

**GUIDING THE DEVELOPMENT AND USE OF COST-EFFECTIVENESS
ANALYSIS IN EDUCATION**

Henry M. Levin

Center for Benefit-Cost Studies of Education

Teachers College, Columbia University

Clive Belfield*

Queens College, City University of New York

December 2013

*Corresponding author: clive.belfield@gmail.com.

The co-authors wish to express their appreciation for the constructive assistance of their colleagues Brooks Bowden and Fiona Hollands.

The research reported here was supported by the Institute for Education Sciences, U.S. Department of Education, through Award Number R305U120001 to Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Abstract

Cost-Effectiveness Analysis is rarely used in education. When it is used it often fails to meet methodological standards, especially with regard to cost measurement.

Although there are occasional criticisms of these failings, we believe that it is useful to provide a listing of the more common concerns and how they might be addressed.

Based upon both previous projects on the subject and a recent project which attempted to construct cost-effectiveness comparisons from the What Works Clearinghouse of the Institute of Educational Sciences of the U.S. Department of Education, we have identified a set of recommendations for practice that should have high priority if we are to construct valid cost-effectiveness comparisons for policy uses when choosing among alternative educational interventions.

1. Introduction

Education evaluation specialists, government agencies, and decision-makers who are responsible for addressing resource allocation receive little training or guidance in how research evidence on cost-effectiveness can be used to enhance practice. Although they may be aware of cost-effectiveness analysis, these agents have little experience in using it or drawing on its evidence base. Moreover, this neglect means that these agents are not adequately prepared to scrutinize or interpret cost-effectiveness analysis results. In turn, researchers performing cost-effectiveness analysis do not receive any feedback in how to improve *their* practice and so make their analysis more useful for policy. The result is little use of cost-effectiveness analysis and limited dialogue between the producers and consumers of such analysis (researchers and policymakers) that might improve such analysis.

Given this context, we review the development of cost-effectiveness evaluation and use in education and address the issue of how it can be made more useful for increasing efficiency in resource use in education. The review draws upon more than four decades of applying cost-effectiveness analysis to the assessment of educational alternatives. We also include lessons learned from recent cost-effectiveness evaluations of educational interventions based on our analysis of evidence from the What Works Clearinghouse of the Institute of Education Sciences of the U.S. Department of Education (Hollands et al., 2014).

Given the substantial cost of the education sector and its growth, one might expect considerable scrutiny on how societal resources can be used more efficiently. Cost-effectiveness analysis (CEA) – an approach that identifies which strategies will maximize outcomes for any given cost or produce a given outcome for the lowest cost – is the most versatile tool for this task. For almost any educational objective, such as

improvement of academic achievement in a specific subject, there are usually many alternatives – better personnel selection and training, new instructional strategies, use of technologies, class size reduction, and curriculum innovations. The alternatives that show greater productivity relative to their costs, i.e. that are more cost-effective and more efficient in the use of social resources, should be preferred for adoption and implemented more intensively. The consequence of using CEA appropriately will be that decision-makers select the most effective interventions they can afford for a given budget, or the least costly interventions that result in a desired outcome. As logical as this strategy may appear, it has only rarely been considered or applied in educational decisions.

Since the 1960's the use of cost-effectiveness analysis has expanded in many government sectors including the military, health, transportation, and criminal justice, but this pattern of acceptance has not been followed in education. Moreover, even where CEA has been applied in education, the attempts have been far from rigorous (Ross, Barkaoui, & Scott 2007). Most evaluations make only a 'rhetorical' mention of costs and cost-effectiveness; very few make any systematic attempt to quantify costs let alone to link them with measures of effectiveness (Clune, 2001). There is almost no direct evidence to support – on efficiency grounds – the many educational interventions or reforms now being proposed for implementation. It may be that the most effective interventions are the most efficient, but there is no direct evidence that this is the case.

Thus, the need for CEA is pressing and its application needs improvement. Here, we offer guidance on how to address both issues – expanding the use of CEA whilst at the same time improving its quality. We review the genesis of CEA and describe the basic principles that were first specified over four decades ago. Next we review methodological developments. We use recent cost-effectiveness analyses for exposition

and draw on literature from the health sciences, where CEA is applied much more frequently; we also describe a Cost Tool Kit that will improve the application of CEA. Finally, we put forward a series of recommendations to improve the practice and extend the influence of CEA.

2. Background

Although the human capital revolution in economic thought and its focus on education go back to the 1960's (Schultz 1961; Becker 1964), this revolution mainly emphasized the economic productivity of human capital investment rather than efficiency in the production of education. Even in the early sixties, public spending on education represented a substantial portion of GNP (at almost 4 percent, rising to about 5.5% today), not including the substantial resources devoted to private education or training by industry or the military. Education now represents a serious economic commitment, amounting to about \$1 trillion today for the public share of investment alone. This investment has increased over time because of both rising enrollments and rising costs per student. Yet, inquiry into which educational practices are most efficient has been extremely limited.

Why resource analysis in education has not kept pace with other areas of social production is partially a mystery (Levin 2001). Governments have expressed deep concern about financing education in an age of austerity and rising costs. In addition to the federal budget deficit, 30 states faced budgetary deficits in 2013 averaging almost 10 percent of their annual budgets or \$55 billion. Just a 5 percent "dividend" from more efficient educational strategies and operations would amount to about \$50 billion in annual savings. Such a gain in efficiency is highly feasible in an area that has been unexplored for productivity. But there are few calls for serious studies of cost-

Cost-Effectiveness Analysis: Methodological Concerns

effectiveness in education by state legislatures or other government bodies, and educational evaluators, policy analysts, and administrators are rarely familiar with the tools of cost-effectiveness analysis.

Cost-effectiveness analysis is distinct from cost-benefit analysis, which has been adopted somewhat more frequently (Levin & McEwan 2001: Chap. 1). Cost-benefit analysis (CBA) refers to a direct comparison of the monetary cost and benefit of a particular intervention where cost and benefit consequences can both be measured in monetary terms. Only alternatives whose calculated benefits exceed their costs are considered, and the priority is given to those with the largest ratio of benefits to costs or that generate the largest net benefits (benefits minus costs). Most benefit-cost studies in education attempt to address the benefits of educational interventions on monetary outcomes in labor markets or other forms of social productivity (Belfield & Levin 2007).

