Empirical Support for Establishing Common Assumptions in Cost Research in Education

Robert Shand & A. Brooks Bowden

View supplementary material

Published online: 07 Jul 2021.

Submit your article to this journal

Article views: 198

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

METHODOLOGICAL STUDIES

🔓 OPEN ACCESS | Check for updates

Empirical Support for Establishing Common Assumptions in Cost Research in Education

Robert Shand[a] and A. Brooks Bowden[b] 🔟

[a]School of Education, American University, Washington, DC, USA; [b]Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

The economic evaluation of educational policies and programs employing the ingredients method for cost, cost-effectiveness, or benefit-cost analysis is no exception to the critique that economic models require an untenable number of assumptions. Educational economists must make assumptions due to two sources of uncertainty: model uncertainty, as in the well-documented debate over the selection of the appropriate social discount rate to calculate present value and empirical uncertainty due to the infeasibility of gathering sufficiently detailed data on all resources. This paper highlights the frequency of empirical assumptions made in the education literature and proposes a set of harmonized assumptions to address empirical uncertainty that can be used to increase comparability of economic evaluation across programs and across studies. By building consensus on a set of reasonable, empirically derived assumptions that are selected so as to minimally distort the results of evaluations, differences in costs, cost effectiveness, and benefit-cost ratios can be more confidently ascribed to meaningful differences in resource use, program implementation, and program effectiveness, as opposed to differences in choices made by the analyst.

Background and Motivation

Economic evaluation in education, comprising cost-effectiveness analysis, benefit-cost analysis, and other related methods, is increasingly recognized as an important tool to inform educational policy decisions and thus is used with increasing frequency (Dhaliwal et al., 2013; Sparks, 2019). Decisions made based on both effectiveness or returns and costs can serve to increase both efficiency and equity of educational outcomes by optimizing outcomes for a given level of resources. However, just as with effectiveness evidence, decisions based upon faulty cost data or without transparent and appropriate assumptions can lead to errors in judgment.

We consider the importance of assumptions about cost data for five general types of economic evaluation methods commonly applied in the field of evaluation: cost analysis,

cost-feasibility analysis, cost-effectiveness analysis, cost-utility analysis, and benefit-cost analysis (Levin et al., 2018). The most foundational of these, cost analysis, provides the basis for the others by assessing the full social cost in resource or opportunity cost terms of an intervention to produce an outcome. This entails capturing the value of all resources required to implement an intervention regardless of who pays for or provides each resource. Cost-feasibility analysis extends this by comparing the estimated costs to a budget constraint to determine if a given project is feasible. Cost-effectiveness analysis comparatively assesses costs relative to some single unit of effectiveness to determine which alternative provides the greatest effectiveness relative to its cost. Cost-utility analysis extends cost-effectiveness analysis to incorporate multiple measures of effectiveness along a common scale of utility based on subjective importance weights. Benefit-cost analysis compares costs to economic benefits—the economic value of the outcomes to society in monetary terms—of an intervention.

As Karoly (2012) notes, in a benefit-cost framework, costs should be estimated with the same rigor and attention to detail as benefits in order to draw valid inferences. Ultimately, the purpose of economic evaluation is to inform specific decisions. Building on Karoly's framework, we consider how assumptions made in the field impact two potential decisions:

- First, is a program economically justified in a general sense, without comparison to other programs? In this case, it may be acceptable to use a range of different assumptions, but those chosen should err on the side of conservatism—when in doubt, lean toward upper-bound costs and lower-bound benefits so that a decision would still be justified even if the assumptions in the analysis prove incorrect.
- Secondly, which program is preferred among alternatives? In this case, it is important to make consistent assumptions so that they are comparable, with consideration of how the exact choice of assumptions changes the results of the comparison.

The most widely recognized approach for estimating the full economic cost of a well-defined intervention, including the opportunity cost of any donated or in-kind resources, is the ingredients method (Levin et al., 2018). The ingredients method is based upon two primary principles: opportunity cost and cost accounting. Evaluation results provide a detailed account of all resources ("ingredients") required to achieve a particular outcome in a specific intervention setting. Data sources may include program documentation, budgets, observations, interviews, and surveys of those who implemented a program (Levin et al., 2018; Kolbe and Feldman, 2018). While these sources will cover the majority of data required, there are almost always instances where data are incomplete.

Recommendations for how to consistently conduct economic evaluations that already exist in the field (see, e.g., Boardman et al., 2018; Farrow and Zerbe, 2013; Karoly, 2012; Levin et al., 2018; Washington State Institute for Public Policy, 2016) are valuable but often of a more abstract or conceptual nature. Prior literature provides guidance on issues such as selection of a discount rate for present value calculations, categories of

costs and benefits to include or exclude, and the need for sensitivity analysis to test the robustness of results to assumptions.

We support these general recommendations; however, there is a need for more concrete guidance on assumptions that do not pertain to high-level modeling decisions, but rather to more empirical questions of how to measure an unknown quantity or characteristic of an ingredient.

Cost estimates can suffer from three related sources of error: specification error, when it is unclear how to attribute general resources such as overhead to specific activities, aggregation error, when it is unclear how to divide costs among activities, and measurement error, when instruments are insufficiently detailed or respondent memory is faulty leading to errors and omissions in relevant costs (Datar & Gupta, 1994).

In this paper, we provide a number of standardized values from reputable sources and offer advice on whether and how to use these values. In general, using these reference numbers will make costs more standardized across evaluations. This guidance pertains to several areas: using local or national average prices, estimating average cost per student, including administrative overhead, making conversions in time units and applying fringe benefits to personnel costs, dividing teachers among an appropriate number of students when class sizes are uncertain, adding fringe benefit rates as a fraction of salary, valuing volunteer time, portioning personnel and facilities, and estimating the average lifespan of common durable resources.[1]

Ideally, these empirical and measurement questions would be answered using data observed during the delivery of an intervention. However, for a variety of reasons, it is not always feasible to obtain valid and accurate data at the level of detail needed to complete a cost analysis. This may be due to time availability for analysts or respondents, but imprecise data could also be due to gaps in memory or knowledge of respondents (Bowden and Belfield, 2015). For instance, in order to accurately obtain a shadow price, or approximation of the economic value, for the use of a classroom, in the absence of good data on rental rates for educational facilities, an analyst would need to know the size and characteristics of the facility and its average lifetime to then estimate annualized new construction costs. If a respondent has records for this information, then often it is preferable to use values that are specific to that evaluation.

However, in many cases a respondent will not know, and a standardized set of assumptions on these values may be preferable to respondents' uninformed guesses.

Given the data collection obstacles to performing a rigorous cost analysis, a standardized set of empirical or measurement assumptions can help reduce measurement error, improve comparability of cost estimates of different interventions, and reduce the burden of data collection in cost analysis. Standardized assumptions will also reduce idiosyncratic error if those assumptions are based on larger representative samples that are more stable. Simplifying data collection for researchers and participants can increase the propagation of rigorous cost studies in education and improving the comparability of cost estimates across programs can improve their face validity. Benefit-cost results are often not trusted in the field (Flyvbjerg and Sunstein, 2015); greater consistency and transparency in assumptions can help alleviate these concerns.

---

[1]This paper aligns with work funded by the Institute of Education Sciences (Hollands et al., 2020; IES, 2020).

Standardized assumptions and shadow prices can at least partially alleviate these problems by making evaluation results more nearly comparable with one another and by increasing trust in the findings reported. Therefore, in this paper, we aim to both establish a need for a common set of assumptions in economic evaluations in education and begin an ongoing conversation about what those assumptions ought to be.

We proceed in four steps to propose such a set of assumptions:

1. Propose a framework for weighing the costs and benefits to the researcher of making assumptions versus gathering additional data in an economic evaluation
2. Summarize the assumptions currently made in the field in recently published evaluations to determine around which areas there is the most uncertainty and variability to establish a need for some level of standardization
3. Propose a set of standardized assumptions to use in cases where it is infeasible or undesirable to gather more data as a starting point
4. Consider how these assumptions might change results using examples from our own prior work and with simulation analysis

Framework for Considering Assumptions

Ultimately, the decision as to whether to seek more detailed data or make an assumption will depend on the relative costs and benefits of additional data. As is conventional with optimization problems, the maximum net benefit in obtaining incrementally more detailed data will be reached when the marginal cost of obtaining more information is equal to the marginal benefit of that information, assuming rising marginal costs and diminishing marginal benefits. Assuming that an analyst will gather the most important information first, additional detail on less prominent resources requires ever-finer grained review of documentation, recruitment of additional respondents, repeated attempts to contact or interview respondents with reduced likelihood of response, survey fatigue, and lower-quality information from respondents who are straining the limits of their own knowledge and memory. Thus, obtaining additional data in an effort to obtain a perfectly accurate and precise estimate of costs is unlikely to meaningfully impact bottom-line results or decision problems stemming from an economic evaluation. For instance, it is far more important to have accurate information about overall personnel time and qualifications, which represent the majority of costs in most interventions, than to have great detail on office supplies or the replacement cycle for staplers.

