# 9

# COST-BENEFIT ANALYSIS AND COST-EFFECTIVENESS ANALYSIS

CLIVE R. BELFIELD

## INTRODUCTION

Cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) are methods for evaluating interventions, reforms, and policies. CBA and CEA supplement commonly used evaluation methods related to effectiveness and efficacy by adding an economic component to the evaluation. There is a strong case for some form of economic evaluation across most government programs, including those in the education sector. Every education program requires resources, and the impact of the program may depend on the quantity of resources applied. For example, a mentoring program for a school of 100 students may have five counselors or 10 counselors; an after-school program may operate in a school playground or have extensive, dedicated facilities; and a teacher professional development program may run for one week or six weeks. In each example, the more resource-intensive program might be expected to improve education outcomes to a greater extent. CBA and CEA incorporate resource use directly into an evaluation to test for efficiency—i.e., whether the resources are being used in the most effective way to generate the greatest possible improvement.

There are two senses of efficiency: one refers to investments that yield outcomes that society values (allocative efficiency); the other refers to using inputs in the most effective way to produce the desired outcomes (technical efficiency). CBA indicates whether an intervention is (allocatively) efficient by looking at whether the dollar costs of the resources needed for implementation are less than the dollar benefits that the intervention produces. CEA indicates which intervention is the most (technically) efficient by looking at which intervention generates the biggest improvement in educational objectives per dollar spent.

CBA and CEA apply the basic economic concept of opportunity cost to educational interventions. All interventions require resources, and the appropriate way to evaluate them is by comparing them to the next best use of equivalent resources. For example, if a school is choosing across a range of curricular interventions to improve outcomes for struggling readers, it should look at the resources required to implement each intervention in conjunction with its effect on reading. Then, the school should choose the most cost-effective intervention—i.e., the one that improves reading outcomes by a given amount at the lowest cost (or by the largest amount for a given cost). If a state is considering expansion of its college programs, it should compare the costs of the programs to the returns from graduates' higher earnings, as well as other benefits. By applying the concept of opportunity cost, these analyses take account of the fact that resources are scarce and that programs should be chosen based either on how effectively they use the available resources or on what return is made on the investment.

CBA and CEA are companion forms of analysis: they share many common principles and require similar data. Both include analysis of how resources are used to implement education programs. But the two methods differ significantly in how they analyze the outcomes of a given program. CEA relates costs to effectiveness, where the latter is denominated using an education metric (such as a graduation rate or reading gain). The results of a CEA are presented as a cost-effectiveness ratio where the total costs are divided by the resulting outcomes. The intervention with the lowest ratio is the most cost-effective, that is, it is "buying" outcomes at the lowest price. Thus, CEA helps identify which intervention is the most efficient at achieving a particular set of desired effects (relative to other interventions). CBA requires an additional step, which is to translate measures of effectiveness into monetary values. After this translation, a benefit-cost (or cost-benefit) ratio can be calculated with both the numerator and denominator expressed in monetary units. The program with the highest benefit-cost ratio is the most efficient, that is, it yields the most dollars back per dollar spent on the program. The advantage of cost-benefit analysis is that it determines whether an intervention is worth implementing at all, that is, whether the benefits exceed the costs (or the benefit-cost ratio exceeds one). The disadvantage is that CBA requires considerably more data and analysis.

CBA and CEA are intended to help education policymakers and professionals make decisions. As educational budgets are constrained and not all programs can be funded, it is critical for decision makers to have information on which interventions are the most cost effective. CBA and CEA can be readily applied in education research to help decision makers evaluate many interventions, reforms, and policies. Their application is especially useful in the public sector where there is no obvious driver (such as the profit motive) to ensure efficiency. Given their differences, an evaluator would typically apply either CBA or CEA and not both together. If a decision maker has already specified the educational objective and is seeking the best way to achieve this objective, cost-effectiveness analysis is sufficient. If the decision maker is uncertain that a program is worth the investment, then cost-benefit analysis is needed.

However, cost-benefit and cost-effectiveness ratios do not compel decision makers to choose a particular policy. Other factors, such as political considerations or budget constraints, may be influential. For example, interventions where the costs exceed the benefits or those with a high cost-effectiveness ratio may still be chosen if they serve important social goals (such as giving all students an equal opportunity to learn or helping disadvantaged students). The presumption is that, absent any policy constraints, alternative policy goals, or equity considerations, the intervention with the lowest cost-benefit or cost-effectiveness ratio should be preferred. But it is still up to the policymaker to make this determination.

The method of cost-effectiveness analysis, at least in its application for education research, was developed by Henry Levin in his seminal textbook of 1983, with a second edition published in 2001 with coauthor Patrick McEwan. (The general method of cost-benefit analysis outside education was established in the 1930s for water resource projects and over time as been applied to a wide range of policy interventions; see Gramlich, 1981, and Boardman et al., 2011.) In terms of basic principles, the respective methods of CEA and CBA have not changed significantly in the intervening years. Yet education researchers have sparsely applied CBA and very sparsely applied CEA. In contrast, the literature on program impacts, which pays no attention to the cost of the programs, is large and methodologically more sophisticated. Thus, policymakers are gaining a better understanding of what works, but the evidence on what interventions, programs, and policies are comparatively most efficient is often not available.