In contrast, cost-effectiveness analysis (CEA) in education is used to compare alternative interventions with similar educational goals such as gains in reading or math achievement or completion of courses or other educational outcomes. Measures of outcomes among alternatives must be similar for making comparisons. When costs are compared to outcomes, priority of adoption is given to those interventions that show the highest effectiveness relative to cost. The cost concept and measurement underlying CBA and CEA studies is identical, but the measures of outcome differ. Most educational outcomes are stated in terms of educational effectiveness such as learning results, rather than their monetary benefits such as earnings, and so are more amenable to CEA than CBA. Also, CEA can be applied to the evaluation of educational interventions designed to improve both cognitive and non-cognitive outcomes such as test scores, attitudes, and behaviors; many of these cannot be easily monetized.

Therefore, both in terms of directness and wider applicability, CEA is more appropriate than CBA for many research questions in education.

The early development of cost-effectiveness was focused on the comparison of performance and costs of weapons systems by military decision-makers. Consideration of productivity in educational resource use was stimulated in 1966 with the publication of the Coleman Report that focused on the effectiveness of schools and families in producing educational achievement (Coleman et al. 1966). The Coleman report also provided a rich and important new dataset for further analysis. The first article that addressed CEA specifically in education enlisted the Coleman data to compare the cost-effectiveness of hiring teachers with more experience versus hiring teachers with higher test scores (Levin 1970). Using earnings functions for the costs of teacher attributes in combination with production function results for the same teacher variables from Hanushek (1968), the study found that selecting teachers with higher verbal scores was 5-10 times more cost-effective than selecting teachers with more experience. This finding has turned out to be consistent with recent findings on teacher effectiveness, although policies regarding teacher selection and teacher salaries still usually favor experience over verbal ability.

At that time there was virtually no presence of cost-effectiveness analysis in educational research and policy circles. But the field of social program evaluation had begun to emerge under the leadership of Donald Campbell, Thomas Cook, Lee Cronbach, and others, and a new society had formed around the subject with the aim of defining the field through a 1975 *Handbook of Evaluation Research*, edited by M. Guttentag and E. Struening (hereafter referred to as the Handbook). Given the development of a new field, there was also a willingness to embrace new subjects such as cost-effectiveness in evaluation as a subject for the new Handbook. Henry Levin

authored a chapter in the Handbook on applications of cost-effectiveness to social programs, but not only in education because the Handbook was designed to be relevant to many fields including crime, health, and public assistance. In an exploration of CEA in other fields, Levin found that incorporation of cost analysis in evaluation research was also rare and of low quality. Even if costs were referred to, there was no detail on the costing methods. Many papers stated the costs without any reference to their source or referred only vaguely to obtaining cost information from a business or accounting office. In contrast, the literature was far clearer on the details of evaluation methods used to derive effectiveness.

3. The Ingredients Method

3.1 Opportunity Cost

In preparing the 1975 Handbook article, Levin perceived that the first order of business was to develop a systematic method by which costs could be measured and compared among alternatives, parallel to the attempt to measure their comparative effectiveness. A defensible method would require a definition of costs, a valid approach to their measurement, and a transparent method of interpretation. Drawing on the economic definition of *opportunity cost*, the value of a resource in its best alternative use, he anchored the conceptual framework to the large body of economic analysis of cost. Opportunity cost is the value of what is sacrificed by using a specific resource in one way rather than in its best alternative use. This is a long-standing theoretical definition in economics, and it is at the heart of micro-economic theory.

Opportunity cost is most frequently measured by the market price, the price that equates supply and demand in the competitive marketplace. But many markets do not have competitive market conditions (many buyers and sellers, accurate and readily

available information on alternatives, transparency of pricing, and ease of entry or exit). In these cases, the observed price is distorted relative to opportunity cost, so a *shadow price* is used which approximates to the value or market-clearing price if the market were competitive. In other cases, a market does not exist (for example, a unique source of talent); so another approach must be used to ascertain value. The advantage of using opportunity cost is that it provides a strong theoretical background and consistency across applications and a basis for societal assessment of cost including the ability to consider externalities (cost or benefits to those not directly engaged in the transactions being valued) which are not accounted for by markets.

A straightforward way of evaluating the cost of an intervention on the basis of opportunity cost is to employ an “ingredients” approach to identify and specify the resource components required for implementation. The ingredients method was initially presented by Levin in the 1975 Handbook, with subsequent refinements and two book-length treatments providing a specific focus on education (Levin 1983; Levin & McEwan 2001). This method has been largely adopted as the standard cost methodology for educational CEA as evidenced by the more than 1,000 published citations to the original Handbook article and the two editions of the methods textbooks (Levin, 1983; Levin & McEwan, 2001). The method has also been adopted by the Poverty Action Laboratory at MIT which is renowned for its experimental interventions to reduce poverty (Dhaliwal et al. 2012). A brief description of the method is provided below, but more expansive discussions, explanations, and methodological considerations can be found in Levin & McEwan (2001 & 2002).

3.2 Estimating Costs with the Ingredients Method

There is a natural tendency of decision-makers or evaluators to assume that cost information on potential interventions is readily available from budgets or business

personnel where the interventions have been developed or applied. Unfortunately, cost information from these sources will almost invariably be incomplete and potentially wrong because the principles on which budgets are based were not designed to give accurate calculations of economic cost. Conventional accounting practices in the educational sector (and the public sector more generally) were designed for purposes other than accounting for costs of instructional or other interventions (Levin & McEwan, 2001: 45-46). Budgetary information is unlikely to account for all of the ingredients of an intervention and their true costs. In contrast, cost estimation using the ingredients method builds on evaluators' understanding of the details of the treatments or interventions and uses that knowledge to construct direct estimates of costs.

The ingredients method of cost estimation involves three main steps to ascertain accurate and consistent measures of costs: identifying and specifying the ingredients required to obtain the evaluation results, identifying their costs, and calculating total program costs and average costs per participant. A fourth step may be used to determine the distribution of the cost burden among multiple constituencies. The underlying rationale and details on procedures for all of these steps are presented in Levin and McEwan (2001). It should be emphasized that cost-effectiveness comparisons can only be valid when the alternative interventions being considered are comparable along a number of dimensions, including the measures of effectiveness employed to determine impact. Ultimately, the costs per unit or level of effectiveness must be evaluated across alternatives, and this requires uniformity in both methods of cost and effectiveness measurement.

