Chapter 1: How cost-effective are lectures? A review of the experimental literature

Celia Brown

University of Birmingham

Clive Belfield

Teachers College, Columbia University

Introduction

As a method of evaluation, cost-effectiveness analysis is relatively simple in its principles, with a thorough presentation given by Levin and McEwan (2001). However, these simple principles belie the complexity of cost-effectiveness analysis in practice. Perhaps these complexities explain the dearth of proper cost-effectiveness studies, a dearth identified by Levin (2001), and – in this book – emphasized in Chapters by King, McEwan, and Clune (Chapters 1, 2 and 3).[1] Indeed, these contributions offer a comprehensive overview of the extant literature; and – inevitably – they respond with calls for more, high quality research. Although this call is understandable, it is made repeatedly (see Levin, 1988).

---

1 Rightly, Schiefelbein et al. (1999, 53) comment on how "current knowledge about cost-effectiveness in education is extraordinarily inadequate". If possible, this may be an understatement. Cost-effectiveness seems to be willfully overlooked in evaluations: there is no cost-effectiveness analysis of important educational programs such as 'Success for All' (see Slavin et al., 1996). And much economic analysis is either methodologically problematic (see Barber and Thompson, 1998; Udharvelyi et al., 1992) or insufficiently lucid (e.g. conflating cost-effectiveness with cost-benefit analysis, as in van der Drift, 1980). The difficulty of drawing inferences simply from cost analysis is illustrated by a review of the cost-effectiveness analysis of non-completion for further education in the UK: Fielding et al. (1998) find that, under further scrutiny, the conclusions from a simple cost analysis undertaken by the Audit Commission

But there is an alternative response, the one adopted here. This response is to *make use of the extant evidence on effectiveness, augmented with cost estimates*. This approach is efficient, because the extant research on effectiveness is sizable. (Russell (1999), for example, reviews over 90 trials on the educational impact of information technology, as against other educational modes). It also allows for generalizations as to what educational practices are cost-effective. Most importantly, it would allow for a reasonably rapid development of the cost-effectiveness literature. The ideal approach would be for costs analysis to proceed simultaneously with effectiveness analyis and for the former to be incorporated into the research design *ab initio*. But this is clearly not happening at the present time; thus a synthetic approach appears much more appealing, at least to offer approximate answers to pressing resource-use questions.

Specifically, in this paper we undertake a cost-effectiveness review of the lecture mode of (higher) education. We compare this mode with four other modes and the measure of effectiveness is how well each mode imparts information to students. Lectures are the main mode of instruction in most universities, and therefore a prime candidate for cost-effectiveness analysis. But a literature review establishes – as anticipated – scant economic research on their efficiency (see Zietz and Cochran, 1990; Bacdayan, 1997; on computer-aided instruction, see Lewis et al., 1985; or Levin et al., 1985). Confining our analysis to specific experimental research, the cost-effectiveness evidence base is not adequate as it stands: practically no experiments either report or analyze costs. However, it may be possible to assess cost-effectiveness if costs are imputed across the evidence on effectiveness. This is the main aim of our paper,

---

(1993) require substantial modification. In particular, cost analysis that fails to adjust for actual student attendance can give misleading results.

reviewing the evidence to offer an economic evaluation of lectures as a mode of education.

Our cost-effectiveness review is structured as follows. In Section 2, we report our research method for analyzing the evidence on the effectiveness of lectures and for imputing costs. This necessitates a robust method of imputation and the construction of a costing template. In Section 3, we present the results of the review, estimating the relative cost-effectiveness of lectures compared to other modes of instruction. In Section 4, we subject these estimates to a sensitivity analysis and to re-investigation as a check for robustness. In Section 5, we discuss the results, with three objectives: (i) to see what stylized facts can be drawn about the cost-effectiveness of lectures; (ii) to evaluate the usefulness of such an evidence base for management and organization of provision; and (iii) to explore the methodological requirements for cost-effectiveness reviews. Thus, our discussions relate to both points adverted to above: they serve to emphasize – as do the other authors in this volume – the necessity of cost-effectiveness analysis, and they illustrate how effectiveness data can be re-used – practically and efficiently – with imputation about costs.