In this chapter I set down the companion methods of CBA and CEA, beginning with a description of costs analysis. I then look at cost-effectiveness analysis, as this is conceptually less complex than CBA. I show how to integrate costs with measures of effectiveness to yield the key metric—the cost-effectiveness ratio. I then identify some important methodological challenges to performing CEA. Next, I conclude with a summary of current practice, along with some case studies as examples for future reference. I perform a parallel exercise for CBA with a discussion of metrics, methodological challenges, and evidence from policy areas where CBA has been used. In the final section I consider how education policy might be improved by more extensive use of CBA and CEA.

## ANALYSIS OF COSTS

Economic evaluations focus on the resources required to implement the program. The monetary value of these resources is referred to as the cost of the program. For both CBA and CEA, costs data should be collected using the ingredients method (Levin & McEwan, 2001; Tsang, 1997). This method requires the specification of all the ingredients or inputs that are used to implement the intervention net of the costs of the alternative or "business as usual" program. Thus, the costs are expressed as incremental beyond what is usually spent. Many education interventions change or supplement only a part of the total education a student receives: an after-school program, for example, adds extra schooling on top of the regular school day; a new reading curriculum affects only class time spent on reading. It is critical to estimate costs in relation to business as usual and with recognition that the program costs are incremental on top of other resources that improve educational outcomes.

All inputs should be costed out, regardless of who pays for them, but the financing source should be identified. Resources for education are funded from many sources, including several levels of government and students' families, as well as local agencies (businesses or community groups). Typically, the total cost of all these inputs is reported. However, costs for each perspective and funding agency (such as a school, district, or level of government) may be reported separately if needed. Conventionally, there are three perspectives. The primary perspective is the social, as this includes all resources. But education programs may also be evaluated from the perspective of the private individual participating in the program. Also, they may be evaluated from the fiscal perspective, that is, accounting only for the resources associated with government (and funded by the taxpayer). Results may differ depending on the perspective. For example, a school might contribute teacher time only to a program where the materials and facilities are funded by the federal government. From the school's perspective, but not necessarily from a broader social perspective, this program might look very cost-effective.

Program ingredients should be grouped into personnel, materials/equipment, facilities, and other. As education is labor intensive, the largest ingredient in most reforms is likely to be personnel and especially teacher time. For some interventions, such as computer-based instruction or online learning, materials and equipment may be significant. Careful attention must be paid to amortizing this equipment over the life of the intervention: one claim made by providers of online learning, for example, is that the high initial investment lasts many years and reduces the costs of other inputs. For other educational programs, parent and student time commitments may be significant. Accurate accounting for student time is critical: many college and high school students may wish to work instead of enroll in supplemental education programs and so lost wages represent their opportunity cost of participation. Typically, collecting this information requires semistructured instruments administered to key personnel responsible for implementing the intervention (or observations of practice). These instruments should be framed to determine what ingredients are needed (not what amounts were spent). All these ingredients should be tabulated in a spreadsheet.

Once tabulated, each ingredient should be priced out using independent prices from appropriate datasets of agencies such as the Bureau of Labor Statistics or the National Center for Educational Statistics. Prices should capture what it would cost to purchase these ingredients in a perfectly competitive market. Many ingredients in education are purchased in distorted markets (e.g., ones with significant government regulations) and so attention must be paid to whether the prices used reflect their opportunity cost. For parent and student time, for example, it may be appropriate to use the market wage as the opportunity cost of time. But parents may enjoy helping out in schools and so require a lower "wage." There is no consensus on how to adjust for this enjoyment; typically, the opportunity cost of parent time is accounted for by using a relatively low wage (e.g., the minimum wage) and applying this uniformly across all parents regardless of their occupation or employment status. All prices can be adjusted for inflation and local price indices applied as appropriate. The cost of the intervention is then reported as the total sum of all ingredients multiplied by their unit prices.

An example of the ingredients method is given in Table 9.1 for the federal Talent Search program (Bowden, 2013). Talent Search, initially established and funded through the 1965 Higher Education Act, is a program for high school students that includes tutoring, financial awareness and career selection training, college tours, and assistance with all aspects of applying to and enrolling in college. The inputs to the program include services provided by Talent Search offices and time and resources from the high schools and local community groups. Table 9.1 shows the inputs and the resources for each input aligned with the agency responsible for funding these resources. The final column shows the total resource requirement for Talent Search. The data are based on the inputs reported from interviews with five currently operating Talent Search sites. Information on the prices for personnel was taken from the National Compensation Survey (Bureau of Labor Statistics) and from the National Education Association salary survey. These prices (wages) were based on the respective occupations of the personnel, accounting for their experience and qualifications, and were also adjusted to account for fringe benefits where appropriate. Information on the prices for facilities was based on the rental rate for comparable office space. Information for other inputs was based on market prices for specific equipment and materials. The opportunity cost of time of other personnel (volunteers) was based on wage rates for workers with comparable skills and experience. As shown in Table 9.1, the largest