Another factor to consider is a difference in possible costs, in terms of introducing error into an analysis, of making an assumption about the quantity or characteristics of a needed resource versus an assumption about the price or market value of a resource. We present examples of both types of assumptions below but suggest that the bar for making the former type of assumption is considerably higher than for the latter. In most cases, it is more important to gather contextual data about ingredients themselves to accurately describe what is needed to replicate a program than about market prices that are less likely to distort final results.

Data and Methods

Many instances of measurement error and uncertainty in cost analysis are the result of gaps in data rendering some questions unanswerable. This paper synthesizes the best available evidence for these missing data challenges so that researchers can employ these estimates as reasonable expected values. We also document the current state of the field of economic evaluation in education with regard to these empirical and measurement assumptions, including consensus values from national surveys. Using previously published studies, we test the relative importance of these assumptions based on how robust results and decision rules are to them.

Review of Assumptions in the Literature

To assess the scope and range of assumptions most commonly made in the literature, we documented measurement and empirical assumptions—assumptions about data, rather than assumptions about methods—used in recent published economic evaluation research in education. We searched on the terms "cost," "cost-effectiveness," "cost-benefit," and "economic evaluation" appearing alongside "education" using Web of Science, Google Scholar, EconLit, and ERIC. We found 3,824 citations, the vast majority only tangentially related to education. We restricted the sample to articles published in peer-reviewed journals and public-facing reports published since 2000. We supplemented this list with additional articles that cited the 2nd edition of *Cost-Effectiveness Analysis: Methods and Applications* (Levin & McEwan, 2001), a leading textbook in the field of economic evaluation in education. This left 113 papers. Nineteen papers only mentioned costs or cost-effectiveness in passing or rhetorically and were thus excluded.

Thirteen papers were excluded because they were books or book chapters inaccessible to the research team or were in languages besides English or Spanish spoken by the research team and not available in translation. We coded the remaining 81 papers inductively, noting the assumptions made about resources or prices and thematically grouped them into categories of assumptions. For a subset of 20 papers, two researchers independently coded assumptions and discussed differences until consensus was reached.

We then gathered consensus estimates for the most commonly used measurement or empirical assumptions in education. These estimates are often either shadow prices for resources that do not have a market, or parameters that must be used to estimate a shadow price. For example, it may be necessary to transform a price into a different unit (annual to hourly) to match quantities of resources with their respective prices. Whenever possible, we used values or prices that were based on large and representative surveys to generalize to the broadest possible population. When those data were not available, we synthesized estimates from across multiple, credible sources and provide a range of values as opposed to a point estimate.

We finally examined the extent to which using these assumptions could potentially impact the results of an economic evaluation. We first provide examples of the ways each individual assumption can impact results using illustrative examples. We then test multiple assumptions in tandem using sensitivity analyses. The results of these sensitivity checks suggest which of these assumptions can be safely applied without dramatically

altering decisions and which assumptions may require more data or additional sensitivity checks.

To address the question of how the assumptions we set forth here could affect an efficiency decision for a single program, we employ a Monte Carlo simulation of a previously published benefit-cost analysis across the range of values for a subset of our assumptions (Bowden et al., 2020). This Monte Carlo analysis simulates 1,000 iterations of the results applying our assumptions to determine whether any particular assumptions, or combinations thereof, dramatically alter the results. The advantage of Monte Carlo is that it can incorporate multiple assumptions across a range of values simultaneously (Boardman et al., 2018).

The benefit-cost analysis example we use is an evaluation of the City Connects comprehensive student support service program which provides trained counselors to schools to evaluate and support the strengths and needs of every student throughout elementary school. A key component of the counselor's role is to pair students with a tailored set of support services provided by the counselors, the school, and community partner agencies. The cost and benefit estimates we use for demonstration purposes here are based upon six years of program delivery at one site in 2013 constant dollars, as reported in Bowden et al., (2020). For simplicity, we explore the importance of various assumptions using cost data from one site rather than the pooled average estimates, which are less helpful for this example because the average masks specific assumptions made about resource quantities and prices at the site level.

To perform the Monte Carlo analysis, we re-estimated costs using simulated values for a range of assumptions drawn from uniform distributions related to the number of hours in an academic year, the size of a classroom, the lifetime of facilities, the fringe benefit rate as a percentage of salary, and any additional overhead or indirect costs related to administration, maintenance, and utilities. Because this demonstration focuses on the costs, we hold the benefits and other assumptions used in the original analysis, such as the discount rate of 3.5%, constant. Using this approach, we drew 1,000 cost estimates and determined the percentage of time the benefit-cost ratio was less than one, examined changes in the minimum and maximum net benefit estimates, and identified which assumptions drove the pattern of results.

To test the sensitivity of these assumptions to the ranking of programs with respect to cost-effectiveness, we also performed a break-even analysis on a number of assumptions using data on the cost-effectiveness of two alphabetics programs for older struggling readers, Wilson Reading and Corrective Reading (from Hollands et al., 2016). Both programs entail small-group instruction by specially trained, certified reading teachers, following a set curriculum. The break- even analysis proceeds by altering parameter values for a series of assumptions to determine at what level the rank ordering of the programs' costs changes.

Measurement and Empirical Assumptions in the Field

Table 1 summarizes the results of our review of assumptions commonly used in recently published economic evaluations in education research. The most common assumptions relate to the setup of the analysis and the basis for estimating personnel and facilities

Table 1. Summary of assumptions observed in sample.

Assumption	Number	%	Details of assumption
Personnel salary	28	34.6	National average (such as BLS) or actual salaries. Assumed standard time such as year, week, or day length in hours.
Perspective	19	23.5	Societal/social perspective. Institutional perspective. Perspective of various funders (state, taxpayer, nonprofits, etc.)
Facility costs	19	23.5	New construction vs. rental costs, lifespan, share of time used.
Overhead	18	22.2	Overhead is specifically excluded or a set percent of partial or total costs.
Volunteer time	16	19.8	Accounting for volunteer time using an appropriate shadow price to represent the opportunity cost.
Development costs	15	18.5	Development costs amortized over their useful life vs. included in initial year vs. specifically excluded.
Beneficiary costs	15	18.5	Costs to beneficiaries included based on opportunity cost principle.
Local/National	14	17.3	National prices used to facilitate comparison. Local prices used when appropriate for program/research questions.
Costs of attrition	13	16.0	Intent to treat and treatment on the treated are both frequent, as is including some variation of both.
Fringe benefits for personnel	11	13.6	Benefits set to a percentage of wage. Some nationally standard rates used (such as BLS or Census). Benefits excluded in isolated cases.
Training costs	11	13.6	Training costs include costs of training, trainee time and may include food and travel. Assumptions made about training length, participation rates, and frequency. Amortization decisions.
Class size	8	9.9	Assumed a standard class size based either on existing statistics/literature or on program design and context.
Transfers and stipends	8	9.9	Transfers and stipends are specifically excluded, or stipends are included as personnel costs.
Recruitment cost	4	4.9	Costs to recruit personnel, and/or participants included as an estimate of costs or as a flat fee.
Unknown personnel	3	3.7	Created informed estimates for time and compensation of unknown workers.
Extrapolation	2	2.5	Extrapolation used to draw conclusions about relatively small programs taken to larger scale.

costs, often the two largest categories of costs in education programs. Other common types of assumptions include the categories of costs such as development or administrative overhead that should be included or excluded. Somewhat less frequent choices about which assumptions need to be made include the use of local versus national average prices, the sample over which to divide costs, and how to treat cash transfers and related transactions that may have a financial implication but do not count as costs in an economic sense. Significant categories of assumptions, for which we make recommendations below, relate to measurement questions, such as the time basis for a personnel position or facility, are only made in approximately one-third and one-fifth of papers, respectively. Note also that these types of measurement assumptions are common and necessary; over half of papers made 3 or more such assumptions, with a mean number of assumptions per paper of 3. When papers do not make assumptions in these areas, they either base their decisions on empirical observation in their particular context or, more commonly, are not sufficiently transparent or detailed to determine what

choices were made. Specific assumptions disaggregated by individual paper are available in the supplemental online appendix.

Recommendations for Assumptions

Tables 2–5 report a range of suggested starting values for a series of assumptions, based upon credible sources and using nationally representative survey data whenever available. If more than one source was available, we provide a low, middle, and high price (which are further examined in the sensitivity analysis, below) with a recommended preferred value, generally based on what is most common or defaulting to the middle value, but also a suggestion to select the value most appropriate for the evaluation context and to test the range of values in sensitivity analysis. This section further discusses each of the assumptions, in turn.

There are a large number of methodological assumptions that must be made to perform a cost-effectiveness or benefit-cost analysis, including the selection of outcomes, the perspective to employ (societal or that of one particular stakeholder), the respective baseline or comparison group, and the appropriate timeframe for follow-up in measuring long-run costs and benefits.

These are discussed at length in Karoly (2012), Levin and Belfield (2015), and many other sources, and thus are not addressed directly in this paper. We do include in our review of assumptions in the literature some research design decisions, such as whether a social perspective or the perspective of particular stakeholders was employed, but do not make specific recommendations for these high-level decisions, in part because these decisions will depend greatly on the context of a particular evaluation. Rather, we seek to draw attention to the more quotidian but still significant range of assumptions that often must be made when estimating costs, which in turn also apply to cost-effectiveness

Table 2. Suggested assumptions and sources: research design and overall costs.