In order to conduct cost-effectiveness analysis, it is best to integrate collection of cost data simultaneously with the evaluation of effectiveness. In practice, however, almost all cost-effectiveness comparisons in education have involved the construction

of cost estimates well after the effectiveness evaluations have been completed and disseminated. This may be due to policymakers only becoming concerned about costs after being convinced that a program may be worth adopting or scaling up on the basis of effectiveness. One early exception was Levin and Woo's (1981) evaluation of computer-assisted instruction (CAI) in the Los Angeles Unified School District. This evaluation represented the first, large-scale attempt to undertake an experiment using random assignment of students to different amounts of CAI and in different subjects. Levin and Woo were able to engage in direct observations and interactions with program personnel and evaluators, allowing collection of data on resource requirements during the course of the experiment. This information was used later to compare the cost-effectiveness of CAI with that of peer tutoring, smaller classes, and longer school days in Levin, Glass, and Meister (1987). But, the fact that a cost-analysis or cost-effectiveness analysis was rarely integrated into the initial evaluation of effectiveness has meant that most of the work on CEA in education has involved the estimation and integration of costs well after the effectiveness evaluation was completed. The method and underlying concepts used for the ingredients approach to cost evaluation are similar whether done contemporaneously or after the effectiveness evaluation. However, accuracy and parsimony in data collection can be improved if cost and effectiveness evaluations are conducted simultaneously.

The first step of a cost analysis involves the identification and specification of the ingredients or resources that are required and used to implement the program being evaluated. For cost-effectiveness analysis, these must be the resources used to obtain the level of impact observed in the effectiveness evaluation. Such resources are identified in detail, not only in quantitative terms, but also along their qualitative dimensions. For example, the precise types and amounts of personnel are specified according to their

qualifications, functions, and time commitments. A similar exercise is carried out for facilities, equipment, and other program inputs as well as required client resources. This information is obtained from three sources: (1) descriptive reports; (2) observations of the intervention in practice if it is still in operation; and (3) interviews with program personnel. It is important to focus on how specific resources are used to implement the “theory of action” addressed by the intervention. Sources are triangulated to test for consistency and discrepancies resolved by further investigation. All ingredients are identified and specified regardless of how they are financed, i.e., it is necessary to identify all resources that contribute to the total cost, regardless of who provided or paid for them.

Second, once the ingredients are identified, the next step entails identifying their opportunity costs. As much as possible, market prices should be used to establish the value of each ingredient. For personnel, information regarding time commitments, qualifications, and performance requirements can be used to find appropriate market prices in the form of salaries and benefits. Personnel usually constitute at least three quarters of traditional educational interventions, and a high proportion of even technology-intensive ones, so careful attention should be focused on this category. This “costing” procedure is followed for each and every ingredient.

In some cases the ingredients may not be obtained through market transactions, for example, space in a building that is owned by the sponsoring entity, or in-kind resources such as volunteers. In these cases, shadow prices must be used, i.e., the estimated value of the resources determined through an alternative procedure such as contingent valuation (Diamond & Hausman, 1994) which is a process for ascertaining what would be paid in a competitive market. Personnel costs may be obtained from expenditure budgets if derived through competitive market transactions, but in most

cases it is advisable to use personnel prices that are based on an average for a regional or national market. For each ingredient category there are accepted methods for determining their costs (Levin & McEwan, 2001: Chap. 3). In the case of educational facilities, because rental rates are not commonly available, annual costs may be derived by identifying costs of new school building construction and amortizing these costs over the expected usable lifetime of the building.

Once the costs of the individual ingredients are obtained, the third step in cost analysis is to aggregate these costs to determine the total program cost, annualized or evaluated over a given period. Average cost per participant and marginal cost per participant (cost of adding each further participant beyond a base number) should also be calculated. In education, the most common cost measure used is the average cost per participant, which can be combined with the average effect per participant to allow a comparison of cost-effectiveness among program alternatives. Cost-effectiveness ratios are usually calculated by dividing the cost per participant by the average effectiveness observed to indicate the cost per unit increase in effectiveness. Decision-makers should give preference to those alternatives with the lowest cost relative to effectiveness, although other important criteria that are not included in the cost-effectiveness computations might be taken into account. For example, an intervention that improves additional outcomes such as equity might be given extra consideration when choosing among program alternatives.

In a final step of the cost analysis, the distribution of the cost burden among different agencies (e.g., state educational agency, local school district, school, parents) may be addressed, including in-kind contributions of ingredients and such cash subsidies as user charges, government subsidies, and philanthropic contributions. These adjustments enable the calculation of net costs per “funding” constituency for each

intervention. The reason to conduct this final step is that some interventions benefit from assistance from particular government or charitable agencies. These subsidies may change the net cost pattern among alternative programs for the adopting organization (usually a school or district). If the adopter is able to get some of the costs of an intervention subsidized through government or private agencies, or program participants, the cost-effectiveness ratios of alternative programs will be modified relative to the unsubsidized cost-effectiveness ratios. It is best to reveal the unsubsidized costs as the true costs of an intervention for society, even if the subsidized cost is used by any decision-maker who benefits from the subsidy.

3.3 Helping Decision-Makers

The fundamental goal of CEA is to help decision-makers improve the allocation of resources. Such improvement is especially necessary when conventional incentives (such as the profit motive) are absent. Thus, a basic test of when to apply CEA is to ask: would the information contained in a CEA affect any decision? We believe that CEA can potentially influence many decisions, but there may be circumstances when a policy commitment has been made in such a way that the cost of the commitment is immaterial.

In order to affect decisions, CEA must satisfy two criteria. First, it should be based on a credible method that has been applied rigorously. We discuss application of CEA in detail below. Second, CEA should be presented to decision-makers in a way that is compelling and useful. That is, the results of the CEA must be expressed in a way that can influence decisions.

Explicit attention should be paid to providing CEA that is linked to the needs of decision-makers. Conventionally, CEAs have been reported using a social perspective, i.e., taking into account all of the resources required to implement the intervention. This

convention follows a long tradition of welfare economics and the adoption of the social perspective for cost-benefit analyses. However, for some decision-makers this social perspective may not be the most relevant. For example, schools may only be interested in the costs that they must pay for, discounting any burden on other parties. Some interventions, particularly those involving peer-tutoring, may look more cost-effective simply on the grounds that the 'labor' (by the peers) is cheap. CEAs should be performed in ways that provide the most use for decision-makers, although a social perspective is valuable for informing the public on the full costs and on optimal scale.

Analysts must convey the specific meaning of the CE ratios they calculate. Typically, although this is rarely made clear, these ratios are *incremental*. For example, an after-school program may yield an increase in test scores, and the CEA may report the gain in test scores per dollar spent on the after-school program. But this gain is an incremental one from extra spending because all students are receiving regular classroom instruction as well. The CE ratio should be interpreted as the gain in test scores from incremental spending. The ratio is not especially informative about what would be the most efficient investment of the entire resource base available for these students (classroom plus after-school). The ratio is also limited in relevance for after-school programs that are either much larger in scale or much smaller in scale. Costs are likely to be very sensitive to the scale of implementation. Rather, the ratio is most useful for the decision-maker who has a discretionary amount of funding that might be allocated to after-school programs at a scale of operation that is similar to that for which there is evidence. Analysts must therefore be very clear about the policy relevance of their research findings.