Research Method for Cost-effectiveness Analysis

Research Question and Protocol

The question at issue is the relative cost-effectiveness at imparting information of lectures as compared to alternative modes of education. The other four modes are: personalized instruction, discussion modes, independent study, and 'other modes'. We apply cost-effectiveness analysis to the extant evidence on effectiveness. The research method for this review follows several discrete stages, as set out in the following protocol.

The first stage of our review is to specify the sample selection strategy. The evidence base for answering the research question is taken from Bligh (2000). Bligh selected research which was experimental, in that it compared lecture-based groups to other forms of education, with randomization or purposive allocation of students to each group. Using experimental evidence has two advantages for assessing cost-effectiveness.[2] First, the differences in cost can be isolated to the actual program (rather than to differences in resource investment by students). Second, it makes the cost analysis easier: all that is required are net cost differences between the lecture-based group and the alternative treatment group. Bligh's selection includes 298 papers from published and unpublished evaluations of lectures as a mode of education. The treatment groups can be classified into four modes: (1) personalized systems of instruction; (2) discussion/enquiry modes; (3) independent study; and (4) other modes. We use these modes for comparison against lectures.[3]

The second stage of the review is to evaluate each study, with an explicit statement of the quality and methods of research that merit inclusion in our sample. Bligh's criteria

---

2 Experimental evidence has several methodological advantages in identifying effectiveness. Trials allow for control of variables that may strongly influence educational achievement and attainment (such as family background, peer effects and parental interest, see Altonji and Dunn, 1996; Feinstein and Symons, 1999). For training programs, Bassi (1984) finds non-random selection of participants generates numerous difficulties for estimating effects. It is also possible that the assessment method in a trial will be uniform across the two modes. In actual education systems, there may be some variation in the assessments applied across providers (for a discussion, see Brasington, 1999). Enrollment may also be strategic: some enrollees may have dips in earnings which serve to allocate or direct them into training programs for example (Heckman and Smith, 1999). Therefore, using trials should satisfy Levin's (1988, 57) assertion that "*cost-effectiveness analyses should consider the quality and appropriateness of the effectiveness evaluations on which they are based*" (emphasis in the original).

3 More specific details on these modes is available, e.g. if the instruction included demonstrations, individualized instruction, mastery learning or computer-aided instruction. So the modes could be sub-

allow use of all the studies that satisfy the initial sample selection strategy, with no compelling grounds for excluding any particular study for its poor quality.

However, for our purposes, it must be possible (a) for the interventions to be grouped and to be comparable under review; (b) for costs imputation to be undertaken; and (c) for effectiveness to be expressed in a standard metric.

On the issue of comparability of interventions and cost-effectiveness meta-analysis, Levin (1988) raises both general and specific concerns. The general concern is that the two approachs – cost-effectiveness analysis and meta-analysis – were developed to address different purposes. Nonetheless, meta-analysis has proved to be highly adaptable for other contexts (e.g. in health research) and for integration with analogous methods of research synthesis (see Hedges and Cooper, 1994). Essentially, meta-analysis requires that the units really are comparable, and this requirement should be assessed pragmatically. The specific concern is that meta-analysis often generalizes from averages (e.g. average effect sizes), and yet these averages do not refer to any actual program which could be costed. In the research review conducted here, each individual research study is analysed. Although our exposition is in terms of four modes of education, the best individual trials are also identified. This approach should then avoid the problems of 'fallacious aggregation' across non-comparable studies.

To undertake costs imputation, sufficient information should be available on the resource use for each trial. Preliminary investigation of the education literature indicates that such information is rarely adequate (for education for health professionals, see Brown et al., 2001). We therefore constructed a simple template against which to

---

categorized. However, four categories are manageable for analysis and yield reasonable sample sizes. A list of the specific interventions is available from the authors.

estimate the additional costs of the experimental arm. This template, discussed further below, allows imputation of costs that are standardized across each trial.