Assumption	Recommended assumption	Source/Notes
Local or National Prices	National average prices	Hutubessy, Chisholm and Edejer, 2003, make a similar recommendation to facilitate international comparison of cost-effectiveness of health interventions
Sample over which to divide costs in the presence of attrition	Divide by participants who actually receive treatment	Conservative assumption to obtain upper-bound cost estimates; vary in sensitivity analysis
Incremental v. Comprehensive	Consider costs to achieve desired outcome relative to alternatives as appropriate for research design and relevant decision	Cellini and Kee (2015) note that all costs and effects in cost-effectiveness and cost-benefit analysis are relative to a status quo
Overhead—administration, maintenance	5.3% of total spending on school administration, 4.3% on general/central administration, 9.57% of facilities costs for maintenance and utilities	Census Bureau Public Elementary-Secondary Education Finance Data (Census File F-33). 38th Annual Maintenance and Operations Cost Study for Schools, American School and University, 2009. http://asumag.com/Maintenance/school-district-maintenance-operations-cost-study-200904

Table 3. Suggested assumptions and sources: personnel.

Assumption	Low	Source	Mid	Source	High	Source
Number of hours in a K-12 academic year	1,195	Contractual workday*work year: 6.64 h per day, 180 days per year. OECD Labor Force Survey	1,440	8-h day, 180 days per year	1,908	Primary Sources, Scholastic and Gates Foundation survey
Number of working hours in a calendar year	1,790	OECD Labor Force Survey	1,894	Average of 8.2 h per workday for full-time workers, according to American Time Use Survey, 2014, applied to 231 workdays per year (260 weekdays minus 9 holidays, 10 sick days, and 10 vacation days)	2,087	Federal government Office of Personnel Management
Number of hours in a higher education academic year	1,280	40 h per week for two 16-week semesters per year for 9-month contract	1,477	46.18 h per week (National Study of Postsecondary Faculty, NCES, 1998) for 2, 16-week semesters	1,952	Time Allocation Workload Study, Ziker (2014)
Fringe benefit rate (% of wages)	45.7	Bureau of Labor Statistics Private Sector			57.9	Bureau of Labor Statistics Public Sector

Notes and sources: Contractual workday: NCES 2007–2008 Schools and Staffing Survey, https://nces.ed.gov/surveys/sass/tables/sass0708_035_s1s.asp; *Primary Sources* http://www.scholastic.com/primarysources/PrimarySources3rdEditionWithAppendix.pdf; OECD average annual hours worked including part-time workers https://stats.oecd.org/Index.aspx?DataSetCode=ANHRS; OPM estimate based on 1981 GAO study of number of workdays per year: https://www.opm.gov/policy-data-oversight/pay-leave/pay-administration/fact-sheets/116278.pdf; Time Allocation Workload Study https://thebluereview.org/faculty-time-allocation/; BLS fringe benefit rates http://www.bls.gov/news.release/ecec.nr0.htm (private sector) and http://www.bls.gov/news.release/ecec.nr0.htm (public sector).

Table 4. Average class sizes in different contexts.

	Urban	Rural	Overall
Panel A. United States			
US Primary	21.2	19.5	21.7
US Middle	27.8	24.8	26.8
US High	27.6	23	26
	Low	Average	High
Panel B. International			
OECD Primary	15.4	21	30.5
OECD Secondary	15.5	22.7	32.2

Sources: Table 7, Average class size in public primary schools, middle schools, high schools, and schools with combined grades, by classroom type and selected school characteristics: 2015–2016. National Principal and Teacher Survey. https://nces.ed.gov/surveys/ntps/tables/ntps_7t_051617.asp; OECD.Stat, "Average Class Size." https://stats.oecd.org/Index.aspx?DataSetCode=EDU_CLASS.

and benefit-cost analyses. Therefore, we make recommendations in three areas: overall design such as the use of local or national prices and inclusion of administrative overhead, assumptions related to personnel time and compensation, and assumptions related to the valuation and availability of facilities.

Overall Design of Cost Analysis

Local or National Prices

The first choice when pricing the ingredients of an intervention is to elect to use local market prices or national average prices. There are advantages and disadvantages to each approach, and the selection will likely depend upon the purpose of the analysis. The major advantage to selecting national average prices is increased comparability between estimates across evaluations of interventions and programs. Since our recommendations are intended to increase comparability, we generally recommend using national rather than local prices unless there is a compelling reason for the latter. The practical importance of this assumption will depend on the share of costs that vary substantially by region or locale, especially personnel. In many cases, the costs will not vary dramatically between national averages and one specific locale, and geographic price indices can be used to approximately adjust between the two.

Example. The Boston area price estimate of the direct costs of City Connects is $1,540 per year and the national price esimtate is $1,370 per year, a difference of approximately 10 percent (Bowden et al., 2020).

Determining the Denominator in per Student Cost

When calculating an average cost per student, it is necessary to choose a number of students that received an educational intervention in the denominator. In the absence of attrition, this choice is straightforward, but if students join and leave a program over time, the analyst must decide whether to divide costs over the number of students who started the program, the number of students who completed the program, or a weighted average of the two. With improved attendance data, it may be possible to look more closely at how participation changed over time. In some situations with high levels of

Table 5. Suggested assumptions and sources: facilities.

Assumption	Low	Source	Mid	Source	High	Source
Average useful life of an educational building	19	National Center for Education Statistics			44	National Center for Education Statistics
Number of hours per year a building or piece of technology is available for use	1,440	8 h per day for 180 day academic year; minimum for primary purpose of school plus after-school programs/tutoring.	3,500	14 h per day, 5 days per week (based on Bloomington, MN school district, extra charges after 10 p.m. or weekends)	8,760	Theoretical maximum – 24 h per day, 365 days per year
Average classroom size	600	Georgia Department of Education, Guidelines for Square Footage Requirements	900	Hawkins, H. L. & Lilly, H. E. (1998). Guide for School Facility Appraisal. Phoenix, AZ: CEFPI.	1,664	Kenneth Tanner, University of Georgia
Average gym size	4,200	Standard basketball court size	6,200	Virginia Department of Education Facilities Guidelines	9,540	Physical Education Secondary Schools Facilities and Basic Equipment, 1966, Massachusetts School Building Assistance Commission
Average library size			3,996	Virginia Department of Education Facilities Guidelines		
Average auditorium size			8,008	Virginia Department of Education Facilities Guidelines		
Average office size	114	Voss, J. (n.d.) Revisiting Office Space Standards			280	Voss, J. (n.d.) Revisiting Office Space Standards

Notes and Sources: NCES Condition of School Facilities, 2012–2013, Average Functional Age of School Facilities (time since construction or major renovation) http://nces.ed.gov/pubs2014/2014022.pdf; http://nces.ed.gov/pubs2014/2014022.pdf—NCES Conditions of School Facilities, 2012–2013, average age of school facilities; University of Georgia average classroom size http://sdpl.coe.uga.edu/research/territoriality.html; VA Department of Education Facilities Guidelines, http://www.doe.virginia.gov/support/facility_construction/school_construction/regs_guidelines/guidelines.pdf; Massachusetts School Building Assistance Commission http://files.eric.ed.gov/fulltext/ED021397.pdf; Voss Revisiting Office Space Standards http://www.thercfgroup.com/files/resources/Revisiting-office-space-standards-white-paper.pdf.

joiners and leavers, the analyst may need to adapt estimates of costs and effects to reflect different dosages (for more information on sample considerations due to joiners and attrition in the estimation of effects, see the What Works Clearinghouse, Standards Handbook, Version 4.1).

In the case of cost-effectiveness analysis, an important criterion is that costs and effects correspond to one another. The selection of how many students to divide costs by can be significant because some programs can easily adjust to changes in scale while those with standardized fixed costs cannot. In education, most services are provided in groups—whole classes or small groups—which require fixed costs, such as teachers and facilities. The conservative approach, which we recommend here, is to assume that to serve the number of students who will ultimately complete the program, provisions need to be made for a larger number of initial students to allow for attrition. Therefore, total costs to provide the program to the planned sample, akin to an "Intent to Treat" estimate, should be divided by the number of students who actually complete the program (akin to a "Treatment on the Treated" approach in effectiveness estimation) to reflect the higher cost of offering the program to reach the students who participated.

Example. In a cost analysis of the demonstration educational and job training program JOBSTART (a nonresidential version of the Job Corps program), Cave et al. (1993) drew a distinction between the costs per "experimental" who were assigned to treatment and the approximately 90% of the treatment sample who actually participated in treatment. In this case, the program was planned and resourced to serve more individuals than were ultimately served due to attrition. They found that the costs per participant was about 15% higher than the cost per "experimental," illustrating that the more conservative approach is to use the number of participants who received treatment to calculate average costs.