Decision-makers can also aid the development of CEA. A critical element of CEA is that the outcomes be expressed in a one-dimensional format, i.e. that all

outcomes are collapsed into a single number. But, this expression should be related to the decision to be made and the ultimate goal of the intervention. If decision-makers precisely specify what improvements are being sought, this greatly helps in creating a one-dimensional outcome. Simply stating ‘reading improvement’ or ‘increased graduation’ as an outcome is, as we show below, insufficiently precise. Therefore, an important contribution of CEA is to encourage or press decision-makers to clarify the goals of any given intervention. In many studies we have found that decision-makers are often surprisingly vague about the goals of reforms.

4. Methodological Issues in Conducting CEA

4.1 Applying CEA to Effectiveness Studies

In the U.S. the most prominent attempt to assess the quality of evidence from educational evaluations is the What Works Clearinghouse (WWC) of the Institute of Educational Sciences (IES) of the U.S. Department of Education. The WWC reviews research on different educational interventions to evaluate the quality of research and validity of the findings with the goal “to provide educators with the information to make *evidence-based decisions*.” The WWC provides systematic and expert reviews of evaluation research on effectiveness, but includes very little information about costs. Critically, evidence-based decisions will not result in the most efficient use of available resources if cost information is absent, even when the effectiveness results are certified as valid.

To address this deficiency, the IES recently sponsored a research study to link best evidence on effectiveness of educational programs to costs, with a focus on two areas: dropout prevention and early childhood literacy. The premises underlying this study were that the WWC had certified the quality of the evidence underlying the

various interventions that it listed in each area, and that the ingredients method of costing would be applied to determine the costs and cost-effectiveness of the programs deemed to be effective. Cost-effectiveness ratios calculated by combining the data on costs and effectiveness would be compared among alternative programs in each of the two domains. The goal of the study was primarily methodological: to see if the ingredients method could be applied across a range of interventions and to demonstrate the challenges inherent in CEA. Full details of the cost-effectiveness calculations for dropout prevention and early childhood literacy are given in Hollands et al. (2013, 2014).

The study raised a number of methodological concerns related to the ability to perform CEA. Perhaps surprisingly, an immediate and significant concern relates to the measurement of effectiveness. WWC aspires to provide an evidence-based source of information on alternative educational interventions that can inform educational decisions. But the studies that it evaluates are designed and implemented independently, so there is no requirement to provide comparable definitions and measurements of ostensibly equivalent outcomes. For example, one might expect that the effectiveness of dropout prevention programs would best be measured by high school completion. But the 13 WWC interventions in the Dropout Prevention category that are listed as having positive or potentially positive effects are arrayed under three different outcomes: staying in school; progressing in school; and completing school. Even within these categories the specific measurement of outcomes is different. For example, assessments of staying in school and progressing in school are conducted at different grade levels using different criteria among studies. Completing school encompasses finishing high school, graduating from high school, or the receipt of a GED certificate. This last outcome, the so-called high school equivalency examination, is not equivalent to

graduation requirements for high school, and its passing grade differs among states. Even high school completion may mean very different things. A report of the National Research Council found virtually dozens of different definitions and measures of what appears to be high school completion or high school graduation (Hauser & Anderson Koenig, 2011). One potential solution is to place values or “utility-weights” on different outcomes so that they can be aggregated into a single measure of effectiveness (Edwards, 1992).

A similar but even more complex challenge was encountered in finding comparable programs for a CEA based on evidence from the WWC studies on early literacy. The WWC classifies early literacy outcomes under four different “domains,” alphabeticity, fluency, comprehension, and general reading achievement, each of which may consist of multiple sub-categories or “constructs.” While we identified 32 programs that had a positive or potentially positive impact on early literacy, only 19 did so for the alphabeticity domain, and far fewer did so for any construct within this domain, such as phonics or phonological awareness. A first task was to establish at what level (domain or construct) outcomes are really comparable.

A second concern is that the results of different studies should only be compared for similar demographic groupings given the wide differences in educational attainment by race, gender, language minorities, and socio-economic status (Rumberger, 2011). Although comparisons of effects may meet internal validity claims by comparing control groups with treatment groups for similar populations, any result is not comparable with studies done for other populations. For example, in the area of dropout prevention, programs serving promising students still in school should not be compared to programs serving teenage, dropout mothers. In the area of early literacy, effectiveness

of reading programs that serve struggling readers should not be compared with effectiveness of reading programs designed to serve average readers.

Integrating costs and effectiveness data is complicated by the use of effect sizes in evaluations. The use of effect sizes to show magnitudes of effectiveness may give an impression of comparability for the “evidence-based” decision-maker. But the underlying measures of what constitutes effectiveness and for who may differ considerably. So, although decision-makers may be tempted to compare directly the effect sizes of different programs to inform their decisions, CEA requires that the alternative programs being compared share common goals, use similar measures and serve similar populations. In addition, an effect size gain will depend on the variance in the sample, and so studies that rely on samples with different underlying variances will have effect sizes that are not strictly comparable.

With respect to costs, a significant challenge is that there is limited information on the resources allocated to the control group. In many studies, the control group might be assumed to receive ‘business-as-usual’ education. If this assumption is correct, the resources for business-as-usual are still not easily determined: control groups are often spread across schools or classrooms (in contrast to the usually more concentrated locations of the treatment group, e.g., in an after-school program). More importantly, the interpretation of the CEA changes substantially if costs are only attributed to the resources required above and beyond business-as-usual. The CE ratio is now an incremental cost-effectiveness ratio (ICER), where extra effectiveness is being generated from expenditure on the sum of business-as-usual and the cost of the intervention. It is also possible that the assumption that the control group benefits from nothing more than business-as-usual is incorrect: the control group members may receive alternative resources in place of those received by the treatment group. For

example, in a study of after-school programs, students not assigned to the program being evaluated might be enrolled in other after-school activities. Rarely is cost data available for business-as-usual services or for alternative activities in which the control group participates.

Another serious challenge arises from the time lag between the initial evaluation of effects and the delayed effort to determine costs. Typically this was three years or more, but some of the delays were as long as one or two decades. Even the most conscientious efforts to identify the implementation process for each alternative program and the program ingredients are hampered by such time lags. For example, archival materials describing the intervention are rarely published and represent fugitive documents that are not easily found after a year or so. Even if staff participants can be found, they are likely to have difficulties remembering details accurately, and often the intervention cannot be observed itself because it no longer exists or has been transformed over time. In some cases there is no extant version at all and no availability of personnel from many years ago who can be accessed. As a result, the lag in timing between assessment of effectiveness and the retrospective assessment of costs represents a serious obstacle to obtaining reasonably complete and comparable cost estimates.