Table 9.1 Cost Spreadsheet Using the Ingredients Method

| | Costs (Inputs x Prices) | | | |
|---|---|---|---|---|
| | Talent Search agency | School site | Other agents | Total |
| **Personnel:** | | | | |
| TS staff: directors | $78,390 | | | $78,390 |
| TS staff: counselors (Level A) | $98,700 | | | $98,700 |
| TS staff: counselors (Level B) | $78,870 | | | $78,870 |
| TS staff: other | $57,470 | | | $57,470 |
| TS Work Study | $4,640 | | | $4,640 |
| TS staff: professional development | $2,860 | | | $2,860 |
| School staff: principals/teachers | | $3,700 | | $3,700 |
| School staff: guidance counselors | | $7,160 | | $7,160 |
| School staff: other | | $5,200 | | $5,200 |
| In-kind personnel | | | $1,460 | $1,460 |
| **Facilities:** | | | | |
| Host college | $15,150 | | | $15,150 |
| School sites | | $3,720 | | $3,720 |
| Overhead | $23,530 | | | $23,530 |
| **Materials/equipment:** | | | | |
| TS site | $10,030 | | | $10,030 |
| Contributed | | | $2,620 | $2,620 |
| **Other Inputs:** | | | | |
| Transportation | $18,010 | | | $18,010 |
| Other TS inputs | $3,670 | | | $3,670 |
| Other in-kind inputs | | | $22,880 | $22,880 |
| **Total annual cost** | $391,320 | $19,780 | $26,960 | $438,060 |
| **Students served** | 790 | 790 | 790 | 790 |
| **Annual cost per student** | $495 | $25 | $34 | $555 |

*Source:* Adapted from Hollands et al. (in press, Table 1). 2010 dollars, rounded to $10. Average across five Talent Search sites.

proportion of the total cost is for personnel. Although the Talent Search agencies incur most of the costs of providing the program, there are significant resource commitments at the school site and from other groups (nonprofit agencies).

Costs analysis of Talent Search shows variation across sites in what resources are used. Other cost analyses have shown variation in how much resource is used. For example, a cost analysis of reading programs using the ingredients method is reported by Levin, Catlin, and Elson (2007). While the reading program was intended to be applied uniformly across sites, Levin et al. (2007) found substantial variation across sites in the total costs of implementation, with three times as much spent at the most resource-intensive site compared to the least intensive.

Crucially, cost estimates should not be derived from budgetary information. Agency budgets are rarely comprehensive, covering all relevant ingredients, and are almost never itemized in a way that clearly identifies how much is spent on a given intervention. Using budgets it is very difficult to disentangle the incremental costs of an education reform from the regular school or college operations. Further, budget statements reflect local prices and not what an intervention would cost if another agency decided to implement it in its local context. If budgets were used, interventions in areas with high prices would be disadvantaged in any economic comparisons.

Some interventions may induce additional resource use after the program is implemented. For example, a high school intervention to improve access to college will lead to increased spending for those students who now attend college. For clarity, it is better to separate out these "induced costs" (sometimes called "indirect costs") from program delivery costs. For CEA, these induced costs may be included in the total costs (depending on whether they are necessary to effect the project goals). For CBA, these induced costs are often more accurately labeled as "negative benefits." This is especially important when the results are reported as a benefit-cost ratio (see discussion below) because labeling them as costs will change the benefit-cost ratio by inflating the denominator.

## COST-EFFECTIVENESS ANALYSIS

CEA is a very flexible method by which to supplement effectiveness analysis. It is possible to use any domain of effectiveness for cost-effectiveness analysis. That is, one can perform CEA for interventions that improve phonics outcomes for struggling readers, high school graduation rates for seniors, or credits accumulated by college students. As long as the outcomes can be validly attributed to the intervention, they are legitimate measures of effectiveness for CEA.

However, effectiveness must be enumerated into a singular measure. If an intervention has multiple outcomes then these outcomes must be weighted to yield a singular measure of overall effectiveness. For example, an after-school program may improve students' academic and social skills, and both outcomes should be included if they contribute to the outcome of interest such as graduation from the program. Weights may be determined by reference to theory or according to decision makers' preferences. For example, educational programs may be funded based on how many students graduate from the program; that funding imperative will influence decision makers' preferences. A more formal approach is to base the weights on an explicit cost-utility analysis (CUA). This type of analysis requires information on the value of each outcome. This value is based on preferences, which are typically elicited from education professionals or expert panels (or sometimes from parents). Once this valuation has been established, CUA follows the same method as CEA with utility substituting for effectiveness.

### *Cost-Effectiveness Ratios*

In principle, the cost-effectiveness ratio is straightforward to calculate. It is an expression of the total costs of an intervention divided by the effects of that intervention. Its interpretation is also straightforward, again in principle. Interventions with the lowest cost per unit of effectiveness (smallest ratios) are the most cost-effective. In some cases, cost-effectiveness may be reported as the increase in effectiveness per dollar spent.

However, one must be very clear about how this ratio is actually derived. CEA is a comparative evaluation method. There must be at least two alternatives for comparison. At a basic level, the scenarios must be comparable in several important respects. As noted above, they must be intended to improve the same outcomes and these must be enumerated in a singular form. As well, the scenarios must be of similar scale, covering equivalent numbers of students. Large interventions cannot be easily compared to small interventions, because cost estimates cannot be easily extrapolated up or down. The example of a small-scale, teacher-led classroom intervention that is to be implemented statewide illustrates the three reasons extrapolation becomes difficult. First, the cost per student will increase: to recruit many extra teachers across the state, higher wages will have to be offered. Second, as the absolute cost of the intervention rises, it is likely that the state will hit a financing constraint or face a higher interest rate to borrow funds to implement the program. Third, the intervention may exhibit economies or diseconomies of scale: implemented statewide, the intervention may be more effective (through peer learning) or lower cost (with savings on administration and management). Input prices, financing constraints, and economies of scale may all be uncertain when extrapolating from a small to a large program.

The comparative aspect of CEA also has implications for the interpretation of the results. For a field trial outcomes for the treatment and control groups are compared. Both groups probably receive some educational resources (in regular classroom settings), but the treatment group receives more or different resources, for example through extra tutoring or in an after-school program. The costs of the intervention are therefore incremental costs, that is, they are extra resources beyond those all children (or all children in the control group) receive. The cost-effectiveness metric is therefore an incremental cost-effectiveness ratio (ICER). It expresses how the extra resources for the treatment generate gains in outcomes relative to the control group. With ICERs, the treatment is not "buying" outcomes but buying improvements in outcomes.