Regarding effectiveness and the need for standardization, the measures were standardized into a Cohen effect size of the alternative mode compared to lectures (Hedges and Cooper, 1994). In most cases, these effect sizes also had to be calculated independently by the authors (sometimes based on p values). From the point estimates (and confidence intervals) of the effect size estimates, a simple trichotomy of more, less or equally effective can also be derived.

The third stage of the review involves the synthesizing of results from the collection of evidence. Two approaches are used. The first approach involves two steps. We identify the cost-effectiveness of lectures as against the other modes of education across four quadrants of a matrix: lectures can either be relatively more or less effective and relatively more or less costly than the alternative mode. Each experiment will therefore fall into one of the four costs-by-effects quadrants. Then, we draw inferences based on the frequencies in each of the four cells via a simple vote count: the numbers in each cell reflect the relative cost-effectiveness of lectures. The second approach is to use the effect sizes and net incremental costs across each trial. This approach allows cost-effectiveness to be estimated numerically and more precisely.

<INSERT FIGURE 1 HERE>

An alternative representation of cost-effectiveness is to map interventions figuratively. The basic ideas are represented in Figure 1. There, cost-effective and cost-ineffective quadrants are identified, along with a fitted diagonal line that divides the set of interventions into those where cost-effectiveness is either increasing or decreasing.

Cost Template

We use a template to estimate costs. Necessarily, for an approach which links imputed cost data to existing effectiveness data to be meaningful, the derivation of the cost data must be rigorous. As part of the research protocol, we therefore describe our estimation of costs in detail here.

To impute costs, a template was derived, breaking the cost items into components or ingredients, as per Levin and McEwan (2001).[4] Costs are divided into four ingredients. One cost is the time of provider staff, which includes instructional time, but also time for material preparation, for induction, and for assessment. Two other costs are the "software" physical inputs (e.g. learning materials) and the "hardware" physical inputs (e.g. premises and overheads). The final cost ingredient is student effort, to include assembly of learning materials, contribution to class and outside class, and psychic engagement in learning. Our approach is to consider an aggregate cost-effectiveness measure, including all costs regardless of who incurred them. Clearly, the cost-effectiveness of particular modes will differ if a provider perspective is adopted: shifting costs from instruction to discussion groups, for example, will lower costs to the provider and raise costs to the student (in terms of preparation and effort in class). This sophistication cannot be incorporated here.

For each intervention, the resources necessary were estimated relative to the lecture mode; this estimate was based on the authors' reading of each study.[5] Across particular

4 In many cases educational effectiveness researchers use classification systems to explain their interventions which do not have a ready interpretation in terms of resource use (see for example Freudenstein and Howe, 1998; Davis et al., 1999). Findings are therefore not always reported in a way that allows for ready translation into the ingredients approach.

5 This task is simplified by the close comparability across the modes; many ingredients are the same in either the lecture or the comparison mode. Some resource requirements do differ but many of the trials are

ingredients, this change could be positive or negative: moving from lectures to independent study, for example, typically reduces provider staff time but increases student effort. These changes to ingredients are expressed in standard units, and prices are attached to these units. An additional hour of staff time, for example, is priced using the appropriate salary scale. This amount is estimated at $220 per instructional hour (and so reflects the hours of preparation and assessment before and after each instructional hour). For the "software" physical inputs, a figure of $44 per hour was estimated. For the "hardware" physical inputs, an amount of $56 per hour was used. Costs are per participant, measured in year 2000 dollars, and were also adjusted for the duration of each course. For students, a wage and materials cost estimate of $12 per hour is used as the opportunity cost.[6] (Rather than deflate the costs to the actual date of each intervention, we assume each intervention was to be implemented at current prices, so as to yield a standard cost metric. This compositional adjustment makes logical sense also).

In general, this imputation method may be sensitive to two factors. One is that the unit changes in resource required are mis-measured; the other is that the prices of the inputs are inaccurate. Typically, the former arises because the studies do not fully describe what materials are used in the intervention; and the latter arises because actual costs were not reported. This sensitivity is addressed in subsequent analysis.

In order to assist the accuracy of imputing costs, a set of stylized facts about the cost function for education programs was applied. Across lectures and the treatment groups,

---

specifically set up so that only the mode of instruction is different (e.g. the premises are the same for both modes).

