# Education and Innovation Theme

## Cost-Effectiveness in School Education
## January 2013


Clive Belfield, City University of New York
Henry M. Levin, Columbia University

# Cost-Effectiveness in School Education

**Clive Belfield**
Associate Professor, Queens College, City University of New York
Research Fellow, Center for Analysis of Postsecondary Education and Employment,
Teachers College, Columbia University
belfield@tc.edu

**Henry M. Levin**
William Heard Kilpatrick Professor of Economics and Education
Teachers College, Columbia University
levin@tc.edu

**Fiona Hollands, Brooks Bowden, Henan Cheng, Robert Shand, Yilin Pan, and Barbara Hanisch-Cerda.**
Center for Benefit-Cost Studies of Education at Teachers College,
Columbia University

**January 2013**

Research referred to in this paper was, in part, funded by the U.S. Department of Education, Institute of Education Sciences Award Number R305U120001. We appreciate comments and feedback from participants at the Grattan Institute Conference, October 2012 in Washington, DC.

## Abstract

The need for cost-effective education is paramount. Yet very little research attention is paid to cost-effectiveness analysis, despite the method being set out over four decades ago and despite its relatively straightforward objective. In this paper we review the method of cost-effectiveness analysis as applied to education. We draw attention to a number of important practical and methodological challenges when linking costs to effectiveness and performing cost-effectiveness analysis. Despite these challenges, this form of analysis can still be valuable for helping allocate scarce resources. Cost-effectiveness analysis is very much underdeveloped – both in method and application - in education research, but under the circumstances of current budget constraints we anticipate greater reliance on this form of analysis to inform decision-making in education.

## Introduction

Almost three-quarters of one trillion dollars of public funding are spent on elementary and secondary school education in the United States; this is approximately 5% of Gross Domestic Product (U.S. Government Spending, 2012). Historically, educational costs have risen at a much faster rate than the consumer price index. Investments in education will therefore need to grow as a proportion of GDP, purely to preserve existing programs and services for students. Alternatively, political and financial constraints will impinge such that less resource will be provided for future cohorts. Indeed, state education budgets have been falling since the recent Great Recession: the Center on Budget and Policy Priorities (2012) reports that for the 2012-2013 school year, 35 states are providing less funding for education than they did five years ago. Under either scenario, however, there needs to be a greater emphasis on costs and efficiency across the entire education system. As budgets grow, small efficiency gains can still be large in dollar terms; and if budgets sink, there will be increased pressure for districts and states to make decisions on which programs to cut. Schools will need to be more cost-effective in allocating resources.

Little attention has been focused on the general issues of cost and productivity of education and how outcomes can be improved relative to the costs. The early literature focused on whether money made a difference for educational outcomes. Unfortunately, phrased as ‘does money matter?’, this research did little to help educational decision-makers. At some level, extra dollars of spending may yield relatively less of an increment in educational outcomes than existing dollars but there was little guidance as to when that point might be reached or whether relatively less was still better than nothing (see Hanushek, 2004). Policy debates reached a stalemate. Researchers rarely claimed that resources should be reduced – and almost never specified where cuts should be made – but instead argued against increases in resources. At best, there are examples such as Hanushek's suggestion that one expensive but ineffective reform (class size) should be sacrificed in favor of another (higher teacher salaries to improve teacher quality). On the other side, researchers would argue that finding a program to be effective was sufficient to justify committing resources to it.

Yet this research overshadowed an alternative method for evaluating efficiency within the school system, namely cost-effectiveness analysis (CEA). In principle, cost-effectiveness analysis is straightforward: it requires that the outcomes of an intervention or program be evaluated in relation to the resources committed to it, and it is intended explicitly to assist decision-makers. This form of analysis for educational evaluations was set out in detail by Levin (1975) and in a subsequent book also by Levin (1983, and then with McEwan, second edition, 2001).

A great deal of attention has been given in both the research literature and policy to the effectiveness of educational alternatives. Although strong arguments have been made that cost also needs to be taken into account (Harris, 2009), there has been little effort devoted toward linking effectiveness to measures of costs (Levin, 2002; Hummel-Rossi, 2001; Clune, 2001). This omission of cost considerations risks the promotion of educational interventions that have only small positive effects, but high costs that exceed those of equally effective alternatives.  More generally, it has led to research into school effectiveness that has limited application for decision-makers who face budgetary constraints.

Cost-effectiveness analysis assists policy-makers in setting priorities among a proposed group of interventions on the basis of their efficiency in resource use at achieving a specific outcome. Although providing data on relative efficiency, CEA does not answer the question as to whether such an outcome is worth investing in; cost-benefit analysis is necessary to provide such an answer. Yet, the evidence on the economic benefits of education is often compelling enough that CEA may be sufficient. If a program can be shown to be effective in improving educational outcomes, it is likely that it will pass a cost-benefit test. But the program would still need to establish that it is more cost-effective than other programs with the same goals. CEA is therefore directly related to the decision-making process for policymakers. In situations where effectiveness data already exist for a program, CEA can often be performed more rapidly than cost-benefit analysis. Thus, a strong case for more frequent use of CEA can be made.