Expressed using these metrics, the value of cost-effectiveness analysis becomes clear. Interventions that are highly effective in the sense of generating big gains in valued outcomes may not be very cost-effective. That is, the reason why they are so effective might be because they require a large amount of resources. Instead, other interventions might be only slightly less effective yet sufficiently cheaper, and thus more cost efficient. CEA can help identify dominated interventions (those that are less effective and more costly) and dominating interventions (those that are more effective and less costly). However, a possible scenario is that interventions are like consumer products: the more expensive the intervention, the more effective it is. If so, the decision maker would need additional information to select an intervention using additional information, such as which interventions are affordable.

### *Empirical and Methodological Challenges*

Cost-effectiveness analysis has been applied in only the most limited fashion in education research. Many studies that claim to have performed such analysis have typically made only a cursory or partial attempt. In part, this situation has arisen because of, as noted by Harris (2009) and in writings by Henry Levin, the empirical and methodological challenges to performing CEA.

Obtaining cost data is challenging. Potentially, it requires access to many personnel, including those in charge of managing an intervention, those delivering it, and the students experiencing it. These personnel will not typically have collected the information needed on costs, and so multiple interview and survey instruments must be designed and administered. Cost estimates should be calculated during the intervention and not after the evaluation of effectiveness. When collected retrospectively, cost data are much less accurate: personnel cannot easily account for their time dedicated to an intervention;

many of the personnel have taken on different roles; and typically there is limited archival data on how an intervention was implemented. Finally, as with effects, costs will vary across participants or sites. However, because costs are bounded at zero, this variation is unlikely to follow a normal distribution. The challenge is compounded by the fact that rarely do sampling frames for evaluations explicitly address variation in resource use. When such variation is discovered, it is then hard to estimate variation across the population. (Most likely, evaluation sites will have been chosen based on their power to detect statistically significant impacts not statistically significant cost-effectiveness.)

Integrating costs and effectiveness data is also challenging. This is the case even when effectiveness has been rigorously identified, for example through a field trial. Very few educational interventions have only one single outcome, so it may be necessary to apply weights or to apportion costs to particular outcomes (separating out the costs into those designed to improve each outcome). Yet there is little consensus on how to weight different effects. Effect size gains (differences in means divided by the standard deviation) are also difficult to interpret in terms of incremental cost-effectiveness ratios. These effect size gains depend on the variance in the populations being studied, so comparisons should only be made across interventions with the same variance in effects. Also, studies tend to report effects either for the intent-to-treat or treatment-on-the-treated samples. This choice has implications for how resources should be calculated. The intent-to-treat sample is often larger than those who actually receive the treatment and so the total costs will be higher.

Finally, cost-effectiveness results should be tested for robustness, that is, to see how ratios change with variations in estimates of costs or effects. Sensitivity analysis may be performed by selecting alternative values for the most important parameter or by reporting best-case and worst-case scenarios. Sensitivity analysis is necessary for several reasons. Many analyses are based on small samples with wide sampling error in estimates both of costs and impacts. Also, the variance in costs may not follow a normal distribution and so may preclude standard statistical tests for differences in means. Given variance in costs and variance in effects, the variance in cost-effectiveness may be difficult to predict. Potentially, the analysis may yield imprecise estimates of cost efficiency (particularly if the results are disaggregated to the site level), and this imprecision should be reported explicitly in a formal sensitivity analysis. As noted above, scale is one important consideration. Another is that input prices may change as technology changes (e.g., for school computing facilities).

## EXAMPLES AND FINDINGS

There are few exemplary cost-effectiveness analyses in education (Levin & McEwan, 2002; Levin, 2001). In a review of over 1,300 relevant academic papers in education on cost-effectiveness, Clune (2002) divided them on a quality scale as follows: 56 percent "rhetorical," 27 percent "minimal," 15 percent "substantial," and 2 percent "plausible." As an indicator of the low quality of the research, Clune's (2002) definition of "substantial" was an "attempt to mount data on cost and effectiveness but with serious flaws." "Rhetorical" was defined as "cost-effectiveness claims with no data on either costs or effects." CEA of education policy has not improved considerably since Clune's (2002) review. This paucity of evidence means that it is only possible to speculate which programs might be cost-effective (Harris, 2009).

One early example of CEA is the investigation by Henry Levin, Gene Glass, and Gail Meister (1987) of computer-assisted instruction. The authors compared the effectiveness of four interventions in improving mathematics and reading performance of elementary school children. These interventions were computer-assisted instruction; cross-age tutoring; reducing class size; and increasing instructional time. Each intervention had been found to be effective at improving achievement, but there was substantial variation in the costs of implementing each intervention. By comparing cost-effectiveness ratios, cross-age tutoring appeared significantly more efficient than computer-assisted instruction, which in turn was more efficient than either reducing class size or increasing instructional time. Peer-tutoring appeared most efficient for two reasons: the peers and the tutors both improve their achievement levels; and the time of the peers is valued at an opportunity cost of zero, as these students would be studying in class anyway.