Finally, CEA introduces a second source of variance into estimation of program impact. Research suggests that what appears to be the same educational intervention can be implemented in dramatically different ways at different sites (Durlak & DuPre, 2008). Yet, not only will effectiveness vary across participants and sites in a study, but so will cost. As costs are unlikely to follow a normal distribution, it is not clear what statistical tests could be applied to identify sampling variance. Among dropout prevention programs, this variance was manifest in cost estimates across sites.

Cost-Effectiveness Analysis: Methodological Concerns

Compounded by variance in effectiveness, cost-effectiveness ratios varied substantially. For example, the cost-effectiveness ratios varied by a factor of more than ten across five sites implementing the Talent Search program. In other work, we found that even a highly formulaic and prescriptive educational intervention licensed by a commercial publisher used vastly different amounts of resources at each of several sites (Levin, Catlin, & Elson, 2007).

In summary, even when rigorous studies of effectiveness are available, it is challenging to provide comparable cost-effectiveness estimates. Only some of the challenges reflect methodological difficulties with respect to assessing costs. The primary challenge – the incomparability of program outcomes, measures of effectiveness, and populations served– relates to effectiveness. A second challenge arising from the practice of delaying cost analysis until after the effectiveness results are available can be more easily addressed by evaluating costs and effectiveness simultaneously. It is important to note that the problems that we identified are not intrinsic to the efforts of WWC, but to the fact that evaluators undertake each assessment independently, and typically without knowledge of how similar interventions are being evaluated or considering the conditions necessary for comparison.

4.2 Learning from Health Sciences

Other research fields have had significantly more success in implementing cost-effectiveness analysis. For example, a search of the CEA Registry in health yields over 500 CEAs on the topic ‘cancer’ and over 200 CEAs on the topic ‘diabetes’. (On injury prevention, see Miller and Levy, 2000; on crime, see Marsh et al., 2008). Simply in absolute terms, other disciplines have applied CEA far more frequently. There is much to learn from this body of research.

Consequently, there are simply many more high quality CEAs in health. There are more exemplar studies and more chances to explore methodological differences in how CEA is performed. In contrast, because there are so few studies in education, there is considerably little investigation of methodology *per se*. Specifically, CEA in the health sciences has paid much greater attention to three key areas: cost-utility analysis; sensitivity testing; and statistical testing. On cost-utility analysis, there is an extremely large literature on how to value a Quality-Adjusted Life Year (QALY). Even as these valuation issues are far from settled (see for example, Drummond et al., 2009), there are likely to be many methodological lessons for education researchers performing cost-utility analysis. In particular, Bayesian statistics and inference (Jackman, 2009) are used to address the relation between information and uncertainty, a condition also present in educational evaluation. On sensitivity testing, there is attention to Monte Carlo testing and boot-strapping methods, see Jain et al. (2011). On sample sizes and power of the test for CEA, Gardner et al. (2000) and Glick (2011) provide considerable detail on how to ensure that tests for statistical significance can be applied. Analogous literature for education researchers is completely absent. As this literature addresses modeling and statistical issues, it seems likely that it does contain valuable information for CEA.

Yet, it is important to recognize that the practice of CEA in the health sciences does not always accord with best practices. Some studies in the health sciences are compromised by the lack of a clear distinction between ingredients and prices and insufficient details on costs. For example, in their Consensus statement on CEA for health sciences researchers, Siegel et al. (1996) state: "For cost data, method for inflation adjustment, type of currency, and year of costs should be indicated. Sources of data on health care utilization (physical units) and unit costs, and methods for measuring and valuing time costs are also important." That is, although it is implied that

Cost-Effectiveness Analysis: Methodological Concerns

ingredients and prices are distinct, it is not clearly stated, nor is it given priority or emphasis. As another example, in their article on the cost-effectiveness of retroviral therapy for HIV, a paper which has been cited over 300 times, Freedberg et al. (2001) devote only a short paragraph to costs and provide no information on how inputs and prices were distinguished. Also, some (perhaps most) studies rely on hospital costing data. The extent to which hospitals accurately cost out procedures is not discussed - nor is the distortion between these hospital accounting costs and market prices. Other studies rely on 'service utilization studies'. The validity of these studies is not discussed, particularly as to how they might apply to new interventions which change practice.

There is some concern in the health literature on improving practice. However, in the ISPOR Health Science Policy (McGhan et al., 2009) review of CEA practices, which was an attempt to improve CEA practice, no mention is made of how to actually collect cost data. More strikingly, in their review of effectiveness research in health, Chandra et al. (2011) conclude that compared to effectiveness “costs are much easier to measure and can be appended at a later date”. We believe this advice is misguided due to the challenges previously described in retroactive collection of cost data.

Finally, we note that, despite its volume and methodological rigor, this health literature has not always been welcomed in policy decision-making (at least in the U.S.). This is partly because of weak links between practice and what is needed for decision-making. Detsky and Laupacis (2007) assert that rarely are CE ratios put into the appropriate comparative context to help decision-makers, for example. But it is also partly because of squeamishness about efficiency in health care decisions. For the national Affordable Care Act, Congress has explicitly proscribed the use of CEA or even cost analysis. The development of CEA in education should therefore be mindful

that simply producing more analyses, even high quality ones, is not enough to improve policy.

Thus, similar problems beset health CEAs as they do CEAs in education. The main difference is that the health sciences have advanced considerably further in addressing these problems.

4.3 Creating Tools for Harmonizing CEA Practice

A secondary goal of the aforementioned IES study was to develop tools for improving the practice of CEA in education. Such tools serve two important functions. The first is to formalize the method of CEA; the second is to make it easier to perform a CEA. In our work with colleagues at the Center for Benefit-Cost Studies in Education (CBCSE), we have created a Cost Tool Kit which is freely available by agreement with CBCSE (see <http://cbcse.org/cost-resources/>).

Since its inception, the ingredients method involved computation of cost data and display of results on financial spreadsheets, even prior to the availability of electronic versions. Application of the method was readily adaptable to a computer-based platform and CBCSE has recently built one such Excel-based platform for use by researchers. This Cost Tool allows analysts to enter ingredients data into templates which guide them on categorization of ingredients by type, entering the appropriate quantity of each ingredient, the process of assigning prices, and making appropriate adjustments for inflation and locale. The Cost Tool calculates the costs of each ingredient and provides summary tables showing total program costs, cost per participant, costs by ingredients category, and costs by year of program operation. The Cost Tool incorporates a series of price indices which make the adjustments for inflation and for price variations across states and metropolitan/non-metropolitan

settings. These indices are applied in ways similar to the application of the CPI for translating nominal dollars into real dollars. If the user enters effectiveness data, the Cost Tool calculates a cost-effectiveness ratio for the program.