Administrative Indirect/Overhead Costs

Many interventions entail costs that cannot be directly observed, even when using the ingredients method. In higher education and the philanthropic world, this is often reflected by applying an indirect or overhead cost percentage to cover such intangibles as office administration, maintenance or utilities, or IT services. There are two dangers in overestimating these costs: one is double counting (e.g., an administrative overhead percentage is applied on top of counting administrator time as an ingredient), and the other is counting these costs as variable costs that increase with the number of participants when they may, in fact, be fixed. In other words, one could argue for including overhead as a cost of educational interventions as some share of the administrative apparatus of the school must be directed to a program. Alternatively, one could also argue against including overhead on the grounds that the school will have the same level of administration, maintenance, utilities, information technology, etc., regardless of whether the program is there or not, and thus administrative overhead is not an incremental cost.

Table 2 shows the estimate for administrative overhead from the Census Bureau's Public Elementary-Secondary Education Finance Data (Census File F-33), where 4.3% of

total school expenditures are on central administration and 5.3% are on school-level administration.

According to the 38th Annual Maintenance and Operations Cost Study for Schools, (American School and University, 2009), approximately 9.6% of expenditures go to maintenance and utilities. Based on these data, we recommend being attentive to double-counting, but as long as administrative and maintenance costs are not counted separately as ingredients, an overhead charge of 19.2% is appropriate for interventions that add to the programming of a school, such as an after-school program. Other interventions may require different approaches to administrative overhead. For instance, a curricular intervention that entails swapping out one math curriculum for another may not have any incremental overhead costs, as they would be the same in the business-as-usual condition, whereas an intervention that entails creating an entirely new organization may in fact have substantially higher overhead costs.

Example. In an evaluation of the costs and benefits of Talent Search, one of the federal TRIO programs designed to promote preparation for and success in higher education especially among historically disadvantaged populations, Bowden and Belfield (2015) include an overhead charge of 8% of federal funding based on the federally required rate; in this case, actual data is preferable to an assumption. However, if such data were not available, using 9.6% for maintenance and utilities overhead on facilities would change the per student cost from $640 to $670, and using the full 19.2% rate would raise the unit cost to $730, a 14% increase.

Personnel

Converting Annual Salary to Hourly Wage

While some programs or curriculua require full-time staff, often educational interventions use personnel on a part-time basis. This may be expressed as a percentage of full-time equivalent ("FTE") or as a number of hours. The wage price value for that ingredient may, however, be expressed as an annual salary rather than an hourly wage. Converting an annual salary to an hourly wage requires an assumption about the number of working hours in a year. In education, this will vary considerably based on whether an individual works over a calendar year, a K-12 academic year (typically 10 months of the year, or 180 7-h days), or a higher education academic year (typically 9 months of the year but for 8-h days).

As seen in Table 3, there is a considerable range in empirical estimates for these values. For full-year workers, the federal government officially uses the figure of 2,087 h per year, (Office of Personnel Management, based on a 1981 Government Accounting Office study), or 40 h per week over a year. However, the American Time Use Survey and OECD estimate relatively lower values of 1,894 and 1,790 h per year, respectively, accounting for holidays and paid leave. For academic year workers, the variance is considerably higher. According to the 2007–2008 Schools and Staffing Survey, the average contractual work year for K-12 teachers is approximately 1200 h per year. This raises the question of how to account for additional unpaid time, with teachers self-reporting almost 2000 working hours per year (*Primary Sources*, a Gates Foundation and

Scholastic survey of K-12 teachers, and the Time Allocation Workload study for higher education workers, by Ziker et al. (2014)).

Given the wide range, the exact choice of hours in a year may depend upon the particular evaluation setting. In general, we recommend the conservative choice of using teachers' contractual working hours as a basis for converting between annual salary and hourly wage to arrive at an appropriate opportunity cost value of teacher time, but then to perform sensitivity analysis within the range. While it is important to acknowledge and take into account the additional hours that teachers voluntarily work, dividing their salary by unpaid work hours would likely be inappropriate and lead to an underestimate of their hourly wage.

Example. In an analysis of the costs of Responsive Classroom, a pedagogical approach to incorporating elements of social and emotional learning into the regular school day, the baseline costs including teacher planning and implementation time were $2,160 per student assuming teachers work 180 days per year for 7 h per day (Belfield et al, 2015). With the higher figure of 1,908 h per year per teacher self-reports, the estimated cost per student declines to $1,650 per student, a 23% decrease.

Average Class Size

In some cases, an intervention may be provided to a number of students divided into classes but the exact number of students per classroom or teacher may be unknown or highly variable. To estimate an individual student's share of teacher time one must make an assumption about average class sizes to determine the number of students over which to divide a teacher's salary to estimate a per student cost. The National Teacher and Principal Survey of 2015–2016, conducted by the National Center for Education Statistics, estimates a range of class sizes for different types of schools, grade configurations, and educational settings in the United States.

The Organization for Economic Cooperation and Development does so similarly for several other countries around the world; we summarize these estimates in Table 4 (US Department of Education National Center for Education Statistics, 2016; Organization for Economic Co-operation and Development [OECD], 2020). Based on these data, we recommend using an average class size of 21 for primary schools and 26 for secondary schools, making appropriate adjustments and performing sensitivity analysis as appropriate to the particular study.

Fringe Benefit Rate

It is widely accepted that fringe benefits should be included as part of the cost of personnel, particularly if the personnel are full-time. Even if personnel are reallocated from one task to another, or are working extra beyond their full time hours (but are not necessarily receiving any additional fringe benefits), fringe benefits costs should be included to reflect the full value of their time in opportunity cost terms and to reflect replication costs as in other settings additional personnel may need to be hired to cover certain responsibilities.

Whether or not to include fringe benefits for part-time workers or volunteers, who may not receive any benefits, is subject to debate. A conservative approach would be to

include benefits on opportunity cost and replication grounds. Even if a volunteer receives no benefits, if a worker would need to be hired to replace him/her to replicate the intervention, that worker would receive fringe benefits. Part-time employees who do not receive benefits incur fringe in the form of employer contributions to payroll taxes and other administrative costs of employment. Another question is determining what fringe benefit rate to use. The Bureau of Labor Statistics has estimated fringe benefit costs for both public and private sector employees in the education sector. Fringe benefits represent 28.2% of total compensation for private sector education workers and 35% for public sector workers. However, when applying these fringe benefit rates on top of wages, they should be applied over the smaller denominator that wages represent as part of the whole: 28.2% of the 71.8% wage share for private sector workers, or 39.2%, and 35% of 65% for public sector workers, or 53.8% (Tare & Brown, 2019).

Example. The social and emotional learning (SEL) program 4Rs incorporates SEL themes into English language arts classes through literary analysis. If the fringe benefit rate were updated from 29.5% to 57.9% of salary, the per student cost shifts from $510 to $580 (Belfield et al, 2015; Long et al., 2015). Depending on the share of costs devoted to personnel, this shift could be significant when comparing with another SEL program. In a comparative analysis, it is important to use a consistent fringe benefit rate that is transparently reported so it could be tested as it applies to different contexts.

Training Costs and Amortization

Just as investments in physical capital should be amortized to annualize their high up-front costs over the year-by-year "use" of the resources, in many cases it also makes sense to annualize training costs as an investment in human capital. Failing to do so when it is appropriate will overstate the up-front costs and understate subsequent year costs. The determination of when it is appropriate, and over what time period, will depend in part on whether the training is highly specific to the intervention in question or whether it represents a general investment in human capital. It will also depend upon whether the program is likely to continue in future years, and if the training will need to be repeated, either as a "refresher" course or due to staff turnover. Based on the finding that approximately half of teachers leave teaching within 5 years (Ingersoll and Smith, 2003), we recommend annualizing the costs of training for a program over five years.

Example. Consider a training-intensive program, such as the one-to-one reading remediation program Reading Recovery. Hollands et al. (2016) estimated the costs of Reading Recovery, amortizing training costs over observed average tenures for teacher leaders (7.4 years) and teachers (4.3 years). If these data were unavailable and training costs were instead amortized over 5 years, the estimated per student cost would reduce from about $3,860 per year to $3,850 per year, suggesting that if an analyst knew the effects of training would last for more than one year but was not sure exactly how long, making a reasonable assumption based on average tenure is unlikely to be highly distortionary. Alternatively, if the length of time the training will last is unknown and training costs were not amortized at all, the per student cost would be $4,390, a much larger difference from the original estimate. This suggests that making the assumption that training will last one year versus multiple years is likely to impact bottom-line results and it

is worthwhile to obtain more data to determine if training lasts for more than one year. However, if the exact length of time beyond one year is unknown, a simplifying assumption is appropriate.

Shadow Price for Volunteer Labor

As has been widely discussed in the literature (see Cook, Chapter 3 in Farrow & Zerbe, 2013; Brodman & Powers, 2014), the shadow price for volunteer time is an important and often overlooked aspect of cost estimation. The "true" shadow price depends in part on perspective. From the perspective of a program sponsor or a decision-maker determining whether to replicate or scale up a program, the best estimate is the market value of the services performed by the volunteers. In other words, what would it cost to purchase similar services on the labor market if unpaid volunteers were unavailable? The shadow price from the volunteer's perspective may be different—on the one hand, volunteers give their time, by definition, voluntarily and thus arguably derive some intrinsic value from their work as a compensating differential. Thus, the shadow price may be lower than the market value of the services provided. On the other hand, their true opportunity cost in terms of foregone wages or leisure may be much higher than the market value of the services. It is the volunteers themselves who are bearing this additional cost, arguably due to the intrinsic pleasure they derive from volunteering, but if particular characteristics are required of volunteers that are not reflected in the market value of their services, these should be noted in the analysis.

