differences in earnings accurately reflect differences in worker productivity; evidence in support of this assumption is not strong (Hough, 1994). Finally, individuals who acquire more education may have other personal attributes that influence their lifetime earnings beyond the direct effects of schooling.

Cost Studies of Single Programs. The nine cost studies in this category each examined a single program, comparing it to a control group or to variations in treatment within a program. Six of the nine used cost-benefit analysis.

Four studies (two were of the highest quality) tracked the effects of the Perry Preschool program. They found that at every time period, from ages 3-4 to 40, the private benefits to the low-income, minority group children enrolled in the program and the benefits to society far exceeded program costs (Barnett, 1992; 1993; 1995; Schweinhart, n. d.). The main findings at age 40 (Schweinhart, n.d.) were that the Perry Preschool participants, compared to control group members: (i) completed a higher level of schooling, (ii) were more likely to be employed, (iii) had higher earnings, (iv) had fewer lifetime arrests, and (v) earned 14% more per person than they would have otherwise. In constant 2000 dollars the economic return of the Perry Preschool program, 36 years after its delivery, was $17.07 per dollar invested. Of that return, 76% went to the general public and the remainder went to program recipients.

The strength of the cost-benefit analysis in these four studies lies in the quality of the original research design that generated the data. The Perry Preschool study has long been recognized as an exemplary evaluation (Wortman, 1995). The Perry Preschool study had random

assignment of participants to experimental conditions, virtually no attrition of study participants, and a plausible, consistent pattern of causes and effects from preschool to adulthood. In addition, the results were robust, showing positive returns even when benefit estimates were reduced by half and the real discount rate was increased to more than 7% (Barnett, 1993).

Another cost-benefit analysis based on rigorous determination of program effects used data from the STAR class size experiments, described by Finn (2002) as "one of the great experiments in education in U.S. history" (Mosteller, Light, & Sachs, 1996, p. 814). Krueger (2003) estimated the benefits of class size reduction using the findings from the original studies (which have been confirmed by re-analyses conducted by Finn, Gerber, Achilles, & Boyd-Zaharias, 2001 and Nye, Hedges, & Kostantopoulos, et al., 2002). Krueger drew upon the estimated correlation between test scores and future earnings to value the benefits of the program. After discounting costs and benefits to account for the time value of money, Kreuger concluded that the average rate of return was $2 for every $1 invested. Krueger also conducted a sensitivity analysis in which he varied the discount rate and predictions of annual earnings growth to demonstrate the robustness of the conclusions.

Other cost studies of single programs, although less rigorous than the Perry Preschool and STAR evaluations, demonstrated that the programs assessed were worth the money invested in them. Stern, Dayton, Paik and Weisberg (1989) found that special schools for at-risk students that combined academic and vocational courses were cost-effective, as were a Mexican technical education program for low income students (Lopez-Acevedo, 2003), a grade 5 mathematics program (Quinn, van Mondfrans, & Worthen, 1984), and a computer-assisted instructional program (Fletcher, Hawley, & Piele, 1990).

The cost studies in this set provided convincing evidence that particular programs justified their cost.[4] The evidence was particularly persuasive for the Perry Preschool evaluations: these findings launched preschool and Junior Kindergarten programs in numerous jurisdictions. In addition, the STAR evaluation strengthened support for class size reductions.

However, small changes in the calculations of costs, benefits or financial decision rules can have profound effects on study conclusions. For example, Stecher, Bohrnstedt, Kirst, McRobbie, and Williams (2001) found that the student achievement effects of reducing class size in California were much lower than the effects reported in STAR, the program on which the California intervention was modeled. The effect size in Stecher et al. was less than the .10 SD level at which the STAR internal rate of return on investment fell below zero.

Another limitation of these cost studies is that few considered the possibility that the effects of programs might change. General equilibrium effects could diminish the achievement benefits of programs if the innovations were implemented on a broad scale (Krueger, 2003) and effects might diminish even within a single site (Cronbach, 1982).