An example of a recently completed cost-effectiveness analysis is given in Table 9.2. This CEA is from Hollands et al. (in press), who reviewed all interventions that have been proven to increase the high school graduation rate and performed cost-effectiveness analyses on those for which both effectiveness and costs data could be obtained. The effectiveness measure was the incremental gain in high school graduates as a result of the intervention: columns 2–4 show that the interventions increased the high school graduation rate by 9–20 percentage points. Cost data were collected from archival sources and, for Talent Search, from a direct application of the ingredients method. Column 6 of Table 9.2 shows the large differences in the cost per participant for each intervention, ranging from $3,290 to $22,290. The cost-effectiveness ratios are reported in column 7. The CE ratio is the total cost across the entire sample (entries in column 1 times those in column 6) divided by the yield of new graduates (column 5). The cost for the entire sample must be used, even as some of these students would have graduated anyway, and some do not graduate even after the intervention. This CE ratio can be interpreted as the incremental cost to yield a new high school graduate. Lower ratios are therefore preferred as these represent lower costs for ensuring a student graduates from high school. There are sizeable differences in the cost-effectiveness of these interventions: Talent Search is the most cost-effective, with the other interventions being less than half as efficient. Per dollar, investments in Talent Search will yield much greater improvements in the high school graduation rate than investments in other interventions. However, as noted above,

**Table 9.2** Cost-Effectiveness Results for Programs That Improve the Rate of High School Completion

| Program | (1) Treatment group size | (2) Treatment group graduation rate | (3) Control group graduation rate | (4) Percentage point gain (2) – (3) | (5) Yield of new graduates | (6) Cost per participant | (7) CE ratio (1)*(6)/(5) |
|---|---|---|---|---|---|---|---|
| *Intervention for students in school:* | | | | | | | |
| Talent Search | 3,930 | 82.5% | 71.8% | 10.8 | 423 | $3,290 | $30,520 |
| *Interventions for youth who have dropped out of school:* | | | | | | | |
| NGYC | 596 | 87.0% | 67.3% | 19.8 | 118 | $14,100 | $71,220 |
| Job Corps | 3,940 | 51.4% | 34.4% | 17.0 | 670 | $22,290 | $131,140 |
| JOBSTART | 1,028 | 44.4% | 29.3% | 15.1 | 155 | $10,460 | $69,510 |
| *Intervention for young mothers who have dropped out of school:* | | | | | | | |
| New Chance | 1,240 | 58.6% | 49.5% | 9.2 | 113 | $17,820 | $194,640 |

*Source*: Hollands et al. (in press, Table 2).

*Notes*: Treatment on the treated (TOT) estimates for percentage point gain of treatment group over control group. 2010 dollars.

a critical factor in interpreting cost-effectiveness ratios is the context of the decision. The interventions listed in Table 9.2 serve very different student groups. Talent Search is an intervention for students who are in school; the other interventions are for students who have already dropped out of school. Getting students who are in high school to complete their studies is much less costly than getting dropouts to complete their high school studies. Nevertheless, cost-effectiveness analysis provides information on the extent of the disparity in resources required to ensure high school graduation for different subgroups.

## COST-BENEFIT ANALYSIS

Cost-benefit analysis can potentially be performed for any educational intervention. Using the ingredients method, the costs of each intervention can be expressed in dollars. For CBA, the impacts and outcomes of the intervention are also expressed in monetary terms and labeled as the benefits. The simple logic of CBA is that policies are preferred where the benefits exceed the costs. The advantage of CBA over CEA is that each outcome of an educational intervention can be explicitly accounted for in that many outcomes over many periods of time can be considered. Consequently, interventions with different goals and different outcomes can be compared, including interventions unrelated to education. For example, the returns from an education program may be compared to those from a public infrastructure project. However, the validity of CBA depends on the extent to which outcomes can be expressed in dollars. If many outcomes cannot be monetized, then CBA may not be an appropriate way to evaluate an intervention, although even in this case it may be helpful to use the framework of CBA to systematically organize the analysis and identify the key unknowns.

Most educational outcomes—such as test score gains or increased rates of college completion—do not have clear market prices. They cannot be "bought," but they are valuable; they lead to a more productive citizenry. Hence, it makes economic sense for society to commit resources to obtain these outcomes. To translate these educational outcomes into dollar values, indirect methods, often called *shadow pricing techniques*, are needed. A *shadow price* is the amount of resource someone would be willing to pay in order to obtain a particular outcome.

There are several ways to calculate shadow prices and they may be calculated from the same three perspectives as used above for costs: individual, fiscal, and social. The most common shadow pricing approach is to add up all the economic benefits that flow from reaching a particular education level (or add up all the burdens from not being educated). This method is akin to the "cost of illness" method in the health economics literature (and might be called the "cost of ignorance" method in the education sector). For example, from the individual's perspective the shadow price of being a high school graduate as opposed to a high school dropout is the additional lifetime earnings associated with graduating from high school. From the fiscal perspective, the shadow price might be additional tax payments from being a high school graduate. This shadow price is conservative even as a measure of fiscal savings because it does not include the value of savings to the criminal justice, health, and welfare systems that arise from having a more educated populace (see Belfield & Levin, 2007). It is obviously even more limited as a measure of social benefits.

As another example, the shadow price of reducing the rate of special education may be calculated as the savings that arise from placing a student in mainstream instruction rather than special education classes. An alternative shadow pricing technique is to derive the economic value of education from the prices of other goods that achieve the same outcomes. For example, the shadow price of a high quality education can be calculated as the extra amount one has to pay for a house that is close to a high quality school. The shadow price for a higher test score might be calculated as the price paid for private tutoring that is needed to generate that test score gain. There are existing shadow prices for some educational outcomes that might be "plugged in" to an impact evaluation. For example, Belfield and Levin (2007) estimate the fiscal shadow price per additional high school graduate at $209,100 in 2005 dollars (see Karoly, 2012). But for most outcomes, the researcher must calculate these prices independently, and even when shadow prices exist, care must be used in transferring them to other contexts.