This paper provides an overview of these issues. We begin by describing the method of cost-effectiveness analysis. We then consider why it has not been widely adopted. Moreover, because of the lack of attention to cost-effectiveness in prior research, there are a series of empirical and methodological challenges in performing this analysis. We itemize these challenges, using examples drawn from recent work with colleagues (Levin *et al*., 2012). If these challenges are addressed there is greater potential for cost-effectiveness analysis to help improve resource allocation in schools. However, even if these challenges are not addressed we believe that the general principles of cost-effectiveness analysis can still be applied and that this will improve resource allocation in schools.

## The Method of CEA

### *Ingredients Method*

Cost-effectiveness analysis is a decision-oriented method of inquiry. The decision-maker's first task is to determine which educational outcome is to be pursued. Reforms and interventions can then be evaluated insofar as they improve that outcome and at what cost. Thus, both effectiveness data and cost data are necessary and they must be combined. Our focus is on the latter two components.

Collecting cost data requires identification of each resource used to deliver a given intervention, which is then labeled as an ingredient. For educational programs, the main ingredient is usually labor services of which the largest component is teacher time. But, facilities, equipment, supplies, outside services, and many other ingredients may be used, and these must also be identified and measured. Student time is also important: older students might otherwise be able to use their time productively in the workplace. Some students may be inputs in helping deliver interventions, such as peer tutoring, to other students. Because the ingredients provide the most fundamental units for ascertaining costs, these must be measured as accurately as possible. Thus, it is usually important to know not only the role or function of a personnel ingredient, but also enough about the skill, education, and experience required to carry out the job in order to determine what such an input or ingredient will cost.

Once ingredients have been identified and stipulated, the next step is to ascertain their prices. All ingredients are assumed to have prices, including donated or volunteer resources, because all have opportunity costs. A crucial requirement for cost analysis is that the cost estimates must be comparable so all ingredients are priced using a standard set of prices. Thus, cost values are independent of the specific year in which the evaluation was done. Also, these values are independent of local and regional differences in costs. Some parts of the country and some localities have lower prices than others for some ingredients and higher prices for others. Prices can change over time with market shifts, e.g., depending on the supply of qualified teachers or principals. These conditions are not intrinsic to the educational intervention and so should not be used in making cost-effectiveness comparisons.

The cost values of all identified ingredients must be accounted for, regardless of their source of finance or in-kind support. It is important to cost out all the ingredients necessary for the intervention over its entire duration. The duration of the intervention must also be taken into account. Educational interventions vary significantly in length, from a couple of days (e.g., teacher professional development) to multiple years (e.g., leadership programs). Finally, these costs should be calculated as incremental costs, i.e., taking into account resources utilized by the intervention beyond those utilized for "business-as-usual". For some interventions, students who do not participate may still access alternative educational resources. If so, these alternatives need to be costed to allow determination of the incremental costs of the intervention.

The outcomes of the cost analysis are cost metrics for each intervention expressed as: a total cost for the intervention at the level of scale studied, i.e. across all sites; a cost per site when site by site data is available; and an average cost per student receiving the intervention. Each of these metrics serves a particular purpose. The total cost metric informs decision-makers of the scale of the intervention, which may be salient if there is a funding constraint. The site level costs indicate whether and how resource requirements vary with differences in program implementation. The average cost per student metric yields an easier comparison between interventions and may be useful when considering the equity of investments across subgroups of students.

The ingredients method and the cost metrics are not complicated, at least in theory. Yet, we are aware of very few studies that apply the formal ingredients method to estimate the costs of education programs. Studies that do include costs typically rely on budget statements or prices given by education providers. Rarely do studies present average costs, total costs, and site-specific costs either in the aggregate or divided across the agencies that fund these resources.

### *Using Cost Analysis in a Decision-oriented Framework*

Even as cost analysis does not tell the whole picture, these metrics are nonetheless very valuable for decision-makers in the context of scarce resources. The initial aim of a cost analysis is to provide comparable cost estimates across different interventions. A later aim is to adapt the standardized findings to apply to situations encountered by different policymakers in varying educational jurisdictions. Ideally, the cost analysis should be set up to indicate how the cost burden is distributed, or "who pays", among school districts, other government agencies, private entities, volunteers and other providers of in-kind resources, and clients or users of the program. This allows a decision-maker to assess the burden that will fall on his/her particular budget, given the particulars of education cost sharing in his/her jurisdiction.