Example. The evaluation by Jacob et al. (2016) of a supplemental reading program that relies upon volunteer tutors to read with students who were behind grade level took two approaches to explore how to value the time of the volunteer tutors. Based on the knowledge of contextual differences in volunteer pools and the relationship between reading skill and teacher effectiveness, they collected data on education level from the volunteers. However, the program was designed to work with any literate adult serving as a volunteer. The analysis used the price of a teaching assistant as the hourly value of the volunteer tutor time. In a sensitivity analysis, they found that the cost was not sensitive to the value of volunteer time.

Salaries

There are a wide range of national average salaries for educational positions based on education level, experience, and special training and qualifications, as well as various national surveys and compendia of such prices such as the Center for Benefit-Cost Studies of Education Database of Educational Resource Prices. The array of possible prices can be overwhelming, especially if details about qualifications required to match to an exact price are unknown, irrelevant, or vary widely. In such cases, we recommend using national average prices based on broadly representative surveys, such as the Occupational Employment Statistics Survey of the Bureau of Labor Statistics, which estimated average salaries of $59,420 for elementary school teachers, $59,660 for middle school teachers, $61,660 for high school teachers over a 10-month academic year, and $96,400 for principals over a 12-month calendar year as of 2019.

Facilities

Rental Rate for Educational Facilities

The most accurate estimate of the economic value of an educational facility space would be the rental rate for that space, for two reasons: even if the program does not need to pay rent for the use of the facilities, the foregone rental income represents the opportunity cost of using the space for the program, and if the program were to be replicated in another location without available space, it would have to pay rent for that space. However, there is not a national market for the rental of educational facilities, and while local markets do exist, they are highly idiosyncratic and do not appear to be tied to the underlying value of the assets or the opportunity cost. Given these idiosyncrasies and the lack of a national average, if using national average prices we recommend annualizing new construction and land acquisition prices over the useful life of a facility as an estimate of facilities costs.

New Construction Prices for Educational Facilities

The most comprehensive analysis of construction costs for educational facilities is by Cumming, which found an overall national average in 2020 of $327 per square foot not including land acquisition, furnishings, and professional fees such as design and permits. Analysts should consider varying these costs by school type (generally middle and high schools have more specialized spaces like science labs and music/art facilities and thus cost more), specialized type of spaces, and geographic location (Cumming, 2020).

Average Useful Life of Educational Facility

Annualizing the price of educational facilities requires an estimate of their useful life. A starting point is the average age of school facilities, which is 44 years according to the National Center for Education Statistics 2012–2013 *Conditions of School Facilities* report, as that is an indicator of time to replacement. There are several reasons, however, for that figure to be biased. On the one hand, if the goal of the cost estimate is to determine the value of facilities adequate to replicate a particular program, many educational facilities in the United States are in inadequate or suboptimal condition. On the other hand, facilities spending and capital plans at the district level are not necessarily optimizing or cost-minimizing, particularly given the design of matching grants from higher levels of government that may incentivize over-spending on capital projects. Finally, even though a school building may not be replaced for 40–50 years or more, NCES reports that it has only been 19 years, on average, since the last major rehabilitation of school buildings in the United States. Given multiple and competing sources of bias, it is likely that the best guess for the useful life of a new educational facility lies somewhere between these values; we recommend using an average of 30 years and performing sensitivity analysis.

Example. In a cost analysis of a university-school-community partnership program (Shand et al., 2018), an intermediate value of 30 years was used for the baseline estimate of $1,560 per student in annual costs, whereas assuming facilities lifetimes of 19 years

and 44 years, respectively, would result in estimates of $1,580 per student and $1,550 per student, suggesting that results are robust to reasonable ranges of values for this assumption.

Number of Hours per Year a Facility is Available for Use

Dividing the annual cost of an educational facility among multiple uses requires an assumption regarding the number of hours the facility is available for use over the course of the year. The theoretical maximum of 8,760 h (24 h per day for 365 days per year) is likely infeasible, given that most educational facilities would lay dormant overnight and even over large chunks of weekends and school vacations. Simply applying the costs to the school day (1,280 h, or 7 h per day for 180 days per year) is likely overly conservative given that most schools can and do make facilities available for use or lease for evening and weekend events. We, therefore, recommend guidelines in between these extremes as proposed by the Bloomington, Minnesota school district, whereby facilities usage requires special charges for additional maintenance and security for late-night and weekend uses; thus, facilities costs are divided among 3,500 h per year, for 14 h per day, 5 days per week.

   Example. In the aforementioned analysis of REACH, replacing the baseline assumption that facilities are available 8 h per day (2,080 h per year) with a more lenient one that allows for evening activities reduces the per student cost by only $5.

Average Facility Size

Many educational interventions require the use of facilities but respondents are not always aware of the actual size of the facilities used, those facilities spaces and sizes may vary, or they may be flexible by design. Therefore, awareness of the average size of various types of educational facilities can provide more accurate estimates when respondents are not aware of intervention-specific sizes. Estimates reported in Table 5 come from facilities guideline documents from multiple states, and range from 600 to 1,600 square feet for classrooms, 5,200–9,500 square feet for gymnasia, 4,000 square feet for libraries and media centers, and 8,000 square feet for auditoriums. We recommend selecting a reasonable value within these ranges based on the context and performing sensitivity analysis with the ranges.

Sensitivity Analysis Using These Assumptions

Given that a number of assumptions suggested in the previous section involved ranges of values, and that these assumptions take the place of measuring these values directly, it would be useful to know how sensitive or robust results are to these assumptions. The two types of sensitivity analysis, break-even analysis and Monte Carlo analysis, complement one another as they align with the two types of decision problems outlined by Karoly (2012): break-even analyses to test at what value an assumption would need to be set to change a decision and Monte Carlo analysis to show the effects on a single intervention of varying multiple assumptions simultaneously and probabilistically. We test the assumptions that are most likely to alter results because they represent large shares of the costs of the interventions in question,

Wilson Reading, Corrective Reading, and City Connects: fringe benefit rates, whether or not to annualize training costs, hours facilities are available for use for reading programs, hours in an academic year, classroom size and lifespan, and administrative overhead for City Connects.

Break-Even Analysis

This section reports break-even analyses on the Corrective and Wilson Reading programs on three key assumptions that would most likely impact results due to the design of the programs: whether or not to include fringe benefit rates, whether or not to amortize training, and how to estimate an hourly facilities rate. The purpose of the break-even analysis is to determine how extreme assumption values would need to be in order to alter a recommendation for a hypothetical decision-maker. In the baseline analysis, Corrective Reading cost $10,108 per student and $45,945 per effect size unit gain in early literacy skills, and Wilson cost $6,696 per student and $20,291 per effect size unit gain, all in 2010 dollars (Hollands et al., 2016).

Fringe Benefit Rates

Since both programs are personnel-intensive small group instruction programs, there is no break-even point for benefit rates—testing fringe benefit rates from 0% to 100% of salary, Corrective Reading had a consistently higher average cost per student than Wilson Reading under a full range of plausible assumptions. Note that this unlikely to be universally true—the more similar the programs being compared, the more robust they will be to shifts in assumptions that affect all programs similarly. If comparing a capital or technology-intensive program (say, a self-directed computer-adaptive tutoring program vs. an in-person tutoring program), then an assumption about fringe benefits will likely be more important, as it will affect one program substantially more than the other.

Annualizing Training

Wilson Reading requires substantially more training than Corrective Reading; therefore, annualizing training costs, as is done at baseline, could be making Wilson appear cheaper compared to Corrective. However, not annualizing training is not enough to break even with Corrective Reading in any case as it increases the cost of Wilson to $7,053 per student, still substantially less than the cost of Corrective Reading.

Hours of Availability for Facilities

Wilson Reading is cheaper and uses more facilities, so there would be no upper bound break-even point here (dividing facilities among more available hours per years will just make the cost estimate for Wilson even cheaper relative to Corrective). At just 55 h per year of available facilities time (which is absurd on the surface, as the facilities are in use for at least 100 h for the programs themselves), the programs break even. This number is sufficiently low that it is unlikely that any reasonable assumption about the available hours for use for facilities will affect the results of an analysis.

Table 6. Representative Monte Carlo results, city connects benefit-cost analysis.