In addition, some of these studies assigned costs to society and benefits to individuals. For example, Krueger (2003) related the public costs of class size reduction to private benefits in the form of enhanced lifetime earnings of students in the program. From an economic perspective this is a reasonable procedure. The Kaldor-Hicks rule states that it is acceptable for someone to lose from a public policy as long as there is a net benefit when winners and losers are aggregated (Nas, 1996). The rule requires that winners be able to compensate losers but they need not actually do so. Not all taxpayers would agree with this perspective. Some would argue that Krueger's analysis supports private funding of smaller classes, an argument made by independent schools. A fairer examination of costs and benefits would relate public costs to

public benefits (e.g., benefits such as reduced expenditures for social services or increased taxes generated from increased lifetime earnings). It may be that policy makers would support the view that distributive justice (Rawls, 1971) requires that resources should be redistributed to benefit the least advantaged members of society. Cost studies should provide sufficient data for policy makers to make such a choice.

Cost studies of Program Alternatives. We found seven evaluations in this category that met our criteria. All but two used cost-effectiveness methods. These studies differed from the previous category in that two or more programs were examined rather than one.

Among the strongest of these studies was an examination of the cost-effectiveness of a primary education improvement project in a developing country (Creemers & van der Werf, 2000). Using a control group design, a battery of student and school level variables, and sophisticated analysis tools (HLM), the authors found a small achievement effect for the program. The use of pretests to adjust for pre-experimental group differences, the battery of student level controls they introduced (particularly their use of both low and high inference observational data), and the employment of standardized measures of achievement increased confidence that the program was responsible for the achievement gains. The evaluators determined cost-effectiveness using the effects divided by costs method. The most useful outcome of the evaluation was a rank ordering of the relative benefit of each of the ingredients that contributed to program cost (teacher development, management, books and materials, and community participation). The authors found that

> relatively cheap interventions with only small effects, like community participation, are extremely cost-effective for all subjects. More effective interventions for some subjects,

like the teacher professional development component for science, on the other hand, are quite expensive and less cost-effective (p. 377).

Other results from this set included Fielding's (1995; 1998) use of multilevel modeling to analyze the cost-effectiveness of four modes of delivering courses leading to advanced secondary school graduation. Lewis, Stockdill, and Turner (1990) found that a computer-assisted instruction program was more cost effective when combined with peer tutoring program than when delivered as a stand alone program, mainly because the costs were twice as high as the stand alone. Levin, Glass and Meister (1987) compared four instructional treatments finding that peer tutoring was the more cost-effective, because of a combination of larger effect sizes and lower costs than the other treatments. Barnett (1985) found that one year of the Perry Preschool program had greater cost-benefit than two and Lewis (1990) concluded that a 5th-year licensure program degree had a better cost-benefit ratio for individuals and society than the other options.

The cost studies in this set deepened our understanding of program effects by providing comparative information about competing program alternatives that went beyond consideration of effect sizes. For example, in a review of four strategies for improving student achievement, the option with the highest effect size (tutoring delivered by adults) was the least cost-effective (Levin et al., 1987). In addition studies in this set demonstrated that cost-effectiveness is moderated by student and school context factors (e.g., Creemers & van der Werf, 2000).

Discussion

Contribution to Understanding of Educational Program Effects

Our primary purpose in this article was to argue that evaluators rarely examine the cost of educational programs and their failure to do so limits our ability to develop a knowledge base of what works in education. Cost studies, as represented by the 30 studies in this review, make a

contribution to our understanding of the effects of educational programs in terms of each of the four purposes identified by Mark et al. (2000). Their first purpose concerns assessment of program worth. The cost studies reviewed here demonstrated that at the macro level educational spending provides a substantial return on investment for individuals and society. At the program level, these evaluations determined that particular programs were cost effective. These findings strengthen public confidence in educational policy making and justify maintenance of educational budgets. Other evaluation approaches do not address this dimension of program worth. As noted earlier, most evaluation textbooks give cursory attention to cost analysis; the discussion of cost and fiscal propriety in the Standards (Joint Committee, 1994) refers to the costs of the evaluation not the program; and the Guiding Principles (AEA, 2000) make no reference to examining program costs. The inclusion of cost considerations enables cost studies to answer a core question of funding agencies, including taxpayers, which would not otherwise be addressed.