In some instances, the policymaker's concern is with the costs of expanding the use of an existing intervention, e.g., to a higher grade level. In this case, costs must be distinguished as fixed or variable. In this situation, the marginal costs of the intervention can be determined by considering only the variable cost component of ingredients. In general, those interventions with high fixed costs such as those with large investments in facilities and equipment (e.g., establishing computer laboratories) will require a high enrollment or utilization to be most efficient. In contrast, interventions that are constituted largely of variable costs such as personnel (e.g., after-school tutoring) will have costs that are less sensitive to the scale of output. Thus, an economic evaluation of alternative interventions that differ in terms of their intensities of fixed versus variable costs may produce very different average and marginal cost results depending on the scale of enrollment or output. Therefore, it is advisable to compare interventions of similar scale.

A related analysis could estimate how much increase in effectiveness can be expected for a fixed monetary investment in a program. This analysis would be useful when the budget is limited and decision-makers need to determine how to maximize their return on investment. In this circumstance, decision-makers would choose the intervention that provides, for example, the greatest increase in high school completion but still falls within their budget limit.

### *Cost-Effectiveness Analysis*

Once the cost metrics have been calculated, a cost-effectiveness ratio is derived. This ratio is calculated as the cost divided by the effectiveness. For example, if a program costs $2 million and yields 100 new high school completers above and beyond what would be expected from a valid comparison group, the cost-effectiveness ratio is $20,000. This ratio shows the cost of 'buying' these extra completers. Lower cost-effectiveness ratios are preferred (if the program yielded 200 extra completers the cost-effectiveness ratio would be $10,000). This ratio is helpful when it is compared against alternative ways to improve the high school completion rate: this example program is the most cost-effective as long as no other program yields extra high school completions for less than $20,000. This ratio may also be helpful because it can be easily related to the value of the program – specifically, whether it is worth spending $20,000 to buy extra completers.

Alternatively, the ratio may be expressed by dividing the measure of effectiveness by the cost figure to provide a gain per dollar spent. For the above example, this ratio could be expressed as the yield of extra completers per $100,000 spent, which is 5. Therefore, the intervention generates 5 extra high school completers for every $100,000 spent. This ratio is helpful when there is a financing constraint (e.g., if the district can only spend $1 million). In addition, when programs have negative effects, e.g., the treatment group actually yields a smaller percentage of completers than the control group, a traditional cost-effectiveness ratio may be impossible to interpret. It is, however, possible to comprehend such a result if it is presented as the number of graduates "lost" for every additional $100,000 spent. Thus, this alternative metric is useful and should be presented alongside the cost-effectiveness ratio where appropriate.

As with all empirical investigations, cost-effectiveness ratios should be tested for the confidence we have in the estimated values. In the case of these ratios, we are interested in how robust the results are to alternative modeling assumptions. This sensitivity analysis should include best and worst case sensitivity testing, as well as Monte Carlo simulation where appropriate.

The primary purpose of cost-effectiveness ratios is to compare interventions. If there are implementations at multiple sites, the cost-effectiveness of the intervention according to the version implemented at each site can be ranked for its efficiency in use of valuable resources. Where there is more than one intervention addressing a particular outcome, the ratios can be compared across all the interventions. However, the cost-effectiveness of a program is reliable only if site-specific variation is small; otherwise the average result may not be obtained when the program is implemented in a given setting. Continuing the above example, imagine that a second intervention aiming to increase graduation has a total cost of $5 million and yields 200 extra graduates. This intervention is more effective overall, but the higher cost renders the comparable cost-effectiveness ratio less favorable ($25,000), i.e., it costs more to obtain an extra high school completer. Cost-effectiveness comparisons do not have to be limited to one type of approach or reform. For high school completion, for example, one can compare interventions based upon diverse approaches such as coaching, curriculum, technologies, professional development, grouping practices, and extended school time.

Although the method of cost-effectiveness analysis is simple, its practical application is challenging. Indeed, it must be because so few evaluations include it and have instead undertaken only effectiveness studies. Yet, this poses a quandary because from an economic perspective, effectiveness analysis is incomplete and therefore potentially misleading. Costs need to be incorporated with effectiveness. Very simply, implementing the most effective program is nonsensical if it is too expensive (e.g., if there are multiple other programs that could be implemented for the same resource investment and that are more effective in combination). There is no guarantee that a more costly program is going to be more effective. There is no guarantee that all interventions are delivered at the same cost (such that only their relative effectiveness is of interest). A more likely scenario is that a given educational objective can be achieved for a much lower cost using one intervention than from using another intervention.

Indeed, there are some examples where the same educational result has been obtained for a fraction of the cost of an existing practice or policy. An early study of teacher selection found that, in order to raise student achievement, it was five to ten times more cost-effective to select teachers with higher verbal test scores than to invest in teachers with additional teaching experience (Levin, 1970). Levin, Glass, and Meister (1987) found that, for raising the achievement of elementary students, peer tutoring was twice as cost-effective as computer-assisted instruction and almost four times as cost-effective as reducing class size or increasing instructional time. See also Borman and Hewes (2002).