Iteration (ranked by NPV)	Acad yr. h.	Class-room size (sq. ft.)	Fac. life	Fringe ben.	Overhead	PV costs	B/C ratio	Net benefits
Low	1,195	600	19	0.44	0			
High	1,908	1,664	44	0.58	0.19			
1	1,838.71	717.39	39.21	0.48	0.01	$5,018.20	1.89	$4,471.80
2	1,876.09	812.52	39.96	0.49	0.01	$5,034.89	1.88	$4,455.11
3	1,835.68	897.52	40.88	0.46	0.01	$5,047.17	1.88	$4,442.83
4	1,867.42	642.27	32.26	0.51	0.01	$5,072.20	1.87	$4,417.80
5	1,650.76	883.92	39.06	0.45	0.01	$5,076.19	1.87	$4,413.81
…	…	…	…	…	…	…	…	…
996	1,201.34	1,515.31	22.12	0.48	0.16	$6,951.20	1.37	$2,538.80
997	1,208.38	1,637.46	23.11	0.51	0.15	$7,001.90	1.36	$2,488.10
998	1,353.55	1,456.58	20.17	0.56	0.17	$7,029.28	1.35	$2,460.72
999	1,214.57	1,377.57	20.15	0.58	0.15	$7,074.12	1.34	$2,415.88
1,000	1,237.28	1,545.42	19.31	0.53	0.18	$7,312.11	1.30	$2,177.89

Notes: Low and High refer to lower and upper limits, respectively, of a uniform distribution from which values for the simulation were drawn in each iteration. NPV: Net present value; Acad yr. h.: Academic Year Hours. Fac.: facilities. Ben.: benefits. PV: present value. B/C: benefit/cost.

Monte Carlo Analysis

The Monte Carlo analysis of the benefit-cost results of City Connects considers the distribution of possible results for that single program under different scenarios. With 1,000 simulations across the range of assumption values, the benefits exceeded the costs of City Connects in all cases. Table 6 shows an abbreviated illustration of the simulations, sorted from high to low by net benefits. The rightmost columns, showing the present value costs, benefit-cost ratio, and net benefits, respectively, and how they vary with different values for the assumptions in the columns to the left. The fact that no combination of assumptions, even simulations that involve extreme values on a range of assumptions, change the sign of the net benefits, provides some assurance that the range of assumptions proposed here are reasonable.

Examining extreme scenarios provides some further insight from the Monte Carlo analysis. The highest net benefits are approximately $4,400 and unsurprisingly tend to occur with more hours in an academic year, smaller classrooms, longer facilities lifetimes, lower fringe benefit rates, and lower overhead percentages. The lowest net benefits of approximately $2,100 tend to occur at the opposite extremes. What is surprising, however, is which assumptions appear to be the drivers of these differences based on examination of the simulation values in outlier cases; cases with low net benefits consistently have fewer hours in the academic years, much larger classrooms, and high overhead rates. Therefore, in cases with benefit-cost ratios closer to one, these assumptions may require closer inspection.

Conclusion

This paper proposes a standardized set of assumptions related to concrete questions of measurement when conducting a cost analysis. These assumptions can help address two concerns in conducting a cost analysis –the comparability and face validity of descriptions of the ingredients for interventions, especially when sufficiently detailed data on resources or ingredients are not available, and the feasibility of conducting cost analyses

using the ingredients method on a larger scale by simplifying the data collection process.

Overall, most importantly we recommend being transparent and, whenever possible, testing the sensitivity of results to different assumptions. Be clear about when you are making an assumption, and if so if you are using a common, standardized assumption for the purpose of comparability with other interventions or argue it is more appropriate to select a value that is unique to your intervention. Beyond the selection of individual parameter values, it is most important that those values can be tested and varied by other analysts as needed for comparability and to determine how an intervention might change in different contexts.

To briefly summarize our recommendations, we recommend the use of national average prices, the final analytic sample that received the treatment and not the original sample at assignment as the basis for dividing costs, estimation of incremental costs relative to a business as usual condition as appropriate for the relevant decision and research design, and inclusion of overhead costs as appropriate while being attentive to double counting and what is truly incremental as high-level design decisions. For personnel ingredients, we recommend using conservative and standardized national average estimates of hours of work and hours for use to convert annual salaries to hourly wages, appropriate fringe benefit rates and shadow prices for volunteer labor, and amoritizing training costs. For facilities ingredients, we recommend using new construction costs and estimating the useful life, size of facility, and time available for use either specific to context or based on national average values from surveys.

The assumptions suggested here are suggested to represent analytically appropriate and defensible average values in the absence of more precise estimates. Nonetheless, they represent a starting point for discussion that can be updated as more and better data on costs of educational interventions become available. The robustness checks presented here—specifically, the break- even analysis and the Monte Carlo analysis— suggest that these assumptions are reasonable in the sense that values within the proposed ranges did not substantially alter the results of cost analyses for three specific programs. In at least some cases, reasonable assumptions based on empirical estimates will not substantially distort results when more precise, program-specific data are not available. Therefore, they can be applied to assist researchers in making high- quality, comparable estimates of cost available to decision makers, but researchers should still be transparent about when they are making assumptions and test sensitivity of their own results for robustness to those assumptions whenever possible.

Acknowledgements

Funding

opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID

A. Brooks Bowden http://orcid.org/0000-0001-6079-5456

References

*denotes Meta-analytic reference in summary of assumptions

American School and University. (2009). *38th annual maintenance and operations cost study for schools*. https://www.asumag.com/page/research/maintenance-operations-cost-study.

*Anderson, R., Ukoumunne, O. C., Sayal, K., Phillips, R., Taylor, J. A., Spears, M., Ricardo, A., Lewis, G., Millings, A., Montgomery, A. A., & Stallard, P. (2014). Cost-effectiveness of classroom-based cognitive behaviour therapy in reducing symptoms of depression in adolescents. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 55(12), 1390–1397. https://doi.org/10.1111/jcpp.12248

*Aos, S., Miller, M. G., & Mayfield, J. (2007). *Benefits and costs of K-12 educational policies: Evidence-based effects of class size reductions and full-day kindergarten*. Washington State Institute for Public Policy.

*Barnett, E. A., Bork, R. H., Mayer, A. K., Pretlow, J., Wathington, H. D., & Weiss, M. J. (2012). *Bridging the gap: An impact study of eight developmental summer bridge programs in Texas*. National Center for Postsecondary Research.

*Barnett, W. S., & Masse, L. N. (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review*, 26(1), 113–125.

*Bartik, T. J., Belford, J., Gormley, W. T., & Anderson, S. (2016). *A benefit-cost analysis of the Tulsa universal pre-K program*. Upjohn Institute Working Papers.

*Bartley, S. J., & Golek, J. H. (2004). Evaluating the cost effectiveness of online and face-to-face instruction. *Journal of Educational Technology & Society*, 7(4), 167–175.

Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit-Cost Analysis*, 6(3), 508–544. https://doi.org/10.1017/bca.2015.55

*Blonigen, B. A., Harbaugh, W. T., Singell, L. D., Horner, R. H., Irvin, L. K., & Smolkowski, K. S. (2008). Application of economic analysis to school-wide positive behavior support (SWPBS) programs. *Journal of Positive Behavior Interventions*, 10(1), 5–19. https://doi.org/10.1177/1098300707311366

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2018). *Cost-benefit analysis: Concepts and practice*. Cambridge University Press.

*Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost-effectiveness of success for all. *Educational Evaluation and Policy Analysis*, 24(4), 243–266. https://doi.org/10.3102/01623737024004243

*Bosse, H. M., Nickel, M., Huwendiek, S., Schultz, J. H., & Nikendei, C. (2015). Cost- effectiveness of peer role play and standardized patients in undergraduate communication training. *BMC Medical Education*, 15(1), 183. https://doi.org/10.1186/s12909-015-0468-1

*Bouguen, A. (2016). Adjusting content to individual student needs: Further evidence from an in-service teacher training program. *Economics of Education Review*, 50, 90–112. https://doi.org/10.1016/j.econedurev.2015.12.004

Bowden, A. B., & Belfield, C. (2015). Evaluating the Talent Search TRIO program: A benefit- cost analysis and cost-effectiveness analysis. *Journal of Benefit-Cost Analysis*, 6(3), 572–602. https://doi.org/10.1017/bca.2015.48

Bowden, A. B., Shand, R., Levin, H. M., Muroga, A., & Wang, A. (2020). An economic evaluation of the costs and benefits of providing comprehensive supports to students in elementary school. *Prevention Science*, *21*(8), 1126–1135.

*Breneman, D. W., Abraham, A. A., & Hoxby, C. M. (1998). Remediation in higher education: Its extent and cost. *Brookings Papers on Education Policy*, *1*, 359–383.

*Brewe, E., Dou, R., & Shand, R. (2018). Costs of success: Financial implications of implementation of active learning in introductory physics courses for students and administrators. *Physical Review Physics Education Research*, *14*(1), 010109. https://doi.org/10.1103/PhysRevPhysEducRes.14.010109

Brodman, J., & Powers, M. (2014). *The value of volunteer work: Shadow prices and opportunity costs*. The Shoreline.

Cave, G., Bos, J., Doolittle, F., & Toussaint, C. (1993). *JOBSTART: Final report on a program for school dropouts*. MDRC.

Cellini, S. R., & Kee, J. E. (2015). Cost-effectiveness and cost-benefit analysis. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 636–672). John Wiley & Sons.

*Chandra, D. G., Borah, M. D. (2012, February). *Cost benefit analysis of cloud computing in education*. Computing, Communication and Applications (ICCCA), 2012 International Conference on (pp. 1–6). IEEE.