The second contribution of cost studies is the guidance they provide on program improvement. For example, cost studies contributed directly to the establishment of publicly funded preschool programs, class size reductions, and support for primary education in developing countries. Cost studies combine evidence of program impact with evidence of cost-effectiveness, providing a broader foundation for allocation decisions at the school, district and state/province levels. The studies by Creemers and van der Werf (2000) and Levin et al. (1987) demonstrate that a ranking of program choices based on a consideration of costs and benefits differs from rankings based solely on effect sizes. Cost studies encourage implementation of low cost, moderate impact programs over high impact programs that may not be feasible on a broad scale or lead to lower net benefit for a given budget. In addition, the studies reviewed here found

that identifiable characteristics of programs, learners, teachers and organizations influence improvement efficiency, providing guidance to optimize the match of program elements with context.

Third, cost studies make a contribution to knowledge development, especially those evaluations that construct and test models of how educational spending influences student learning (e.g., Elliott, 1998; Wenglinsky, 1997). Others test previously developed models in novel contexts (e.g., Creemers & van der Werf, 2000).

Finally, cost studies make a lesser contribution to oversight and compliance. Cost studies rarely examine whether a program meets legal and financial requirements. Such issues are the domain of audits (Mayne, 2006). However, cost studies address the fiscal prudence of particular program structures and procedures, which could lead to the rewriting of regulations to make programs more efficient.

Assessing Quality in Cost Studies

Our secondary purpose of the article was to establish and apply criteria for judging the quality of cost studies. We developed eight criteria. One set focused on the credibility of the claim that an educational program had an effect on outcomes. It addressed internal validity (design and database) and external validity (population) as a condensed version of the 33 criteria for quantitative studies identified by Wortman (1994). The second set focused on study characteristics derived from cost-benefit/effectiveness manuals (especially Levin & McEwan, 2001).

We found that only 30 of the 102 cost studies published in refereed journals met our minimal criteria for inclusion; i.e., they had a credible design that enabled evaluators to distinguish program impact from other explanations for study outcomes and they related costs to

benefits using a defensible procedure. Among the 30 we found only five that could be considered exemplary, mainly because very few of the studies used experimental designs. Although we found that the majority of the 30 studies provided a rationale for the calculation of costs and benefits, only a slight majority (*N*=16) adjusted their estimates to account for the time value of money and a minority (*N*=13) conducted a sensitivity analysis to test the robustness of their assumptions. These findings suggest there is a problem of both quality and quantity in educational evaluators' consideration of costs.

Limitations of Cost Studies

Although we believe that cost studies provide three feasible options for illuminating the worth of programs (i.e., cost-benefit, cost-effectiveness and the less frequently used cost-utility approach), a key question yet to be resolved is whether cost studies are themselves cost-effective (i.e., if they meet Standard F3, that the information benefits of the evaluation warrant its costs). Cost evaluators must conduct two rigorous studies. First, evaluators need to attribute benefits to a program, a notoriously difficult task in education, especially when random assignment to treatment and control groups is not possible. Second, evaluators need to assess costs and relate them to benefits. The techniques required are costly, difficult to implement, require data not usually collected in a program evaluation, and involve assumptions that may be hard to justify. It makes sense to address the two evaluation phases sequentially—if there is no program effect, the cost-effectiveness component is moot. Whether a given cost study is cost-effective depends on its use. Use is a function of an evaluation's credibility to consumers (Patton, 1997). Unfortunately, cost studies combine the mystery of statistics with the inscrutability of economics.