**Challenges to Performing Cost-Effectiveness Analysis**

The rarity of cost analysis and CEA, in contrast with the ubiquity of effectiveness studies, suggests that cost analysis and CEA must be difficult to perform. Indeed, in our own investigations we have identified a set of practical challenges and a set of methodological challenges to performing CEA. We describe these below. However, we caution that these challenges do not undermine the need to perform CEA. This need still holds even if these challenges cannot be met in full. Moreover, we note that these challenges, which are not widely acknowledged, do not seem to be the main barrier to performing CEA in education.

### *Practical Challenges to Cost-Effectiveness Analysis*

The first challenge is to recognize that budgetary data are not sufficient for costing out programs. There are many reasons why budgets are inadequate for determining costs (Levin & McEwan, 2001, pp. 45-46). The overriding concern of accounting and budget reporting in education is to establish transparency in how money is spent, primarily for auditing purposes. Often, capital improvements that last many years are charged to the budget in the year that the improvement is made or over a fixed time period rather than being spread out over all the years of serviceability. Also, resources that are received from non-school sources (such as volunteers, gifts, use of facilities belonging to other agencies) are often not accounted for at all. In some states the pension system for schools is charged to the state and does not appear as a cost in local school budgets. By accounting for the ingredients used in a program or intervention, most errors of omission or cost distortion can be avoided.

Using budget data almost always understates the full cost of implementing an intervention. Even more egregiously, costs are sometimes reported from the developer's perspective, e.g., the amount an educational software program costs to buy. No attention is then paid to the much larger cost – the school resources needed to implement the software (on the discrepancy, see Levin *et al*., 2007). Thus, programs may appear inexpensive, when in fact much of the burden of implementing the program falls on the school personnel.

A second challenge is to use standard prices for particular inputs. For example, if a new teacher with a BA and graduate training is required, the prevailing labor market price should be used. This ensures that interventions can be compared appropriately. Unless standard prices are used, any intervention delivered in New York City will be 30% more costly than one delivered in Wyoming (CPI differences in prices). Unless standard prices are used, school districts will not easily be able to know how much the intervention would cost in their locality. They would have to know all the sites where the intervention was delivered and calculate their own regional price index to apply to their school district. Collecting these standard prices is a challenge, although in the U.S. there are sufficiently detailed datasets that allow for estimating salaries of teachers with many different educational qualifications and training and with varying levels of experience.

A third practical challenge is persuading evaluators that collecting cost data is as important as collecting effectiveness data and that both should be collected simultaneously. Many important and effective educational interventions have already been rigorously evaluated for impact, but starting over in order to collect the costs presents additional cost and time delays, especially as rigorous evaluations of impact can take several years to complete. In a few cases, costs are indeed collected at the time of implementation, as we recommend, or some time later. However, retrofitting costs with effectiveness data – as is almost always required given the absence of contemporaneous cost analysis – creates a significant research challenge. Useful information on the educational intervention that is readily available at the time of its implementation may no longer be available at a later date or may not be fully accurate. The program may have changed and the inputs required may no longer be useful (e.g., computers from the 1990s). This leaves aside the additional effort required to obtain costs data from programs that may have moved site, changed personnel or have changed as to how they are implemented. Certainly, by incorporating the ingredients method at the time of implementation of the intervention, the costs can be obtained with greater accuracy and less effort.

### *Methodological Challenges to Cost-Effectiveness Analysis*

The primary methodological challenge is that interventions may only be compared in a cost-effectiveness analysis if they measure at least one outcome in common. Assumptions about the equivalency of outcomes need to be reviewed carefully. For example, many 'dropout prevention' programs have much broader goals, some of which may be more important than completing high school. In these cases it is usually impossible to disentangle costs associated with one outcome from the costs associated with other outcomes. But the challenge is bigger than this: the interventions must be evaluated using the same scale. Two reading programs may be compared in terms of general effectiveness in 'reading', but in order to be compared for cost-effectiveness the outcomes must be measured using the same scale (e.g., the TOWRE reading scale). It is also important to consider whether the interventions serve similar populations of students in similar settings, are delivered at similar scale, and are funded at similar levels (so that one does not exceed the district's budget, for example). Each of these factors may affect both costs and effectiveness of a program, and they almost certainly influence the decision-making process. To make a viable comparison, similarity in each of these characteristics is desirable. Yet, it is often hard to find programs that are genuine alternatives to each other and so may be meaningfully compared.

Our review of interventions to reduce the dropout rate exemplifies this primary challenge (see Levin *et al.,* 2012). Within the dropout prevention topic area, the Institute for Educational Sciences (within the U.S. Department of Education) has identified and fully reviewed 13 interventions with positive or potentially positive effects (IES, 2012). Thus, we might anticipate that dropout prevention programs might be compared for cost-effectiveness. However, these interventions are actually categorized into three outcome areas: completing school (graduating from high school or earning a GED); progressing in school (moving up a grade); and staying in school across any grade. These are not the same outcome. Improvements in the latter two outcomes – progressing and staying in school – are necessary but not sufficient for a student to complete high school.