Shadow prices for educational outcomes must be adjusted to account for when these outcomes occur. So, a program in 10th grade that increases high school graduation rates (impacts) should lead to higher earnings (benefits) in adulthood. But these benefits lag behind the costs: the costs of the program were incurred in 10th grade and the flow of benefits will not occur for at least two years. The benefit flow must therefore be discounted on the grounds that money received in the future is worth less than money received in the present (see Moore et al., 2004). The practice of discounting involves progressively lowering the value of dollar amounts as they occur further into the future (and follows the practice illustrated in accounting textbooks). This weighting is reflected in the choice of the discount rate. Choosing a high discount rate means future benefits will be worth less than if a low discount rate were chosen. The value of the discount rate may vary depending on the types of benefits that accrue, but typically values of 3–7 percent are chosen for educational interventions (provided the benefits are reported in constant dollars—they have been deflated for price increases). When benefits and, if necessary, costs have been discounted, they are referred to as *present values*; these are the value figures that should be compared for the CBA.

### *Cost-Benefit Metrics*

The comparison of costs and benefits can be expressed using three different metrics. The first is *benefits minus costs*; as these are expressed in present values, this difference is called the net present value (NPV) of the educational investment. Interventions where the net present value is positive are worth undertaking, but given competing interventions the one with the largest net present value is preferred. The second metric is the *benefit-cost (B/C) ratio*, which is the benefits divided by the costs. If this B/C ratio exceeds one, the benefits are greater than the costs and the investment is worth undertaking. Typically, the intervention with the highest B/C ratio is preferred. The B/C ratio is quoted most often in CBAs because it is a useful summary of the economic value (as in "for each dollar spent on the program, $X dollars are returned in benefits"). Importantly, and unlike the net present value, the B/C ratio is independent of scale and so must be interpreted carefully. An education program that costs $1 million but yields benefits of $3 million has a B/C ratio of 3 and a net present value of $2 million. By comparison an educational intervention that costs $1,000 but yields benefits of $10,000 has a B/C ratio of 10 and a net present value of $9,000. In monetary terms, the former intervention would be preferred even as it has a lower B/C ratio.

The final metric is the *internal rate of return*. This is the discount rate that ensures that the present value of benefits equals the present value of costs. Instead of assuming a discount rate (of say 3 percent), calculating the present values of costs and benefits, and reporting the difference, the researcher derives a value for the discount rate such that the difference between the costs and benefits is equal to zero. This value of the discount rate is called the *internal rate of return (IRR)*. Again, educational interventions with high IRRs are worth undertaking, and the one with the highest IRR is preferred. For this metric, the IRR might be compared to an outside standard such as the rate of return on Treasury bonds or shares. So, if the government can borrow money at 4 percent and provide a reading program that yields a 10 percent return (IRR), then the program is a worthwhile investment.

Each of these metrics can be calculated from each perspective. For the private individual, most education interventions easily pass a cost-benefit test because the student typically pays only a small proportion of the costs but reaps a large proportion of the benefits (e.g. higher earnings). From the fiscal perspective, the calculus is less clear-cut; taxpayers often pay a large proportion of the costs but reap a relatively small proportion of the benefits (e.g. increased tax revenues on higher earnings). For society, the full resource costs must be compared against the full social benefits.

### *Empirical and Methodological Challenges*

As with CEA, there are both empirical and methodological challenges to overcome in conducting CBA. As CBA applies the same method of costing as CEA, the same challenges as noted earlier apply in obtaining costs data and linking it to the impacts of the intervention.

However, different challenges arise in estimating benefits because of the difficulty in accurately estimating shadow prices for educational outcomes. Even if the outcomes of an educational intervention are identified (e.g., from a random assignment trial), the economic value of these outcomes may not be precisely bounded. For example, the effect size gain in achievement from the Tennessee STAR experiment to reduce class size may be precisely identified (see, for example, Finn et al., 2005), but there is much less certainty over the economic value of reducing class size.

Broadly, educational outcomes can be grouped into those related to attainment (e.g., years of schooling or college), achievement or cognitive test scores, and behavioral change (e.g., reduced delinquency or increased socioemotional knowledge). Presently, there are well-established protocols for estimating shadow prices for the first of these groups. That is, there is reasonable evidence on the value in terms of higher incomes from getting students to graduate from high school and from each additional year of education (at least the income effects). However, there is much less information for the other two groupings. Notably, despite substantial policy attention to raising test scores and reducing achievement gaps (e.g., No Child Left Behind), there is no consensus on the economic value of achievement gains. In order to identify these gains, it is necessary to separate out confounding factors—students with higher achievement tend to have higher attainment—and to specify how initial cognitive gains influence subsequent cognitive skills (for a discussion, see Belfield & Levin, 2009). For delinquency, there are shadow prices for specific crimes (e.g., violent assault) and there are some estimates of the cost consequences of being an at-risk youth (Cohen & Piquero, 2009). However, these prices are averages across all youth and it may be that educational investments are influential only for intermittent offenders and not chronic offenders. Few shadow prices for socio-emotional skills have been calculated.