[6] Full salary costs per instructional hour are derived from Ragan et al. (1999), with an assumption of 100 instructional hours per year. The "software" and "hardware" costs were estimated as proportions of these salary costs, based on a review of cost function studies (Koshal and Koshal, 1999b). The earnings opportunity costs for students are derived from high-school earnings from the National Longitudinal

the distribution of provider costs was held to closely approximate that for generic education programs: two-thirds of costs being for teaching staff; one-tenth involving curriculum materials and other forms of support; and the remainder on physical premises etc. This dis-aggregation is reasonably robust for the standard lecture-based mode (Koshal and Koshal, 1995, 1999a). Generally, treatment modes that increase the amount of human resource are likely to be particularly costly.[7] For the treatment modes, it was conjectured that: (a) personalized systems of instruction requires more resources for instruction/staffing and for planning; (b) discussion/enquiry modes require more study effort than lectures; and (c) independent study requires less instruction and more study by the enrollees. This stylized facts assisted the authors in imputing net incremental costs to each intervention.

Results

The Cost-effectiveness Evidence Base

From the papers reported by Bligh (2000), the final sample of sufficient quality is 54 lecture-comparison interventions across 38 papers. The sample of 54 treatment modes breaks down across the four modes as follows: 4, personalized system of instruction; 15, other modes; 17, independent study; and 18, discussion modes. The interventions are relatively small-scale: the mean (standard deviation) number of students in the

---

Survey of Youth reported in Light (1999), with supplemental materials costs added. All figures are transformed into 2000 dollars. All costs are subject to sensitivity analysis, as reported below.

7 Inouye et al. (1997) compare the costs of five important US elementary school programs to improve reading: the variation in their costs reflects the differing intensity of the programs in terms of resources for instructors. Waterford and DISTAR were relatively intensive in terms of hardware investments; Writing to Read was moderately labour-intensive in its use of para-professionals; and Success for All and Reading Recovery – the high cost programs – were highly labour intensive.

experimental group is 47 (44); and the mean (standard deviation) number of instructional hours per intervention is 13 (10).

Research was rejected from our sample for a number of reasons. For 63 papers, these modes were essentially no different from the lecture mode, e.g. because of a cross-over design in instruction. For 92 papers, the relevant paper could not be traced by the authors (the main reason being that the publication was not in a peer-reviewed journal or was in a discontinued publication). For 89 papers, the relevant paper was unpublished and the only printed evidence was a citation abstract of one paragraph. For 16 papers, either the outcome was not a learning outcome, or there was no actual lecture delivery, or it was not possible to calculate effect size. Of these reasons, the main concern is publication bias against the untraceable or unpublished papers (92 + 89). We consider this bias below.

The Effectiveness of Lectures

We begin with an assessment of the relative effectiveness of lectures, across all of the 298 trials reported by Bligh (2000). Over half of the studies (52.0%) show no significant difference in effectiveness between lectures and other modes; the lecture mode is more effective in over one quarter (27.2%) of the studies, and less effective in one-fifth (20.8%). Based on 45 trials, lecture delivery appears to be clearly more effective than personalized systems of instruction (as 44.4% of trials show lectures to be more effective). From 109 trials, lecture delivery appears similarly effective to discussion and enquiry modes (19.1% favor the former mode, but 23.4% favor the latter, and 57.5% are inconclusive). Similarly, 40 trials show lectures and independent study appear equivalent in effectiveness (with 25.0% favoring lectures, and 22.5% against).

Finally, from 104 trials, lecture delivery appears more effective than "other" modes: 26.0% (19.2%) of trials show lectures to be more (less) effective. This much can be inferred from effectiveness analysis applied at a very broad level of generality (see Bligh (2000) for a more thorough exposition). However, this discussion of effectiveness is preliminary to a full economic evaluation.