Limitations of Our Review

Our review is limited by our initial search procedures that emphasized articles in refereed journals over books and unpublished monographs. Our selection of 30 studies led to important omissions. For example, we excluded cost studies that compared nonequivalent groups without statistical adjustment. This decision eliminated the demonstration of the cost-utility analysis of special education services by Lewis, Bruininks, and Thurlow (1988), as well as several studies that found that educational spending and particular programs were not worth the money.

Directions for Future Research

We suggest two priorities for cost study applications. The first is to examine the effectiveness of strategies for reducing the achievement gap of disadvantaged groups, as required by the No Child Left Behind Act (2001) in the U.S. and comparable legislation in other countries. Current evaluations focus exclusively on the size of gap reduction. For example, Balfanz, MacIver and Byrnes (2006) reported the effects of a program to reduce the gap in mathematics achievement through whole school reform. Although they provided some data on the cost of the program, there was no consideration of cost-effectiveness. In addition, the extensive literature on the appropriateness of the NCLB standards (e.g., Kim & Sunderman, 2005) has not examined cost-benefit issues.

The second proposed direction for future research is to extend cost-benefit/effectiveness comparisons from within education to comparisons between education and spending on other public priorities. We found only one study that compared the cost-effectiveness of educational spending to spending on other needs. Hy (2000) examined the effect of redirecting $50 million from less-essential non-educational spending to education, in Arkansas. Hy found there would be net gains in total personal income, in employee compensation, in taxes collected, and in job

creation. But Hy did not identify this less-essential spending except by state department, making it impossible to determine what social benefits might be lost by the redirection. Nor did Hy consider what the benefits of additional educational spending might be for students, such as improved achievement and future earnings. The challenge of making comparisons among government agencies lies in the valuing of the outputs generated by these agencies. For example, how does one compare increments in student achievement to reductions in death from breast cancer? Cost-utility procedures, rarely used in education, provide strategies for dealing with such questions. Doing so would enable evaluators who consider costs to provide policy makers with information highly relevant to budget construction.

Conclusions

Our review found that cost analysis is rarely included in educational evaluations at any level, much less to the degree of rigor required to use cost data persuasively. Inattention to cost analysis weakens evaluator's usefulness to policy makers and constrains the development of knowledge of what works in education. The 30 studies in this review demonstrate that cost studies investigate the worth of educational programs more completely than evaluation approaches that fail to do so. Effect sizes alone, even when used to compare the impacts of similar programs, are insufficient to guide decisions about the allocation of public funds.

The vast majority of cost studies in education are conducted by economists, not educational evaluators. We suggest five strategies to increase the frequency and quality of cost analysis by evaluators. First, funding agencies need to include in requests for proposals sufficient funds to enable evaluators to collect and analyze data on costs. Second, evaluator training programs need to include cost analysis techniques in their curriculum (as proposed by Levin, 2001). Third, the Standards and Principles need to include explicit attention to cost analysis.

Fourth, evaluators need to mine existing databases. For example, Kreuger (2003) conducted a cost analysis, rated as one of the top five in this review, in which he calculated the internal rate of return of reducing class size. Krueger combined impact data from an exemplary study of program effects, findings from studies reporting correlations between test scores and future earnings, and public data on per pupil costs at various levels of class size. Fifth, there are several excellent resources available to guide the novice through the techniques of cost analysis (e.g., Levin & McEwan, 2001; Nas, 1996; Rossi et al., 2003). Educational evaluators need to use them.

References

[Asterisked items are the 30 studies included in the review]

American Evaluation Association. (2000). Guiding principles for evaluators. *American Journal of Evaluation, 21*(1), iv-v.

Balfanz, R., MacIver, D. J., & Byrnes, V. (2006). The implementation and impact of evidence-based mathematics reforms in high-poverty middle schools: A multi-site, multi-year study. *Journal for Research in Mathematics Education, 37*(1), 33-64.