*Coady, D. P., & Parker, S. W. (2004). Cost-effectiveness analysis of demand-and supply-side education interventions: The case of PROGRESA in Mexico. *Review of Development Economics*, *8*(3), 440–451. https://doi.org/10.1111/j.1467-9361.2004.00244.x

*Cohen, A., & Nachmias, R. (2006). A quantitative cost effectiveness model for Web-supported academic instruction. *The Internet and Higher Education*, *9*(2), 81–90. https://doi.org/10.1016/j.iheduc.2006.03.007

*Cohn, E., & Huches, W. W. (1994). A benefit-cost analysis of investment in college education in the United States: 1969–1985. *Economics of Education Review*, *13*(2), 109–123. https://doi.org/10.1016/0272-7757(94)90002-7

Cumming. (2020). *U.S. costs per square foot of gross floor area 2020*. https://ccorpinsights.com/costs-per-square-foot/

Datar, S., & Gupta, M. (1994). Aggregation, specification and measurement errors in product costing. *Accounting Review*, *69*(4), 567–591.

*Deogan, C., Zarabi, N., Stenstrom, N., Hogberg, P., Skarstrand, E., Manrique-Garcia, E., Neovius, K., & Mansdotter, A. (2015). Cost-effectiveness of school-based prevention of cannabis use. *Applied Health Economics and Health Policy*, *13*(5), 525–542. https://doi.org/10.1007/s40258-015-0175-4

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. *Education Policy in Developing Countries*, *17*, 285–338.

*Dollahite, J., Kenkel, D., & Thompson, C. S. (2008). An economic evaluation of the expanded food and nutrition education program. *Journal of Nutrition Education and Behavior*, *40*(3), 134–143. https://doi.org/10.1016/j.jneb.2007.08.011

Farrow, S., & Zerbe, R. O. (Eds.). (2013). *Principles and standards for benefit-cost analysis*. Edward Elgar Publishing.

Flyvbjerg, B., Sunstein, C. R. (2015). The principle of the malevolent hiding hand; or, the planning fallacy writ large. *Social Research*, September 1. https://ssrn.com/abstract=2654423

*Foster, J. M., Toma, E. F., & Troske, S. P. (2013). Does teacher professional development improve math and science outcomes and is it cost effective? *Journal of Education Finance*, *38*(3), 255–275.

*Gallard, A. J., Albritton, F., & Morgan, M. W. (2010). A comprehensive cost/benefit model: Developmental student success impact. *Journal of Developmental Education*, *34*(1), 10.

*Gilpin, G. A., & Bekkerman, A. (2012). Cost-effective hiring in US high schools: Estimating optimal teacher quantity and quality decisions. *Applied Economics Letters*, *19*(14), 1421–1424. https://doi.org/10.1080/13504851.2011.631886

*Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 5). North-Holland.

*Golberstein, E., Hirth, R. A., & Lantz, P. M. (2011). Estimating the education-health relationship: A cost-utility approach. *B E Journal of Economic Analysis & Policy*, *11*(3), 7.

*Graziose, M. M., Koch, P. A., Wang, Y. C., Gray, H. L., & Contento, I. R. (2017). Cost-effectiveness of a nutrition education curriculum intervention in elementary schools. *Journal of Nutrition Education and Behavior*, *49*(8), 684–691. https://doi.org/10.1016/j.jneb.2016.10.006

*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. https://doi.org/10.3102/0162373708327524

*Hollands, F., Bowden, A. B., Belfield, C., Levin, H. M., Cheng, H., Shand, R., Pan, Y., & Hanisch-Cerda, B. (2014). Cost-effectiveness analysis in practice: Interventions to improve high school completion. *Educational Evaluation and Policy Analysis*, *36*(3), 307–326. https://doi.org/10.3102/0162373713511850

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness*, *9*(1), 30–53. https://doi.org/10.1080/19345747.2015.1055639

Hollands, F. M., Pratt-Williams, J., Shand, R. (2020). *Cost analysis standards & guidelines 1.0*. Cost Analysis in Practice (CAP) Project. https://capproject.org/resources

*Hollands, F. M., & Tirthali, D. (2014). Resource requirements and costs of developing and delivering MOOCs. *International Review of Research in Open and Distance Learning*, *15*(5), 113–133.

Hutubessy, R., Chisholm, D., & Edejer, T. T. T. (2003). Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost effectiveness and resource allocation*, *1*(1), 1–13.

*Iatarola, P., & Fruchter, N. (2006). An alternative method for measuring cost-effectiveness: A case study of New York City's Annenberg Challenge Grant. *Journal of Education Finance*, *31*(3), 276–296.

Ingersoll, R., & Smith, T. (2003). The wrong solution to the teacher shortage. *Educational Leadership*, *60*(8), 30–33.

Institute of Education Sciences. (2020). Cost analysis: A toolkit (IES 2020-001). U.S. Department of Education. Washington, DC: Institute of Education Sciences. https://ies.ed.gov/pubsearch/.

*Isaranuwatchai, W., Alam, F., Hoch, J., & Boet, S. (2016). A cost-effectiveness analysis of self-debriefing versus instructor debriefing for simulated crises in perioperative medicine in Canada. *Journal of Educational Evaluation for Health Professions*, *13*, 44. https://doi.org/10.3352/jeehp.2016.13.44

*Izadi, H., Johnes, G., Oskrochi, R., & Crouchley, R. (2002). Stochastic frontier estimation of a CES cost function: The case of higher education in Britain. *Economics of Education Review*, *21*(1), 63–71. https://doi.org/10.1016/S0272-7757(00)00044-3

Jacob, R., Armstrong, C., Bowden, A. B., & Pan, Y. L. (2016). Leveraging volunteers: An experimental evaluation of a tutoring program for struggling readers. *Journal of Research on Educational Effectiveness*, *9*(sup1), 67–92. https://doi.org/10.1080/19345747.2016.1138560

*Jenkins, D., Speroni, C., Belfield, C., Jaggars, S. S., & Edgecombe, N. (2010). *A model for accelerating academic success of community college remedial English students*. CCRC Working Paper No. 21. Community College Research Center, Columbia University.

*Jung, I. (2005). Cost-effectiveness of online teacher training. *Open Learning: The Journal of Open, Distance and e-Learning*, *20*(2), 131–146.

Karoly, L. A. (2012). Toward standardization of benefit-cost analysis of early childhood interventions. *Journal of Benefit-Cost Analysis*, *3*(1), 1–45. https://doi.org/10.1515/2152-2812.1085

*Kesztyus, D., Schreiber, A., Wirt, T., Wiedom, M., Dreyhaupt, J., Brandstetter, S., Koch, B., Wartha, O., Muche, R., Wabitsch, M., Kilian, R., & Steinacker, J. M. (2013). Economic evaluation of URMEL-ICE, a school-based overweight prevention programme comprising

metabolism, exercise and lifestyle intervention in children. *The European Journal of Health Economics*, *14*(2), 185–195. https://doi.org/10.1007/s10198-011-0358-3

*Kim, J. S., Guryan, J., White, T. G., Quinn, D. M., Capotosto, L., & Kingston, H. C. (2016). Delayed effects of a low-cost and large-scale summer reading intervention on elementary school children's reading comprehension. *Journal of Research on Educational Effectiveness*, *9*(sup1), 1–22. https://doi.org/10.1080/19345747.2016.1164780

*Kivela, J., Haldre, K., Part, K., Ketting, E., & Baltussen, R. (2014). Impact and cost- effectiveness analysis of the national school-based sexuality education programme in Estonia. *Sex Education*, *14*(1), 1–13. https://doi.org/10.1080/14681811.2013.813386

*Kline, P., & Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *The Quarterly Journal of Economics*, *131*(4), 1795–1848. https://doi.org/10.1093/qje/qjw027

Kolbe, T., & Feldman, R. C. (2018). Evaluating education policy & program costs. In *Complementary research methods for educational leadership and policy studies* (pp. 281–304). Palgrave Macmillan.

*Kuo, E., Stoep, A. V., McCauley, E., & Kernic, M. A. (2009). Cost-effectiveness of a school-based emotional health screening program. *The Journal of School Health*, *79*(6), 277–285. https://doi.org/10.1111/j.1746-1561.2009.00410.x

*Lare, D., Jablonski, A. M., & Salvaterra, M. (2002). Block scheduling: Is it cost-effective? *NASSP Bulletin*, *86*(630), 54–71. https://doi.org/10.1177/019263650208663007

*Lee, K. W., Kim, D. H., & Lee, H. K. (2016). Is the Meister Vocational High School more cost-effective? *International Journal of Educational Development*, *51*, 84–95. https://doi.org/10.1016/j.ijedudev.2016.07.003

*Lee, Y. (2006). To dream or not to dream: A cost-benefit analysis of the development, relief, and education for alien minors (DREAM) act. *Cornell Journal of Law and Public Policy*, *16*, 231.

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness*, *8*(3), 400–418. https://doi.org/10.1080/19345747.2014.915604

*Levin, H. M., Glass, G. V., & Meister, G. R. (1987). Cost-effectiveness of computer-assisted instruction. *Evaluation Review*, *11*(1), 50–72. https://doi.org/10.1177/0193841X8701100103

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd ed.). SAGE.