<INSERT TABLE 1 HERE>

More detailed analysis of effectiveness is reported in Table 1. This is based on direct inquiry into each paper, but only using the available sample of interventions. These show the mean effect size across the four modes, and we use a fixed effects weighting to adjust for the quality of each study.[8] Three modes – independent study, personalized systems and discussion modes – are less effective than lectures; only the other mode has a positive effect size. None of these modes appears to be an improvement on lectures, using tests of statistical significance. Pooling the sample, lectures appear to be relatively effective, with a net mean effect size from using an alternative mode at -0.1563. However, measured per instructional hour, the net mean effect size is -0.0002 (i.e., the lecture mode involves more hours of instruction).

Costs of Intervention

Relative incremental costs per participant are reported in Table 2. The range of cost differences is -$2308 to +$3366, with a mean (standard deviation) cost difference of $52.38 ($1137.66) across all modes. On average, two of the modes are less costly than lectures – independent study and discussion – and two of the modes are more costly –

[8] This is based on the variance of the effect size (Shadish 1994).

personalized systems and other modes (although the last of these has a large standard deviation).

<INSERT TABLE 2 HERE>

Most interventions involved substituting one input for another, with changes to the input mix rather than investing more resources into a particular mode. Of the 54 interventions, 33 involved changes to the staffing mix of delivery; 38 to the materials; 40 to student effort; and 11 to the physical inputs to education. However, also of importance is the duration of the intervention. Duration has a clear influence on costs: 10 of the interventions were one hour (or less), but 30 of the interventions were the length of a higher education module/semester.

Cost-effectiveness of Lectures

We now link the effectiveness and costs information together. As a simple exposition, we look at the proportion of interventions in each mode that may either be clearly accepted as cost-effective or clearly rejected as not cost-effective relative to lectures. Modes that are unambiguously cost-effective are ones where the costs are lower, but the point estimate of effectiveness is higher. Modes that are definitely not cost-effective are ones where the costs are higher, but effectiveness is lower.

These categorical statements can be applied to our sample of interventions. Of the 18 discussion mode papers, 5 would be immediately rejected as not cost-effective and 3 would be clearly accepted as cost-effective. Of the 4 personalized system papers, 2 would be rejected and none would be accepted. Of the 17 independent study papers, 4 would be rejected and 6 accepted. Finally, for the 15 other modes, 9 would be rejected and only 2 accepted as more cost-effective.

On this cost-effectiveness vote count, only independent study emerges as more likely to be a cost-effective mode relative to lectures. Discussion modes and the personalised system appear less cost-effective than lectures. Lastly, the other mode appears to be the least cost-effective. (Even where there is correspondence between effectiveness and cost-effectiveness evidence, however, this does not invalidate the argument that cost analysis should be incorporated into evaluations).

In the aggregate, 37% of interventions would be rejected as clearly not cost-effective and only 20% are clearly cost-effective. For the remainder – almost half of the interventions – a financing constraint should be applied to justify any extra investment: these interventions are more effective, but also more costly.

A graphical representation of our results is given in Figure 2, plotting each intervention against the lecture mode in terms of relative costs and relative effects. One approach to measuring cost-effectiveness is to look at the distribution of interventions in each of the four quadrants in Figure 1. The top left quadrant identifies the most cost-effective modes and the bottom right quadrant the least cost-effective. This method is analogous to our simple tabulation of cost-effectiveness. However, it also indicates modes where there are large gains for only a small increase in costs. Three of the independent study interventions, for example, show particularly large effect sizes, for zero or trivial changes to costs. Such interventions are likely to be highly cost-effective.

<INSERT FIGURE 2 HERE>

Overall, Figure 2 shows that there does appear to be a cost-effectiveness trade-off: where alternative modes to lectures are more effective, they are also more costly. Given that many interventions did not involve substantially greater resource investment, it may in fact be legitimate to use effectiveness analysis – critically, though, this legitimacy must be established (and the variance of costs is high).

A third approach would be to use cost-effectiveness ratios. These ratios must be defined precisely, because the interpretation depends on how the ratio is calculated. One approach is to use an incremental cost-effectiveness ratio, derived as the ratio of per participant net costs over the effect size. This is the measure used by Gomel et al. (1996). However, this ratio would have to be adjusted to compensate for the fact that both costs and effect sizes can be either negative or positive. Thus, a negative ratio could arise because costs are lower but effect size positive (this would be cost-effective) or because costs are higher and effect size lower (this would be cost-ineffective). Also, because many interventions have zero effect size, this incremental cost-effectiveness ratio is not always determined.