A second component of the Cost Tool Kit is a Database of Educational Resource Prices. A critical step in the application of the ingredients method – one that, as noted above, is often elided – is the separation of ingredients from prices. By providing prices for a range of education-related ingredients (from wages for special education teachers to square footage construction rates for school buildings), this separation is made more explicit. Thus, we have created a database of over 200 educational prices, from sources including the National Center for Educational Statistics (NCES), the American Federation of Teachers (AFT), and the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor.

A User Manual accompanying the Cost Tool Kit provides step-by-step instructions in how to use the Cost Tool and Database of Educational Resource Prices to conduct a basic cost and cost-effectiveness analysis. In addition, an example cost analysis is provided of a hypothetical education program to illustrate some of the required calculations for establishing costs of ingredients. We hope these tools can be adopted by other researchers so that performing CEA is not only easier, but more formalized, and can be harmonized with emerging literature. Other tools – to address statistical testing for costs or sensitivity analysis – should also be developed.

5. Recommendations for CEA Evaluations In Education

Our work, spanning over four decades, has produced an array of publications that use the ingredients method of costs, both for CEA and CBA (See www.cbcse.org). Yet, the use of CEA in educational evaluation is still relatively rare, even if the

ingredients method is widely accepted. Educational evaluators usually lack familiarity with the technique, and decision-makers have not incorporated CEA results into evidence-based decisions. Even among economists, the recent flurry of sophisticated studies to identify educational impacts of alternative interventions rarely include a parallel attempt to measure costs and conduct a cost-effectiveness comparison.

Both our extended experience addressing issues surrounding cost-effectiveness of educational interventions and the CEA (and CBA) studies that have been undertaken provided valuable lessons for us to share with evaluators, decision-makers, and the WWC. The IES, National Science Foundation, and private foundations supporting research and evaluation in education could have a powerful influence by requiring an integration of both cost and effectiveness dimensions into overall educational evaluations. Yet these agencies would need to have confidence that the method is rigorously applied. In this section we propose a number of recommendations which, if adopted, would improve the production, validity, and use of cost-effectiveness studies in education.

Recommendation One—Concurrent Evaluation of Costs and Effectiveness

Educational evaluations should encompass costs and effectiveness of interventions in a combined process. Both are central to an evidence-based decision process when any choice has both resource and outcome implications.

By conducting the evaluations of costs and effectiveness concurrently, the two dimensions can be mutually informed of the details of the implementation process. This entails careful observations of the actual implementation of the intervention in terms of the resources that are used and how they are used, i.e., by learning in detail what the intervention is. Data on resources can be collected systematically, providing a useful record in the event that questions arise after the evaluation is completed. With such a

record of ingredients and process, it is possible to construct a complete picture of the program components and their integration. This information is far more difficult to reconstruct from partial documentation and vague recollection as time passes. It is more likely to be incomplete and distorted when relying on historical records. Finally, the marginal costs of adding a cost evaluation dimension to an overall assessment of effectiveness of an intervention are very small in comparison with a completely separate cost evaluation executed later. By delaying the cost evaluation, information is lost completely or must be reconstructed independently when it was readily obtainable with little effort at the time of the initial evaluation. Not only is accuracy compromised, but the expense of adding a cost dimension to evaluation is also increased substantially by undertaking it separately. Finally, unless these analyses are performed concurrently it may prove next-to-impossible to calculate the resources available to the control group.

Recommendation Two—Work with Common Goals, Measures, and Populations

Cost-effectiveness analysis compares alternative interventions that address the same goals and serve similar populations with regard to efficiency in the use of social resources. The method of accounting for costs and the desirable outcomes and their measures must be similar and for comparable populations. Effect sizes among interventions dedicated to such broad goals as dropout prevention or early childhood literacy cannot be compared unless the outcome measures are the same and they are evaluated across equivalent populations. Simply taking effect sizes under a broad category of educational endeavor and comparing them is inappropriate. Even the most sophisticated research designs with equivalent groups must be limited to comparisons for the populations and measures of outcome that are represented.

Evaluation research studies submitted to WWC are restricted to the populations addressed in those independent studies and the outcome measures used in those studies.

Any interpretation of each study as representing a generic intervention is misleading. At best one can evaluate different interventions being compared in a single study where the populations and metrics of effectiveness are identical. Even in these cases, the cost analysis may be handicapped in that new interventions are often compared with a “conventional” alternative that is not clearly described in terms of its cost implications. Of course, Recommendation One addresses this situation by documenting all processes and ingredients including the “traditional” and the alternatives.

A potential solution to the challenge of comparable populations and outcome measures rests with those agencies that fund evaluation research in education. For every major substantive category of inquiry on educational effectiveness, these agencies can charge a group of experts to set out a range of specific goals that are of educational and social value and specific metrics for each that are considered to be of high quality and reasonably interchangeable or capable of being equated in statistical comparisons. The funding agencies can require that in addition to the “most appropriate” measures of outcome chosen by the researchers, that the studies also incorporate at least one of the more accepted measures that are comparable among all studies with similar goals. The expert group can also attempt to set definitional categories of study populations by race, socio-economic status, age, dependency (e.g., teenage mothers) and other criteria so that comparisons can be restricted to or reasonably adjusted for comparable populations. Of course, this is a long-term quest, but incentives can be provided in requests for proposals to meet some general standards of comparability, even in the short run, and to establish a precedent.

Recommendation Three—Use Multi-Site Studies for Evaluation

Multi-site studies are increasingly common in order to gain statistical power or to seek average effects among sites representative of different regions or populations.

As noted above in our studies, several interventions that we evaluated for CEA were implemented across multiple sites. In all these cases, costs varied among sites as did effect sizes, and the cost-effectiveness ratios were influenced by both types of variation.

For an evidenced-based decision where the results vary so much among sites, we should be far more concerned with the unique characteristics of each evaluation site as they match the site of the decision-maker in terms of demography, organization, leadership, and resource allocation. This information can suggest what might be expected in a particular decision situation rather than inferring from the average effect or average cost-effectiveness. At the very least, confidence intervals need to be established to forewarn decision-makers on the variability of outcomes.