Moreover, when we look at the five interventions that do address school completion, these interventions differ in many ways such as scale, target population, duration, nature and intensity of services provided, and nature and timing of outcomes sought. One is an extant national add-on program that complements existing schooling for students who are expected to finish high school and attend college (Talent Search); two were limited-period demonstration programs for youth who had dropped out of school and needed job training and other life skills (JOBSTART and New Chance); two more are ongoing, intensive residential programs also targeted at youth who have already dropped out of school and need significant additional services beyond educational interventions (National Guard Youth ChalleNGe, Job Corps). In addition to program characteristics, the definition of school completion is not uniform: in some programs students are motivated to pass the GED; others help students graduate from high school while yet others do both. While the economic consequences of earning a GED are inferior to those from earning a high school diploma, most studies of these programs combined the two outcomes as if they were equivalent. Treating GED receipt as equivalent to high school completion with a diploma is problematic. There is substantial evidence that the economic benefits from possessing a high school diploma far exceed those from possessing a GED (Heckman, Humphries, & Mader, 2010).

More fundamentally, the intended outcomes of the programs were not uni-dimensional: all programs had multiple goals beyond high school completion, for example to increase employability and earnings. These other outcomes are more effectively captured in cost-benefit analyses. One intervention (Talent Search) is not directly comparable with the other four interventions: it is not directed at the same student population as the others (they are for dropouts); and its effect is incremental beyond the services students receive from their high schools.

There are also challenges in interpreting the effectiveness of a program and linking this effectiveness measure to the appropriate costs. We identify four additional challenges therefore, noting that these arise even where there is only one effectiveness outcome being measured.

Pooling information across studies is problematic. Some interventions may have been evaluated through multiple studies, each showing different results and effect sizes. Pooling results from multiple studies is unhelpful for cost-effectiveness analysis because it is likely that the cost of implementing the intervention varied across studies. Also other information that accounts for differences in results is lost such as differences in population, differences in base support for education, differences in implementation. The problem is more one of aggregation bias of very different situations and variation in provision of the intervention. More resource-intensive implementations of the intervention are more likely to be more effective. Hence, the cost-effectiveness ratio from pooled results may be a biased estimate of the overall cost-effectiveness of the intervention. We recommend calculating cost-effectiveness only when an individual effectiveness study can be matched up with costs of the specific implementation(s) evaluated.

It is also difficult to pool information across sites for a given study. Some evaluation studies report on multiple sites implementing the same program. In these cases, some sites appear to be effective while others are ineffective. As with multiple studies, the question arises as to which sites best represent the impact of the program for a cost-effectiveness analysis. Cost-effectiveness can be presented for the overall group and also for the subset of effective sites to show how the program in question compares with other programs when implemented optimally. Alternatively, if only some sites show statistically significant differences in outcomes for the treatment group compared with the comparison group (either positive or negative), another analysis can include only those sites as opposed to including all sites. We recommend presenting a number of different such analyses so that a policymaker can choose the analysis that makes most sense in his or her context.

Another problem is how to define the sample for costing out the program. Effectiveness research typically distinguishes between participants who are assigned to the intervention and those who participate in the intervention. For experimental studies, it is important to determine whether the evaluation includes all participants assigned to the treatment group, regardless of whether they actually attended the program (the Intent-to-Treat or ITT observations), or only those who actually participated in program activities (the Treatment-on-the-Treated or TOT observations). Our preference is for cost-effectiveness analysis based on the TOT observations although the implications for ITT cost-effectiveness should also be explicitly considered. Although interventions are typically allocated resources based on the ITT observations, these resources are actually used on the TOT observations. Where fewer students participate than expected (i.e., there are fewer TOT observations than ITT observations), resources are often not returned to the funding agency but instead are spread across the participants. The intervention is therefore more resource-intensive per participant. Critically, it is the actual resources used that will determine effectiveness. For example, if the ITT observations are 2,000 and the intervention has a budget of $1 million, then the unit cost is $500. If there are only 1,000 TOT observations, the unit cost is likely to be $1,000, assuming the agency will spend the entire $1 million. Therefore, the intervention's effectiveness reflects $1,000 of resources. In contrast, it may be hard to know how many resources would have been spent on the ITT individuals (or at least the subgroup who did not participate). It is unclear how the decision to use ITT *vs.* TOT observations will drive the cost-effectiveness results. Clearly, TOT effectiveness is likely to be greater than ITT effectiveness, as in the former case the impact is only measured for those students participating in program activities, who are presumably more motivated. However, the per-participant costs for TOT will also be higher because the total cost of the program is divided by fewer observations.