*Barnett, W. S. (1993). Benefit-cost analysis of preschool education: Findings from a 25-year follow-up. *American Journal of Orthopsychiatry, 63*(4), 500-508.

*Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational Evaluation and Policy Analysis, 7*(4), 333-342.

*Barnett, W. S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources , 27*(2), 279-312.

*Barnett, W. S. (1986). Methodological issues in economic evaluation of intervention programs. *Early Childhood Research Quarterly, 1*(3), 249-268.

*Berhman, J. R., & Birdsall, N. (1983). The quality of schooling: Quantity alone is misleading. *American Economic Review, 73*(5), 928-946.

*Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and characteristics of public schools in the US. *Journal of Political Economy, 100*, 1-40.

*Coleman, P. (1986). The good school district: A critical examination of the adequacy of student achievement and per pupil expenditures as measures of school district effectiveness. *Journal of Education Finance, 12*(1), 71-96.

*Creemers, B., & van der Werf, G. (2000). Economic viewpoints in educational effectiveness: Cost-effectiveness analysis of an educational improvement project. *School Effectiveness and School Improvement, 11*(3), 361-384.

Cronbach, L. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

*Dolan, R. C., & Schmidt, R. M. (1987). Assessing the impact of expenditure on achievement: Some methodological and policy considerations

. *Economics of Education Review, 6*(3), 285-299.

*Elliott, M. (1998). School finance and opportunities to learn: Does money well spent enhance students' achievement? *Sociology of Education, 71*(3), 223-245.

*Fielding, A. (1995). Institutional disparities in the cost-effectiveness of GCE ``A'' level provision: A multilevel approach. *Education Economics, 3*, 12-36.

*Fielding, A. (1998). Perspectives on performance indicators: GCE advanced level and differences between institution types in cost-effectiveness. *School Effectiveness and School Improvement, 9*, 218-231.

*Figlio, D. N. (1999). Functional form and the estimated effects of school resources. *Economics of Education Review, 18*, 241-252.

Finn, J. D. (2002). Missing the mark. *Phi Delta Kappan, 84*(3), 227-229.

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias. (2001). The enduring effects of small classes. *Teachers College Record, 103*(2), 145-183.

*Fiszbein, A., & Psachropoulos, G. (1993). A cost-benefit analysis of educational investment in Venezuela: 1989 update. *Economics of Education Review, 12*(4), 293-298.

*Flanigan J. L., Marion, R. A., & Richardson, M. D. (1997). Causal and temporal analyses of increased funding on student achievement. *Journal of Research and Development in Education, 30*, 222-247.

*Fletcher, J. D., Hawley, D. E., & Piele, P. K. (1990). Costs, effects, and utility of microcomputer-assisted instruction in the classroom. *American Educational Research Journal, 27*(4), 783-806.

*Geski, T. G., & Zuelke, D. C. (1982). Cost-effectiveness analysis: A practical tool for educational planning and changing. *Planning and Changing, 13*(1), 3-10

Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361-396.

*Hansen, W. L. (1963). Total and private returns to investment in schooling. *Journal of Political Economy, 71,* 128-140.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164.

*Harter, E. A. (1999). How educational expenditures relate to student achievement. Insights from Texas elementary schools. *Journal of Education Finance, 24*(3), 281-302.

Hough, J. R. (1994). Educational cost-benefit analysis. *Education Economics, 2*(2), 93-128.

Howell, E. M., & Yemane, A. (2006). An assessment of evaluation designs: Case studies of 12 large federal evaluations. *American Journal of Evaluation, 27*(2), 219-236.

*Huang, G. G., & Yu, B. (2002). District fiscal policy and student achievement: Evidence form combined NAEP-CCD data. *Education Policy Analysis Archives, 10*(38).

Hummel-Rossi, B., & Ashdown, J. (2002). The state of cost-benefit and cost-effectiveness analyses in education. *Review of Educational Research, 72*(1), 1-30.