Even better would be further analyses that might identify possible causes for the differences. The cost-effectiveness of the same “generic” intervention is affected by both the unique implementation process and resource allocation in the applied setting. Cost analysis can reveal differences in ingredients patterns and in total cost among sites explaining some of the differences in results. But, the differences among sites that account for the vast disparities in effectiveness should also be tentatively identified in evaluations so that decision-makers can consider this information in assessing how their results might differ from the average effect size. In the future, we might wish to assist decision-makers in using Bayesian inferences to make these assessments.

By undertaking a concomitant evaluation of both effectiveness and costs, information on implementation and resource allocation patterns will be recorded. That is, the richness of the information that captures the setting, implementation process, and resource patterns can help explain why costs and effects differ, in some cases substantially, from site to site in multi-site studies. We believe that the role of systematic qualitative evaluation is central to both a study of costs and effectiveness and

that expertise from qualitative research needs to be more fully utilized in many CEA studies.

Recommendation Four—Distinguish Local from General or National Prices

Studies should differentiate between the costs of alternatives for a local decision and for a generalized national comparison. Local prices of particular ingredients – typically those prices charged to the actual intervention – will vary in systematic ways that reflect local markets. For example, in some localities the costs of personnel may be reduced or inflated relative to the costs of facilities or equipment. In the case where CEA is being carried out for a local site or region, local prices should be used. By contrast, national averages of prices should be used to calculate costs that one might expect if the intervention were delivered nationally so that CEA comparisons are not biased by local or regional prices. For education programs, alternative price adjustments by setting should also be considered. Notably, input prices for interventions delivered in urban settings are likely to differ from those in rural settings.

Recommendation Five—Accommodate New Evaluation Methods

The basic conceptual and measurement approach for cost-effectiveness analysis has been tested for almost four decades. It has undergone modest improvements and adjustments over this time. Yet, a continuing effort must be made to accommodate new evaluation methods. Currently, the cost model matches up well to effectiveness research based on field experiments and quasi-experimental designs. But, as evaluation designs gain more sophistication, cost analysis must adapt to new requirements. For example, in randomized field experiments, a typical design seeks an excess of applicants for an intervention and then randomly assigns them to treatment or control groups. But, not all of the assignees comply with the assignment. For example, there are those assigned with intent to treat (ITT) who do not participate or who actively seek a different treatment;

and the treatment on the treated (TOT) population may be only a markedly reduced percentage of the ITT population. Of course, there are ways of handling these differences on the effectiveness design, but there are also questions on the cost side. Should the average cost of the treatment be determined by the ITT or TOT population? We believe that the logic is on the side of dividing total costs by the TOT which yields a larger cost per participant.

A second example is that of whether the control group is receiving anything different in place of not being selected for treatment by the intervention being evaluated. In some cases the treatment is not just an add-on for the treated, but is compensated for in the control group by some kind of intervention. Or, alternatively, parents and teachers and students in the control group may choose other activities that compensate for their failure to be chosen for treatment, rather than remaining inert. Those activities need to be identified, and their implicit costs need to be acknowledged in the cost-effectiveness analysis. This type of information can be gathered by a qualitative evaluator or ethnographer when documenting the process and ingredients for both treatment and control students using the approach that we present in Recommendation One.

Recommendation Six—Apply Sensitivity Analysis

More extensive attempts need to be undertaken to ascertain the sensitivity of both cost and effectiveness estimates and to address uncertainty. Evaluations require multiple assumptions about treatments and uses of resources. Some of these assumptions will be more fully testable if observers document process and resources for both treatment and control participants. But, still there will be a range of plausible assumptions that knowledgeable experts may have different opinions about in terms of measuring and interpreting results. We believe that more attention should be devoted to the testing of results, using different assumptions. Of particular importance is the quest

to see if the cost-effectiveness ratios and rankings change appreciably when assumptions are varied (Wang et al., 2003). If possible, we need to derive better measures of distributions of results to provide confidence intervals under different assumptions, in addition to the average CE ratios.

Recommendation Seven— Explore Interdisciplinary Research

Education researchers need not reinvent the wheel or, rather, reinvent modifications to the wheel. Research in health sciences and other social sciences can offer valuable methodological and practical guidance. With respect to methods, such guidance might span across statistical testing, cost-utility analysis, and sensitivity analysis. With respect to practice, guidance might cover how to make policy-makers, decision-makers, and research sponsors more amenable to the findings from CEA.

Recommendation Eight—Train Evaluators in Cost Analysis and Use of the Cost Tool

Most education evaluators have not received training in economic concepts and measurement. But such training appears necessary, given the casual treatment and rhetorical nature of most cost analyses. In principle, CEA is relatively straightforward. But in practice there are many challenges and problems that the researcher needs to address. Examples of good practice are helpful, as are cost tools. In combination with modest training and materials, an evaluator can build on information collected at the time of the effectiveness evaluation. Specific scrutiny devoted to the implementation and ingredients of any particular intervention can provide inputs to a cost tool to estimate the cost component needed for CEA. We recommend that both intensive and longer-range educational opportunities be offered to provide training and tools to evaluators and decision-makers on the construction and use of cost and cost-effectiveness methods for assessing efficiency of resource use in education.

Recommendation Nine—Work with Decision-makers to Structure the CEA

The fundamental test of the utility of CEA is whether or not it influences decisions in a useful and constructive manner. Although researchers may wish to perform research unaffected by the constraints of decision-making, it is important to acknowledge such constraints and incorporate them into the research design. Such incorporation might involve recognizing that schools do not typically have the resources to implement large-scale reforms; that comparing a very low cost reform with an expensive one is uninformative; and that decision-makers have prior beliefs about what should be implemented. It is also important to consider how cost-saving or effectiveness-enhancing information can be persuasive in overcoming political bottlenecks. In this respect the presentation and demonstration of CEA findings should be considered for their accessibility and understanding to those who might influence educational policy such as citizens and the media.

6. Conclusions

Our overall recommendation – to include cost analysis and CEA as part of all educational evaluations – is not simply an employment program for economists. Rather, it reflects our belief that with formal application of the ingredients methods, an understanding of methodological sensitivities, and the use of cost tools, evaluators from across the social science disciplines should be able to undertake cost evaluations of the quality necessary for CEA. Only when such cost analyses are conducted with frequency will evaluations help decision-makers generate efficiency gains in the provision of education.