As with CEA, the response to these challenges is to perform extensive sensitivity testing. Alternative values for key parameters and best-case/worst-case sensitivity tests may be performed. Indeed, worst-case sensitivity analysis may be sufficient to substantiate the results of a CBA: if the benefits still exceed the costs with a worst-case scenario, then it is very likely that the program is allocatively efficient. Similarly, if a CBA includes only a few of the benefits of an education program, and yet these nonetheless exceed the costs, then this too generates a strong conclusion about efficiency.

In addition to these sensitivity tests, for CBA Monte Carlo simulations may be applied. These simulations use information on the distributions of costs, impacts, and shadow prices. By repeatedly drawing new estimates from the distributions of values for each parameter, the NPV, B/C ratio, or IRR can be re-derived. (Typically, 500 or 1,000 draws are taken.) Monte Carlo simulations are helpful when the impacts of some interventions are not statistically significant from the comparison group. The wide variation in these impacts will be reflected in the variance in the economic metrics. These simulations provide information not only on the precision of the net present value but also on the probability that this NPV is negative (i.e., that the costs exceed the benefits).

### *Examples and Findings*

Evidence on the costs and benefits of educational programs is more readily available than evidence on their cost-effectiveness (Hummel-Rossi & Ashdown, 2002). Much of the available CBAs are for early childhood programs, although there are now more examples of CBA at the high school and college levels.

The Chicago Child-Parent Center program is one example of a preschool program that has been subjected to a full CBA. (Others include the High Scope/Perry Preschool program and the Abecedarian Child Development Program; see Heckman et al., 2010, and Barnett & Masse, 2007.) The preschool program is intended to help promote children's academic and social skills with the intention that this will enhance the children's development, leading to improved economic well-being in the short, medium, and long run. Specifically, investments in preschool are expected to: help families with child care responsibilities; improve resource use in school (e.g., by reducing the need for special educational services); and, by enhancing the child's human and social capital, lead to higher incomes and lower delinquency. These consequences can all be expressed in monetary terms.

The costs and benefits of the Chicago Child-Parent Center program are shown in Table 9.3 (from Reynolds et al., 2011). These are based on follow-up of Center preschool participants at age 26 compared to follow-up of a quasi-experimentally matched control group. Three perspectives are reported: those of the child participant, the general public (those other than the participants), and society (the sum of the child and general public impacts). The preschool program costs $8,512 per student, which is paid by the general public and not the child's family. The benefits of preschool are substantial and diverse across the different perspectives. For the individual, the benefits include: childcare savings for the family, improved child welfare (lower abuse/neglect), and significantly higher earnings over the lifetime as the child becomes an adult. In total, the present value

Table 9.3 Chicago Child-Parent Center Program (Pre-School): Cost-Benefit Analysis

| | Perspective | | |
|---|---|---|---|
| | Participants | General Public | Society |
| Program Costs [C] | – | $8,512 | $8,512 |
| Program Benefits: | | | |
| Child care | $4,387 | – | $4,387 |
| Child abuse/neglect | $4,240 | $3,090 | $7,330 |
| Education | $(98) | $6,001 | $5,903 |
| Earnings | $22,445 | $6,399 | $28,844 |
| Criminal behavior | – | $42,462 | $42,462 |
| Health | – | $3,294 | $3,294 |
| Total Benefits [B] | $30,974 | $61,246 | $92,220 |
| Net Present Value (Benefits-Cost) | $30,974 | $52,734 | $83,708 |
| Benefit-Cost Ratio | – | 7.2 | 10.8 |
| IRR | | | 18% |

*Source:* Reynolds et al. (2011, Table 4, Appendix C). 2007 dollars. Discount rate 3 percent. Education benefits are savings in grade retention and special education net of additional spending on college. Health benefits are for reduced depression and substance abuse.

benefits to the child are $30,974, and these are also the net benefits because the program is at zero cost to their families. The general public perspective shows there are many other benefits from preschool. These include: improved child welfare, savings to education budgets, increased earnings, and reductions in government spending on criminal behavior and health. The total present value benefits to the general public are $61,246. The net present value is therefore $52,734 ($61,246 minus $8,512) and the benefit-cost ratio is 7.2:1. From the social perspective, the benefits measured across all these domains are substantial at $92,220; the net present value is $83,708, and the benefit-cost ratio is 10.8:1. Finally, from the social perspective, the internal rate of return is 18 percent. Using any of these metrics, the Chicago Child-Parent Center preschool program easily passes a cost-benefit test. Also, the more intensive is the program, the greater are the net benefits. Programs of one year or less generated net benefit amounts of $68,347 versus $83,708 for programs lasting more than one year (Reynolds et al., 2011, Table 5).

CBAs of early childhood interventions have been summarized by Karoly (2012), and the results for social CBAs are presented in Table 9.4. These results show that for each program the benefits substantially exceed the costs, in some cases by a factor of more than 10. A general finding of the CBA literature is that early interventions yield high net present values and IRRs.

Table 9.4 Social Benefit-Cost Ratios: Early Childhood Interventions

| Program | Benefit-Cost Ratio |
|---|---|
| Nurse-Family Partnership | 2.88 |
| Abecedarian Child Intervention | 2.49 |
| HIPPY USA | 1.80 |
| Chicago CPC | 10.83 |
| Perry Preschool | 7.1–12.2 |

*Source:* Adapted from Karoly (2012, Table 2). Discount rate 3 percent.