Levin, H. M., McEwan, P. J., Belfield, C. R., Bowden, A. B., & Shand, R. (2018). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (3rd ed.). SAGE.

Long, K., Brown, J. L., Jones, S. M., Aber, J. L., & Yates, B. T. (2015). Cost analysis of a school-based social and emotional learning and literacy intervention. *Journal of Benefit-Cost Analysis*, *6*(3), 545–571. https://doi.org/10.1017/bca.2015.6

*Ludwig, J., & Phillips, D. A. (2008). Long-term effects of Head Start on low-income children. *Annals of the New York Academy of Sciences*, *1136*(1), 257–268. https://doi.org/10.1196/annals.1425.005

*Machin, S., & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, *92*(5–6), 1441–1462. https://doi.org/10.1016/j.jpubeco.2007.11.008

*Marouani, M. A. (2010). More jobs for university graduates: Some policy options for Tunisia. *Applied Economics Letters*, *17*(10), 933–937.

*Massoni, S., & Vergnaud, J. C. (2012). How to improve pupils' literacy? A cost-effectiveness analysis of a French educational project. *Economics of Education Review*, *31*(1), 84–91. https://doi.org/10.1016/j.econedurev.2011.08.013

*McEwan, P. J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness*, *4*(2), 189–213. https://doi.org/10.1080/19439342.2011.649044

*McEwan, P. J., & Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis*, *22*(3), 213–239. https://doi.org/10.3102/01623737022003213

*Meyer, K. A. (2014). An analysis of the cost and cost-effectiveness of faculty development for online teaching. *Journal of Asynchronous Learning Networks*, *18*(1), n1.

*Ono, H. (2007). Does examination hell pay off? A cost–benefit analysis of "ronin" and college education in Japan. *Economics of Education Review*, *26*(3), 271–284. https://doi.org/10.1016/j.econedurev.2006.01.002

Organization for Economic Co-operation and Development. (2020). *OECD State.* https://stats.oecd.org/Index.aspx?DataSetCode=EDU_CLASS

*Papay, J. P., & Johnson, S. M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy*, *26*(5), 696–729. https://doi.org/10.1177/0895904811417584

*Perez-Arce, F., Constant, L., Loughran, D. S., & Karoly, L. A. (2012). *A cost-benefit analysis of the national guard youth ChalleNGe program.* RAND.

*Perraton, H., Creed, C., & Robinson, B. (2002). *Teacher education guidelines: Using open and distance learning. Technology, Curriculum, Cost, Evaluation.* UNESCO.

*Pretlow, J., III, & Wathington, H. D. (2012). Cost of developmental education: An update of Breneman and Haarlow. *Journal of Developmental Education*, *36*(2), 4.

*Psacharopoulos, G. (1980). *Higher education in developing countries: A cost benefit analysis.* World Bank staff working papers number 440.

*Reynolds, A. J., & Temple, J. A. (2008). Cost-effective early childhood development programs from preschool to third grade. *Annual Review of Clinical Psychology*, *4*, 109–139. https://doi.org/10.1146/annurev.clinpsy.3.022806.091411

*Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 cost-benefit analysis of the Title I Chicago child-parent centers. *Educational Evaluation and Policy Analysis*, *24*(4), 267–303. https://doi.org/10.3102/01623737024004267

*Rice, J. K., & Hall, L. J. (2008). National Board certification for teachers: What does it cost and how does it compare? *Education Finance and Policy*, *3*(3), 339–373. https://doi.org/10.1162/edfp.2008.3.3.339

*Rodriguez, P., Nussbaum, M., & Dombrovskaia, L. (2012). Evolutionary development: A model for the design, implementation, and evaluation of ICT for education programmes. *Journal of Computer Assisted Learning*, *28*(2), 81–98. https://doi.org/10.1111/j.1365-2729.2011.00419.x

*Romeo, R., Knapp, M., Tyrer, P., Crawford, M., & Oliver-Africano, P. (2009). The treatment of challenging behaviour in intellectual disabilities: Cost-effectiveness analysis. *Journal of Intellectual Disability Research*, *53*(7), 633–643. https://doi.org/10.1111/j.1365-2788.2009.01180.x

*Ryan, S. M. (2008). Reference transactions analysis: The cost-effectiveness of staffing a traditional academic reference desk. *The Journal of Academic Librarianship*, *34*(5), 389–399. https://doi.org/10.1016/j.acalib.2008.06.002

*Samuels, F., Kivela, J., Chetty, D., Herat, J., Castle, C., Ketting, E., & Baltussen, R. (2013). Advocacy for school-based sexuality education: Lessons from India and Nigeria. *Sex Education*, *13*(2), 204–213.

*Senior, E., Carr, S., & Gold, L. (2016). Strengthening support to families: Basing a family support worker at a primary school in Melbourne. *Child and Adolescent Social Work Journal*, *33*(6), 499–511. https://doi.org/10.1007/s10560-016-0445-5

*Serrano, E., McFerren, M., Lambur, M., Ellerbock, M., Hosig, K., Franz, N., Townsend, M., Baker, S., Muennig, P., & Davis, G. (2011). Cost-effectiveness model for youth EFNEP programs: What do we measure and how do we do it? *Journal of Nutrition Education and Behavior*, *43*(4), 295–302. https://doi.org/10.1016/j.jneb.2010.03.164

Shand, R., Muroga, A., Rodriguez, V., Levin, H. M. (2018). *REACH cost analysis report.* https://www.cbcse.org/publications/reach-cost-analysis.

*Sindelar, P. T., Dewey, J. F., Rosenberg, M. S., Corbett, N. L., Denslow, D., & Lotfinia, B. (2012). Cost effectiveness of alternative route special education teacher preparation. *Exceptional Children*, *79*(1), 25–42. https://doi.org/10.1177/001440291207900102

Sparks, S. D. (2019, 9 April). More education studies look at cost-effectiveness. *Education Week.* https://www.edweek.org/ew/articles/2019/04/10/more-education-studies-look-at-cost-effectiveness.html

Tare, K., & Brown, M. (2019). A look at teachers' job requirements, employer costs, and benefits. *Beyond the Numbers: Pay & Benefits*, 8(14), 1–14. https://www.bls.gov/opub/btn/volume-8/teachers-job-requirements.htm

*Tengs, T. O., Osgood, N. D., & Chen, L. L. (2001). The cost-effectiveness of intensive national school-based anti-tobacco education: Results from the tobacco policy model. *Preventive Medicine*, 33(6), 558–570. https://doi.org/10.1006/pmed.2001.0922

*Thanassoulis, E., Sotiros, D., Koronakos, G., & Despotis, D. (2018). Assessing the cost- effectiveness of university academic recruitment and promotion policies. *European Journal of Operational Research*, 264(2), 742–755. https://doi.org/10.1016/j.ejor.2017.06.046

*Tolsgaard, M. G., Kulasegaram, K. M., & Ringsted, C. (2017). Practical trials in medical education: Linking theory, practice and decision making. *Medical Education*, 51(1), 22–30. https://doi.org/10.1111/medu.13135

*Trepanier, S., Early, S., Ulrich, B., & Cherry, B. (2012). New graduate nurse residency program: A cost-benefit analysis based on turnover and contract labor usage. *Nursing Economics*, 30(4), 207.

*Tsang, M. C. (1988). Cost analysis for educational policymaking: A review of cost studies in education in developing countries. *Review of Educational Research*, 58(2), 181–230. https://doi.org/10.3102/00346543058002181

U.S. Department of Education, National Center for Education Statistics. (2016). *Schools and staffing survey (SASS), public school teacher data file, 2015–16.* https://nces.ed.gov/programs/digest/d17/tables/dt17_209.22.asp

*Villar, A., & Strong, M. (2007). Is mentoring worth the money? A benefit-cost analysis and five-year rate of return of a comprehensive mentoring program for beginning teachers. *ERS Spectrum*, 25(3), 1–17.

Washington State Institute for Public Policy. (2016). *Benefit-cost technical documentation.* WSIPP.

*Whalen, T., & Wright, D. (1999). Methodology for cost-benefit analysis of web-based tele-learning: Case study of the bell online institute. *American Journal of Distance Education*, 13(1), 24–44. https://doi.org/10.1080/08923649909527012

*Wingard, D. L., Garman, K. A., & Reznik, V. (2004). Facilitating faculty success: Outcomes and cost benefit of the UCSD National Center of Leadership in Academic Medicine. *Academic Medicine*, 79(10 Suppl), S9–S11. https://doi.org/10.1097/00001888-200410001-00003

*Yeh, S. S. (2007). The cost-effectiveness of five policies for improving student achievement. *American Journal of Evaluation*, 28(4), 416–436. https://doi.org/10.1177/1098214007307928

*Yeh, S. S., & Ritter, J. (2009). The cost-effectiveness of replacing the bottom quartile of novice teachers through value-added teacher assessment. *Journal of Education Finance*, 34(4), 426–445.

Ziker, J. P., Wintermote, A., Nolin, D., Demps, K., Genuchi, M., & Meinhardt, K. (2014). *Time distribution of faculty workload at Boise State University.* Boise State University.