Indeed, the effort needed to promote CEA is likely to be substantial and require contributions from many groups. Fundamentally, there is a glaring need for more CEAs

Cost-Effectiveness Analysis: Methodological Concerns

to be performed. Given the variation in effectiveness and costs of interventions, and the limited relevance of CEAs at one educational level to others, provision of a few exemplary CEAs are unlikely to be sufficient. Exemplars may provide models for other CEAs, but these other CEAs will be necessary to guide decision-makers working at different education levels. Once a sizeable body of evidence has been amassed, then systematic attention can be paid to methodological issues. Yet, we caution that all such effort will be of limited power if the analysis is not performed both competently and in a manner that is useful for decision-makers.

References

- Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago, University of Chicago Press.
- Belfield, C. & Levin, H. M. (2007) *The Price We Pay: The Economic and Social Cost of Inadequate Education* (Washington, DC: The Brookings Institution).
- Chandra, A. Jena, A. B., & Skinner, J.S. (2011) The pragmatist's guide to comparative effectiveness research, *Journal of Economic Perspectives*, 25(2), 27-46.
- Clune, W. H. (2002) Method strength and policy usefulness of cost-effectiveness research, in H. M. Levin & P. McEwan, *Cost-Effectiveness and Educational Policy* (New Rochelle, NY: Eye on Education).
- Coleman, J. S. et al. (1966) *Equality of Educational Opportunity*, U.S. Office of Education (Washington, D.C.: U.S. Govt. Printing Office).
- Detsky, A.S. & A. Laupacis. (2007). Relevance of cost-effectiveness analysis to clinicians and policy makers. *Journal of the American Medical Association*, 298, 2, 221-224.
- Dhaliwal, I. Duflo, E., Glennister, R. & Tulloch, C. (2012) *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries* (Cambridge, MA: MIT Poverty Action Lab (J-Pal)).
- Diamond, P. A. & Hausman, J. A. (1994) Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives*, 8(4), 45-64.
- Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A. and E. Nord. (2009) Toward a consensus on the QALY. *Value in Health*, 12, S31-S35.

Cost-Effectiveness Analysis: Methodological Concerns

- Durlak, J. & DuPre, E. (2008) Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation, *American Journal of Community Psychology*, 41, 327-350.
- Edwards, W., ed. (1992) *Utility Theories: Measurement and Applications* (Boston: Kluwer Academic Publishers).
- Freedberg, K.A. et al. (2004). The cost effectiveness of combination antiretroviral therapy for HIV disease. *The New England Journal of Medicine*, 344, 11, 824-831.
- Gardner, J.C., Huebner, M., Jetton, J. and C.J. Bradley. (2000) Power and sample size assessments for tests of hypotheses on cost-effectiveness ratios. *Health Economics*, 9, 227-234.
- Glick, H.A. (2011) Sample size and power for cost-effectiveness analysis (part 1). *Pharmacoeconomics*, 29, 189-198.
- Hanushek, E. (1968) *The Education of Negroes and Whites*, unpublished doctoral dissertation, Department of Economics, MIT.
- Hauser, R. M. & Anderson Koenig, J., eds. (2011) *High School Dropout, Graduation, and Completion Rates: Better Data, Better Measures, Better Decisions* (Washington, DC: National Academies Press).
- Hollands, F., Pan, Y., Shand, R., Cheng, H. Levin, H. M., Belfield, C. R., Kieffer, M., Gowden, A. B., & Hanisch-Cerda, B. (2013) *Improving Early Literacy: Cost-Effectiveness Analysis of Effective Reading Programs* (New York: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University,).
- Hollands, F., Levin, H. M., Belfield, C. R., Bowden, A. B., Cheng, H., Shand, R., Pan, Y. and B. Hanisch-Cerda (2014) Cost-Effectiveness Analysis of Interventions

Cost-Effectiveness Analysis: Methodological Concerns

that Improve High School Completion. *Educational Evaluation and Policy Analysis*, forthcoming.

Jackman, S. (2009) *Bayesian Analysis for the Social Sciences*, (New York: John Wiley & Son).

Jain, R., Grabner, M. and E. Onukwugha. (2011) Sensitivity Analysis in Cost-Effectiveness Studies From Guidelines to Practice. *Pharmacoeconomics*, 29, 297-314.

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

Levin, H. M. (1975) Cost-Effectiveness Analysis in Evaluation Research, In M. Guttentag & E. L. Struening, eds., *Handbook of Evaluation Research*, Vol. 2, Chap. 5 (Beverly Hills, CA: Sage Publications).

Levin, H. M. (1983) *Cost-Effectiveness Analysis* (Beverly Hills, CA: Sage Publications).

Levin, H. M. (2001) Waiting for Godot: Cost-effectiveness analysis in education, In *New Directions for Evaluation*, R. Light, ed. (55-68), (San Francisco: Jossey-Bass).

Levin, H. M. & McEwan, P. (2001) *Cost-Effectiveness Analysis: Methods and Applications* (Monterey Park, CA: Sage Publications).

Levin, H. M. & Woo, L. (1981). The costs of computer-assisted instruction, *Economics of Education Review*, 1(1), 1-25.

Levin, H. M., Catlin, D., & Elson, A. (2007) Costs of Implementing Adolescent Literacy Programs, In D. Deshler et al., *Informed Choices for Struggling*

Cost-Effectiveness Analysis: Methodological Concerns

Adolescent Readers: A Research-Based Guide to Instructional Programs and Practices (Newark; DE: International Reading Association), Chap. 4.

Levin, H. M., Glass, G. V., & Meister, G. (1987) Cost-effectiveness of computer-assisted instruction, *Evaluation Review*, 11(1), 50-72.

Marsh, K., Chalfin, A. and J.K. Roman. (2008) What does cost-benefit analysis add to decision-making? Evidence from the criminal justice system literature. *Journal of Experimental Criminology*, 4, 117-135.

McGhan, W.F. et al. (2009). The ISPOR Good Practices for Quality Improvement of Cost-Effectiveness Research Task Force Report. *Value in Health*, 12, 8, 1086-1099.

Miller, T.R. and D.T. Levy. (2000) Cost-outcome analysis in injury prevention control: Eighty-four recent estimates for the United States. *Medical Care*, 38, 562-582.

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

Rumberger, R. (2011) *Dropping Out: Why Students Drop Out of High School and What Can Be Done About It* (Cambridge, MA: Harvard University Press).

Schultz, T.W. (1961). Investment in Human Capital. *American Economic Review* 51 (March): 1-17.

Siegel, J.E. et al. (1996). Recommendations for reporting cost-effectiveness analysis. *Journal of the American Medical Association*, 276, 16, 1339-1341.

Cost-Effectiveness Analysis: Methodological Concerns

Wang, S. J., M. Blackford, Prosser, L. A. et al. (2003) A cost-benefit analysis of electronic medical records in primary care, *The American Journal of Medicine*, 114(5), 397-403.