Hy, R. J. (2000). Education is an investment: A case study. *Journal of Education Finance*, *26*(2), 209-218.

*Ilon, L. (2004). Cost-effectiveness of government inputs to schooling: technical and policy contexts. *Compare*, *34*(1), 87-101.

Joint Committee on Standards for Educational Evaluations. (1994). *The program evaluation standards: How to assess evaluations of educational programs.2/E*. Newbury Park, CA: Sage.

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, *34*(8), 3-13.

Knapp, T. R., & Knapp, L. T. (1990). A benefit-cost analysis of New York State's "Bundy Aid" program. *Economics of Education Review, 9,* 31-37.

*Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, *113*(485), F34-F63.

Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. *New Direction for Program Evaluation, 90,* 55-68.

*Levin, H. M., Glass, & Meister, G. (1987). A cost-effectiveness analysis of computer-assisted instruction. *Evaluation Review, 11,* 50-72.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications (2nd edition)*. Thousand Oaks, CA.: Sage.

*Lewis, D. R. (1990). Estimating the economic worth of a 5th-year licensure program for teachers. *Educational Evaluation and Policy Analysis, 12,* 25-39.

Lewis, D. R., Bruininks, R. H., & Thurlow, M. (1988). Using benefit-cost analysis in special education. *Exceptional Children, 55*(3), 203-214.

*Lewis, D. R., Stockdill, S. J., & Turner, T. C. (1990). Cost-effectiveness of micro-computers in adult basic reading. *Adult Literacy and Basic Education, 14*(2), 176-189.

*Lopez-Acevedo, G. (2003). A Reassessment of Technical Education in Mexico. *Journal of Career and Technical Education, 19*, 59-81.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.

Mark, M. M., Henry, G. T., & Julnes, G. (1999). Toward an integrative framework for evaluation practice. *American Journal of Evaluation, 20*(2), 177-198.

Mayne, J. (2006). Audit and evaluation in public management: Challenges, reforms, and different roles. *Canadian Journal of Program Evaluation, 21*(1), 11-46.

Monk, D. H., & King, J. A. (1993). Cost analysis as a tool for education reform. In S. L. Jacobson, & R. Berne (Eds.), *Reforming education: The emerging systemic approach* (pp. 131-152). Thousand Oaks, CA: Corwin.

Murnane, R. J., Willet, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics, 77*(3), 251-266.

Nas, T. (1996). *Cost-benefit analysis: Theory and application*. Thousand Oaks, CA: Sage.

Nye, B., Hedges, L. V., & Kostantopoulos, L. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123-152.

Patton, M. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks, CA: Sage.

Posavac, E., & Carey, R. (2007). *Program evaluation: Methods and case studies. 7/E.* Englewood Cliffs, NJ: Prentice-Hall.

*Quinn, B., van Mondfrans, A., & Worthen, B. R. (1984). Cost-effectiveness of two math programs as moderated by pupil SES. *Educational Evaluation and Policy Analysis, 6*(1), 39-52.

Rawls, J. (1971). *A theory of justice*. Boston: Harvard University Press.

Rossi, P., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach. 7/E*. Beverly Hills, CA: Sage.

*Schweinhart, L. J. (n.d.). *The High/Scope Perry Preschool Study Through Age 40: Summary, Conclusions, and Frequently Asked Questions. (Monographs of the High/Scope Educational Research Foundation)*. Ypsilanti, MI : High Scope Press.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15-21.

Stecher, B., Bohrnstedt, G., Kirst, M., McRobbie, J., & Williams, T. (2001). Class-size reduction in California: A story of hope, promise, and unintended consequences. *Phi Delta Kappan, 82*(9), 670-674.

*Stern, D., Dayton, C., Paik, I.-W., & Weisberg, A. (1989). Benefits and costs of dropout prevention in a high school program combining academic and vocational education: Third-year results from replications of the California Peninsula Academics. *Educational Evaluation and Policy Analysis, 11*, 405-416.