Finally, most programs are incremental on top of existing programs and the evaluator must recognize the extent of the increment. Most interventions use resources above and beyond what is already being spent. For example, a dropout prevention program in high school uses incremental resources beyond regular instructional resources but these regular instructional resources may still help prevent dropouts. Thus, the program is only an increment to what is already being spent. Another example is a boot camp: dropouts participate in the camp and receive all the resources of the boot camp but the costs should measure only the incremental resources beyond what the non-participants receive. Potentially, these non-participants may be enrolled in other training programs, receive welfare, or re-enroll in school. The cost-effectiveness of the boot camp depends on its incremental effectiveness and its incremental costs – both effects and costs must be expressed incrementally. However, in most studies we find that the resource use of the non-participants is rarely calculated. Failure to measure incrementally is likely to bias interventions toward less favorable cost-effectiveness.

Each of these four extra methodological challenges – within the context of the broader question of the legitimacy of comparing programs – has implications for cost analysis. One implication is that cost data should be collected in a way that is consistent with how effectiveness is measured. Ideally, cost data should be collected based on the actual implementation of the intervention from which the effectiveness data is derived. Where there are multiple sites, cost data should be collected for the actual sites for which effectiveness data is presented in order to determine what resource use is required to obtain observed levels of impact. Cost data should be collected with cognizance as to what resources are allocated to participants who are actually treated, as distinct from the resources allocated to those whom the intervention is intended to treat. Finally, cost data should be calculated with respect to the incremental resources of the program relative to what would be allocated under the *status quo*.

### *Facing these Challenges: The Example of Dropout Prevention Programs*

In our work on dropout prevention programs, we have found that each of these challenges is real. The practical challenge of collecting costs data is substantial. Dropout prevention programs that have been found to be effective were often implemented over a decade ago.

Collecting information on resource use that long ago is hard. Information on input prices is difficult to ascertain: ideally, we would want information on the prevailing prices at the time the program was implemented. These prices are not typically available, and it is necessary to use current prices.

The methodological challenges are also evident. We have referred above to the real challenge of comparing programs that serve very different populations and that have multiple objectives. We should note that this challenge is not specific to cost-effectiveness analysis but spans the general evaluation literature. However, CEA brings this challenge into greater focus by directly asking how multiple objectives can be compared.

Each of the dropout prevention programs that we investigated exhibited significant site-specific variation in both costs and effectiveness. (Insofar as we only used research evidence that showed the program to be effective, we did not factor in study-specific variation). Some sites had effects that were deleterious – the dropout rate increased – and some sites used resources that were almost three times the amount used at other sites.

Each program had some sample attrition, such that there were meaningful differences between the ITT and the TOT samples. Although we were able to adjust for these differences in our calculations, the difference – or at least its significance – was not clearly identified in the reported evaluations.

Finally, as noted above, one of the dropout programs we looked at, Talent Search, was very different from the others. Talent Search was delivered within the school setting. The costs of the program were therefore an increment on top of the resources already being used in the school. In contrast, the other four programs were delivered out of school.

The resources used for these programs represented the entirety of the investment in these students to help them complete school.

### *Sensitivity Testing*

In light of the above challenges to performing CEA, and to comport with general practices in social science, the results from any CEA should be subject to sensitivity testing. It is hoped that the sensitivity analysis would identify that a cost-effectiveness ratio could be reliably estimated and that alternative assumptions would not greatly influence the policy implications for decision-makers.

Although there are accepted methods for performing sensitivity analysis, there are no clear statistical tests that can be applied (such as an F-test or t-test between cost-effectiveness ratios). Moreover, there are multiple sensitivity tests that might be applied. These include best-case, worst-case scenarios and Monte Carlo simulation (Boardman, Greenberg, Vining, & Weimer, 2011). Here, too, however, there is no external standard against which to evaluate these sensitivity tests. Ultimately, the decision-maker must decide on how robust the findings are on the basis of the sensitivity tests. In the case of our analysis of dropout prevention programs, we found that the overall cost-effectiveness ratio for a program was not a reliable indicator of cost-effectiveness at the site level. There were few straightforward policy conclusions on the relationship between costs and effectiveness for programs that reduce the high school dropout rate. This was the case both across programs and across sites within a given program.

## 4. Conclusions

For many reasons, however, cost-effectiveness analysis has not been widely adopted by the research or policy community. Levin (2001) outlined these reasons a decade ago and they are still valid. We suspect that one important reason is that policymakers do not want to hear what CEA might tell them. Most policymakers who advocate for particular educational investments or programs focus on whether the program is effective or not. Having passed a test of effectiveness, policymakers do not then want to have a pass a second test with regard to resource use. CEA raises the possibility that even effective programs may be too expensive to warrant implementation. This argument also applies to those who implement programs and those who might have a stake in the programs recognition for funding. Given this antipathy, the demand for CEA is low. As a result, few colleges supply trained cost-effectiveness analysts.

Of course, we recognize that cost-effectiveness analysis has its challenges. Indeed, we have added to these challenges both at the practical and methodological level. However, we do not believe that these challenges undermine the importance of CEA. Instead, we believe that CEA can make a significant contribution in helping decision-makers allocate scarce resources.