## CONCLUSIONS

Ultimately, the value of CEA and CBA rests on the extent to which it improves policy decision making. Purposefully, cost-effectiveness analysis and cost-benefit analysis are intended to help decision makers allocate resources to educational programs as efficiently as possible. The conditions under which CEA is appropriate are strict and the data requirements for CBA are quite challenging. Nevertheless, even an analysis of costs can be helpful in clarifying how an intervention is actually delivered on the ground and by showing how much it costs when all resources are accounted for. It is possible that some reforms fail to be effective because they do not involve much change (e.g., curricular reforms). It is also possible that some reforms cannot be implemented because they are just too expensive (e.g., reducing class size or whole school reform) or rely on volunteer labor that is not available in large scale. Costs analysis is informative on these possibilities. Even where CEA and CBA are imprecise, they may still provide information on programs that clearly do not pass an efficiency test or on ones (such as early education) that clearly do.

For CEA and CBA to be more widely useful, policymakers need to be able to draw general conclusions from the available evidence. That is, they will want to compare the returns from one investment to the returns from another so as to make the most efficient choice. Therefore, it is necessary for the evidence to be expressed in metrics that are comparable. Indeed, this is a significant advantage of CBA: because the evidence is expressed in monetary units, then potentially any intervention can be compared against another. Yet, if analyses are not harmonized there is a danger that policymakers will draw false inferences about efficiency. Of course, as there are so few CEAs and CBAs, policymakers might also be concerned about the many interventions whose efficiency has not yet been investigated.

There are many dimensions on which CEA and CBA studies need to be harmonized (Karoly, 2012). These dimensions include the range of outcomes considered—a study will have a higher benefit-cost ratio simply if it counts more benefits (e.g., if both labor market earnings and health status gains from high school graduation are counted instead of just the former). Karoly (2012, Table 6) catalogs the many differences in outcomes evaluated in CBAs of early childhood interventions. Other important dimensions are the time horizon of the evaluation—looking over a longer time horizon will likely yield a higher estimate of benefits—and the discount rate for deriving present values. On the latter, Karoly (2012) proposes a 3 percent discount rate should be the standard, with sensitivity analysis performed using discount rates of 0, 5, 7, and 10 percent. More generally, most CEAs and CBAs do not include any distributional considerations. An intervention to help struggling students in college is evaluated using the same tools and metrics as an intervention to help struggling readers in kindergarten. The units (dollar costs, impacts, and dollar benefits) do not differ across these two interventions. Although this equivalence is explicit in CEA/CBA, it is likely that decision makers will find it hard to adjudicate between programs serving very different student groups.

Fundamentally, it would be useful for more CEAs and CBAs to be performed within the education sector. The large aggregate resource commitment—both in money and the time of K–12 and college students—makes it important that decisions about educational investments are made efficiently. CEA/CBA, as well as basic analysis of costs, can contribute toward improving the quality of decision making across the sector.

## REFERENCES


Barnett, W. S., & Masse, L. N. (2007). Early childhood program design and economic returns: Comparative benefit-cost analysis of the Abecedarian program and policy implications. *Economics of Education Review, 26*, 113–125.

Belfield, C. R., & Levin, H. M. (2007). *The price we pay: The economic and social consequences of inadequate education.* Washington, DC: Brookings Institution Press.

Belfield, C. R., & Levin, H. M. (2009). Some economic consequences of improving mathematics performance. Monograph, Center for Benefit Cost Studies in Education, Teachers College. Available at www.cbsce.org.

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

Bowden, B. (2013). The cost-effectiveness of Talent Search. PhD thesis, Teachers College, Columbia University.

Clune, W. (2002). Methodological strength and policy usefulness of cost-effectiveness research. In H. M. Levin & P. J. McEwan (Eds.), *Cost-effectiveness and Educational Policy* (pp. 98–121). Larchmont, NJ: Eye on Education.

Cohen, M. A., & Piquero, A. R. (2009). New evidence on the monetary value of saving a high-risk youth. *Journal of Quantitative Criminology, 25*, 25–49.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology, 97*, 214–223.

Gramlich, E. M. (1981). *Benefit-cost analysis of government programs.* Englewood Cliffs, NJ: Prentice Hall.

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

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High/Scope Perry Preschool Program. *Journal of Public Economics, 94*, 114–128.

Hollands, F., Bowden, A. B., Belfield, C., Levin, H. M., Cheng, H., Shand, R., Pan, Y., & Hanisch, B. (in press). Cost-effectiveness analysis of interventions that improve high school completion. *Educational Evaluation and Policy Analysis.*

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.

Karoly, L. A. (2012). Toward standardization of benefit-cost analysis of early childhood interventions. *Journal of Benefit-Cost Analysis, 3*, 1.

Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. In R. Light (Ed.), *Evaluations that surprise.* San Francisco: Jossey-Bass.

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

Levin, H. M., & McEwan, P. J. (Eds.) (2002). *Cost-effectiveness and educational policy.* Larchmont, NJ: Eye on Education.

Levin, H. M., Glass, G. V., & Meister, G. R. (1987). Cost-effectiveness of computer-assisted instruction. *Evaluation Review, 11*, 50–72.

Moore, M., Boardman, A., Vining, A., Weimer, D., & Greenberg, D. (2004). Just give me a number! Practical values for the social discount rate. *Journal of Policy Analysis and Management, 23*, 789–812.

Reynolds, A. J., Temple, J. A., White, B.A.B., Ou, S.-R., & Robertson, D. L. (2011). Age 26 cost-benefit analysis of the Child-Parent Center Early Education Program. *Child Development, 82*(1), 379–404.

Tsang, M. C. (1997). Cost analysis for improved policy-making in education. *Educational Evaluation and Policy Analysis, 19*(4), 18–24.