Thompson, M. (1980). *Benefit-cost analysis for program evaluation*. Beverly Hills: Sage.

*Toh, M. H., & Wong, C. S. (1999). Rates of Return to Education in Singapore. *Education Economics, 7*, 235-252.

Tsang, M. C. (1997). Cost analysis for improved educational policymaking and evaluation. *Educational Evaluation and Policy Analysis, 19*(4), 318-324.

Verstegen, D. A., & King, R. A. (1998). The relationship between school spending and student achievement: A review and analysis of 35 years of production function research. *Journal of Education Finance, 24*(2), 243-262.

*Wenglinsky, H. (1997). How money matters: The effect of school district spending on academic achievement. *Sociology of Education, 70*(3), 221-237.

What Works Clearinghouse. (nd) *WWC Study Review Standards Revised September 2006* [Web Page]. URL http://www.whatworks.ed.gov/reviewprocess/study_standards_final.pdf [2007, April 19].

Wortman, P. (1995). An exemplary evaluation of a program that worked: The high/scope Perry Preschool Project. *Evaluation Practice, 16*(3), 257-265.

Wortman, P. (1994). Judging research quality. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 98-109). New York: Russell Sage Foundation.

Appendix

Initial Coding Scheme

1. StudyPurpose - describe purpose of the study and/or research questions addressed by the study.
2. StudyDesign - describe design: 0) no control group or nonequivalent groups without statistical adjustment, 1) random assignment to treatment and control groups, 2) nonequivalent groups with statistical adjustment, 3) interrupted time series, 4) correlational (i.e., designs in which statistical controls were introduced to isolate program effects from the effects of other variables using procedures such as hierarchical regression, path analysis, HLM) 5) other (e.g., regression-discontinuity design).
3. StudyTheory - describe theoretical framework of the study (i.e., constructs and their relationships).
4. StudySample - describe sample (e.g., N=?? schools).
5. SampleRep - classify sample: 1) representative of a population (identify) or 2) other.
6. StudyVar - describe variables in the study.
7. VarMeas - describe how variables were measured.
8. VarRelVal - describe evidence of reliability and validity of measures.
9. CostAssess - describe how costs were assessed.
10. BenefitAssess - describe how benefits were assessed.
11. RelateCost-Ben - describe how costs and benefits were related.
12. StudyResult - describe findings of study.
13. DesignThreats (see Wortman, 1994 for categories): (a) construct validity (b) external validity (c) internal validity (d) statistical conclusion validity

14. AccStand - describe & infer accuracy of study (Joint Committee, 1994; summative judgment using 1-5 scale in which 5 is high)
15. UtilityStand - describe & infer utility of study (Joint Committee, 1994; 1-5 scale in which 5 is high):
16. FeasibilityStand - describe & infer feasibility of study (Joint Committee, 1994; 1-5 scale in which 5 is high).
17. EthicalStand - describe & infer probity of study (Joint Committee, 1994; 1-5 scale in which 5 is high).
18. VFMBenefit - benefits of considering costs in the evaluation.