Our analysis illustrates a broad set of challenges in trying to provide accurate information about cost-effectiveness to decision-makers. We hope these challenges can serve as lessons for future research. Perhaps the most important lesson that we can offer to others attempting to conduct cost-effectiveness analysis is that, whenever possible, costs and effectiveness for an intervention should be assessed concurrently at the same sites and based on the same sample of study participants and the same time period. Just as it is important to determine effectiveness for a comparison group, it is also important to determine costs of business-as-usual in order to correctly identify the additional resources required by an intervention. Most interventions are 'incremental' in the sense that they represent investments beyond what students have already received or are otherwise entitled to. Data on impact and costs should be accompanied by descriptions of program implementation such that decision-makers can understand what inputs are required. These descriptions are needed at the site-level and evaluators must address site-specific variation in effectiveness and cost-effectiveness. In light of all the above cautions, a last important lesson is that sensitivity testing is necessary, particularly when effectiveness and costs data are not well integrated. Due to site-level variation, overall or pooled estimates of cost-effectiveness may not be reliable. That is, it is less predictable whether the intervention is likely to have the effect that is estimated for the overall program because of the high variance in results from site to site.

It is almost certain that these challenges apply to many other interventions intended to improve educational outcomes. Indeed, our selection of dropout prevention programs was motivated by the belief that, in some respects, it would be relatively easy to apply cost-effectiveness analysis to them – and that the effectiveness of these programs had been independently validated. We strongly suspect that the challenges we identify are even greater for broader, more wide-sweeping reforms. For example, school choice and competitive reforms change not only the amounts of resource used, but also how they are allocated (see Levin & Driver, 1997). Similarly, accountability standards and exit exams involve many different changes to how education is delivered and assessed (Figlio & Rouse, 2006; Dee & Jacob, 2006). Perhaps the most intractable example is the 'whole-school' reform movement (see Levin, 2002). Whole-school reforms are often advocated as a way to change the culture and organization of schools to ensure greater learning. Overall, economic analysis of whole-school reforms is incomplete, despite the substantial cost involved in implementing them. Many of the challenges in conducting economic analysis in relation to a comparison group are especially pertinent to whole-school reform. Thus, the research field appears to be far short of helping policymakers create the cost-effective school or school system.

Ultimately, cost-effectiveness analysis is intended to help decision-makers allocate limited resources to maximum effect. Performed in a timely and judicious manner, we believe that it can be a great help. But we recognize that, although it is a key piece of information, it is not the only criterion that should be used. If the decision-maker needs to work within a limited budget, as might be the case for a school principal, a table of cost-effectiveness ratios for the interventions addressing the outcome of interest can help determine which intervention would provide the most impact for the lowest cost. If the budget limit is not yet established, as might be the case for a legislator with flexibility to allocate funds across social program areas, the data could be used to identify the interventions that provide the greatest desirable impact and are also politically most feasible. If the decision-maker faces an external constraint, such as a union contract or class-size limit, the data are only useful if they lead to decisions that satisfy this constraint. (For example, the STAR experiment that found smaller classes were much more effective is far too costly for most districts or states to implement, see Finn & Achilles, 1999; Brewer, Krop, Gill, & Reichardt, 1999; Prais, 1996). Another consideration is whether the resources required to implement an intervention would be easily available at comparable costs in the decision-maker's jurisdiction. If, for example, an intervention required additional specialists in reading and the decision-maker was aware of a shortage of such individuals, (s)he might anticipate that the local costs would be higher than the national average and the program's cost-effectiveness would be lower. In sum, the decision-maker must use his/her knowledge of what programs are likely to be acceptable in his/her area of jurisdiction from all these standpoints: financial; regulatory; political; and contextual.

Decision-makers must therefore recognize how to interpret evidence on cost-effectiveness. Indeed, decision-makers can use contextual information to analyze possible reasons for drastic differences between cost-effectiveness ratios for various alternatives. For example, a comparison between the cost-effectiveness ratio for a highly targeted program with a single outcome and a broad-based program with multiple outcomes may be misleadingly unfavorable to the program with multiple outcomes, as the cost in question is, in effect, "buying" more than the single measured effect. Programs may also differ in their level of targeting; for example, dropout prevention programs that successfully target high school students just on the margin between dropping out and graduating will, all else being equal, appear to be more cost-effective than those that serve a broader population. Decision-makers must consider whether such targeting could realistically be expected when implementing programs in their own context, or, if a program may have benefits to participants beyond the immediate intended effect, if that targeting is even desirable.

As noted above, cost-effectiveness data is only comparable across interventions that address the same educational outcome. However, if the decision-maker is attempting to decide how to allocate a budget among programs addressing different educational outcomes, for example reducing dropouts *vs*. increasing early literacy, the required comparison would be a cost-benefit analysis where the different types of outcomes would need to be translated into financial benefits (e.g., higher earnings). For some policy evaluations, cost-effectiveness may be superseded by cost-benefit analysis.

Despite the methodological challenges to, and circumscribed role for, cost-effectiveness analysis, it does convey important information about educational investments and this information cannot be obtained using any other method of evaluation. Thus, it is essential that such analysis be performed rigorously and transparently and directed to help inform policy decisions that promote educational efficiency.

## References


Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2011). *Cost-benefit analysis: Concepts and practice* (4th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Borman, G.D. & G.M. Hewes. (2002). The Long-Term Effects and Cost-Effectiveness of Success for All. *Educational Evaluation and Policy Analysis* 24(4), 243-266.

Brewer, D., Krop, C., Gill, B. & R. Reichardt. (1999). Estimating the Costs of National Class Size Reductions Under Different Policy Alternatives. *Educational Evaluation and Policy Analysis* 21(2), 179-192.

Center on Budget and Policy Priorities. (2012). *New School Year Brings More Cuts in State Funding for Schools.* Retrieved from: http://www.cbpp.org/files/9-4-12sfp.pdf

Clune, W. H. (2002). Methodological strength and policy usefulness of cost-effectiveness research. In Levin, H. M., & McEwan, P. J. (Eds.), *Cost-effectiveness and educational policy*, p. 55-70. Larchmont, NY: Eye on Education.

Dee, T. & B. Jacob. (2006). Do High School Exit Exams Influence High School Achievement and Labor Market Performance? NBER Working Paper, #12199.

Figlio, D. & C. Rouse. (2006). Do Accountability and Voucher Threats Improve Low Performing Schools? *Journal of Public Economics* 90(1-2), 239-255.

Finn, J. & C. Achilles. (1999). Tennessee's Class Size Study: Findings, Implications, Misconceptions. *Educational Evaluation and Policy Analysis* 21(2), 97-109.

Harris, D. (2009). Toward policy-relevant benchmarks for interpreting effect sizes combining effects with costs. *Educational Evaluation and Policy Analysis, 31*(1), 3-29.

Hanushek, E. (2004). What if There are No 'Best Practices'? *Scottish Journal of Political Economy* 51(2), 156-172.

Heckman, J., Humphries, J. & Mader, N. (2010). *The GED*. National Bureau of Economic Research (NBER) Working Paper 16064.

Hummel-Rossi, B., & Ashdown, J. (2002). The State of Cost-Benefit and Cost-Effectiveness Analyses in Education. *Review of Educational Research, Vol. 72 (1),* p. 1 - 30.

Institute of Education Sciences (IES) (2012). *WWC Dropout prevention intervention reports.* Retrieved on July 10, 2012 from: http://ies.ed.gov/ncee/wwc/topic.aspx?sid=3

Levin, H. M. (1970). A cost-effectiveness analysis of teacher selection. *The Journal of Human Resources, 5*(1), 24-33.

Levin, H. M. (1975). Cost-effectiveness analysis in evaluation research. In M. Guttentag, & E. L. Struening (Eds.), *Handbook of evaluation research (Volume 2)*. Beverly Hills, CA: Sage.

Levin, H. M. (1983). *Cost-effectiveness analysis: A primer*. Beverly Hills, CA: Sage.

Levin, H.M. (2001). Waiting for Godot: Cost-Effectiveness Analysis in Education. In Richard Light (Ed.) *Evaluations that Surprise.* San Francisco, Cal.: Jossey-Bass.

Levin, H.M. (2002) Issues in Designing Cost-Effectiveness Comparisons of Whole-School Reforms. In Levin, HM & PJ McEwan (Eds.) *Cost-Effectiveness Analysis and Educational Policy*. Larchmont, N.J.: AEFA Yearbook: Eye on Education.

Levin, H.M., Belfield, C.R., Hollands, F., Bowden, B., Cheng, H., Shand, R., Pan, Y-L., and Hanisch-Cerda, B. (2012). *The Cost-Effectiveness of High School Completion Programs.* Monograph, Center for Benefit-Cost Studies in Education, www.cbcse.org.

Levin, H.M., Catlin, D., & Elson, A. (2007). Costs of implementing adolescent literacy programs. In D. Deshler, A. S. Palincsar, G. Biancarosa & M. Nair (Eds.), Informed choices for struggling adolescent readers: A research-based guide to instructional programs and practices (pp. 61-91). Newark, DE: International Reading Association.

Levin, H.M. & Driver, C. (1997). Costs of an Educational Voucher System. *Education Economics* 5(3), 303-311.

Levin, H.M, Glass, G.V & Meister, G.R. (1987). Cost-Effectiveness of Computer-Assisted Instruction. *Evaluation Review* 11(1), 50-72.

Levin, H.M. & P. McEwan (2001) *Cost-Effectiveness Analysis* (2nd Edition). New York: Sage Publications.

Prais, S. (1996). Class Size and Learning: The Tennessee Experiment – What Follows? *Oxford Review of Education* 22(4), 399-414.

U.S. Government Spending (2012). *U.S. Government spending*. Retrieved from http://www.usgovernmentspending.com/us_education_spending_20.html.