Table 1 Coding of Cost Evaluations in the Review

Author	Design	Population	Database	Time Adjust	Sensitivity	Cost Rationale	Benefit Type	Decision Rule
Effects of Educational Expenditures on Student Achievement								
Card & Krueger, 1992	correlational	identifiable	high quality	no	yes	with rationale	private	CB
Coleman, 1986	correlational	identifiable	high quality	no	no	no rationale	public	CE
Dolan & Schmidt, 1987	correlational	identifiable	low quality	no	yes	with rationale	public	CE
Elliott, 1998	correlational	identifiable	high quality	yes	no	with rationale	public	CE
Figlio, 1999	correlational	identifiable	high quality	no	yes	with rationale	public	CE
Flanigan et al., 1997	correlational	identifiable	low quality	yes	no	with rationale	public	CE
Geski & Zuelke, 1982	correlational	identifiable	high quality	no	no	no rationale	public	CE
Harter, 1999	correlational	identifiable	high quality	no	no	no rationale	public	CE
Huang & Yu, 2002	correlational	identifiable	high quality	yes	no	with rationale	public	CE
Ilon, 2004	correlational	identifiable	low quality	no	no	with rationale	public	CE
Wenglinsky, 1997	correlational	identifiable	high quality	no	yes	with rationale	public	CE
Studies of Private and Social Benefits of Educational Attainments								
Berhman & Birdsall, 1983	correlational	identifiable	low quality	yes	yes	no rationale	public	CB
Fizbein & Psachropoulos, 1993	correlational	identifiable	low quality	yes	no	with rationale	pub/priv	CB
Hansen, 1963	correlational	identifiable	high quality	yes	yes	with rationale	pub/priv	CB
Toh & Wong, 1999	correlational	identifiable	high quality	yes	no	no rationale	pub/priv	CB

Cost Evaluations of Single Programs								
*Barnett, 1992	experiment	identifiable	high quality	yes	yes	with rationale	pub/priv	CB
*Barnett, 1995	experiment	identifiable	high quality	yes	yes	with rationale	pub/priv	CB
Barnett, 1993	experiment	unknown	high quality	no	no	with rationale	pub/priv	CB
*Fletcher et al., 1990	experiment	identifiable	high quality	yes	yes	with rationale	public	CE/CU
*Kreuger, 2003	experiment	identifiable	high quality	yes	yes	with rationale	private	CB
Lopez-Acevedo, 2003	QED	identifiable	low quality	yes	no	with rationale	pub/priv	CB
Quinn et al., 1984	correlational	identifiable	high quality	no	yes	with rationale	public	CE
Schweinhart, n.d.	experiment	identifiable	high quality	yes	no	with rationale	pub/priv	CB
Stern et al., 1989	QED	identifiable	high quality	yes	no	with rationale	pub/priv	CB
Cost Evaluations of Competing Program Alternatives								
*Barnett, 1985	experiment	identifiable	high quality	yes	yes	with rationale	pub/priv	CB
Creemers & van der Werf., 2000	QED	identifiable	high quality	no	no	with rationale	public	CE
Fielding, 1995; 1998	correlational	identifiable	high quality	no	no	with rationale	public	CE
Levin et al., 1987	QED	unknown	high quality	no	no	with rationale	public	CE
Lewis, 1990	QED	identifiable	high quality	yes	no	with rationale	pub/priv	CB
Lewis et al., 1990	QED	unknown	low quality	no	yes	with rationale	public	CE

*These five cost studies had the highest ratings across all eight criteria. QED=Quasi-experimental Design; CB=Cost-benefit; CE=Cost-effectiveness; CU=Cost-utility

Endnotes

[1] Although inter-rater statistics were not compiled, reliability was enhanced by 1) training coders on use of the code book; 2) each review was read by the senior author who sent feedback to the reviewers; 3) questions posed by individual reviewers were explicitly addressed and circulated within the team.

[2] We view these three models of cost studies as equivalent, even though we presented them in the order of CB, CE, and CU.

[3] Behrman & Birdsall (1983) conducted a regression analysis in which the proportion of the variance in rate of return explained by years of schooling declined when a third variable, quality of teaching (meaning average number of years of teacher training) was introduced into the equation. Teaching quality was a moderator variable that predicted both years of schooling (schools with better trained teachers retained students for longer periods) and rate of return (schools with better trained teachers had a better rate of return on educational investment).

[4] We located studies that concluded that a program was not the worth the money. For example, Knapp and Knapp (1990) found that a program to provide direct state aid to certain private colleges and universities in the state of New York was not cost-effective. However, this study and others were deleted from the review because it was a post-only, single group study with no controls.