However, cost-effectiveness rankings can be derived from Figure 2. Any of the interventions in the top-left quadrant should be implemented over lectures: all cost less than lectures and yet are more effective. None of the interventions in the bottom-right quadrant should be implemented. For the more cost-effective trials, a cost-effectiveness ratio can be derived as costs over the effect size. This ratio gives a measure of the resource impact from generating a unit effect size.

There are six trials that use independent study modes that are clearly cost-effective. Two of these are more cost-effective because they are lower cost but have an effect size of zero. For the other four studies, the resource savings are $4578, $1784, $1266, and $28. There are two trials that use the other mode, and these generate a resource saving of $262 and $191. For the discussion mode, cost-effectiveness is only generated because the costs are lower, but the effect size is zero. For personalized instruction, there are no clearly cost-effective studies. Thus, three independent study trials generate the best cost-effectiveness ratio scores; and these are followed by two other modes; and then the 'remaining' independent study trial.

Sensitivity Analysis

Validity Checks

The purpose of the review has been to seek general statements about the relative cost-effectiveness of the lecture mode over alternative modes. However, we recognize that a review of this form may be sensitive to the way the data are combined.[9] Our check addresses the internal validity of the evidence base on the cost-effectiveness of lectures: specifically, we have identified potential sources of bias or measurement error from publication and from imputation of resource costs.[10]

Publication Bias

One approach to addressing publication bias is to consider the inclusion criteria for analysis. There are three main concerns over the inclusion criteria. First, some of the

---

9 Another consideration is to restrict the sample to recent interventions. Some of the studies are over 30 years old and may be rejected as being out-of-date for the purposes of cost-effectiveness analysis. One may question whether or not the causal links still obtain — lectures are now much more interactive or visually stimulating than the basic 'chalk and talk' mode which prevailed at the time some studies were conducted. Of course the other mode may have advanced technologically also: the point here is that the rate of change of technology may not have been stable across the two modes, and it is unlikely that the prices of the inputs for the two activities have remained the same. Critically, the price of a computer-based distance-learning program relative to the price of an hour of lecturer time has changed substantially since 1980 (Levin et al., 1985).

10 Few other studies have been subject to sensitivity analysis or compared using equivalent methods. This is important, because decisions as to what is cost-effective may differ, depending on the assumptions used to estimate costs. Krueger (2000), for example, finds that the STAR experiment yields a rate of return of

studies are not available to us and this may generate publication bias. As a simple test, we plot the effect sizes against sample size to identify if they are distributed in a funnel shape (Hedges and Cooper, 1994). This plot in Figure 3 is reasonably comforting: the spread of effect sizes (positive and negative) increases as the sample size decreases and there is no obviously missing "chunk" of interventions from the bottom left hand corner of the plot, as typically occurs with publication bias (Shadish, 1994).

<INSERT FIGURE 3 HERE>

An alternative investigation of publication bias is to compare the effectiveness of our sample with Bligh's population of 398 studies, as described above. Bligh's population shows 27.2% of interventions were relatively effective; the respective figure for our sample is 28.7%.[11] Plausibly, our sample includes more research where the alternative mode is a statistically significant improvement on lectures. On inspection of all the trials, this over-sampling arises because publication bias favors such results (rather than arising because of a bias in our methods of retrieving and collating the research). However, as our analysis finds that the alternative modes are less cost-effective than lectures, the bias works against our conclusions rather than for them.

Finally, the studies in our sample are on average from research conducted more recently. In part this reflects the improved reporting conventions used in research, so that effect sizes were calculable, and this in turn suggests that fewer of the unavailable studies were likely to meet our quality criteria. (For the obtained sample, 16 of 54 papers were rejected on quality grounds). As well, older studies are likely to have less external

---

6% (and so is applying cost-benefit analysis, not cost-effectiveness analysis); Prais (1996), looking at the same experiment, finds that there are many more cost-effective re-organizations to provision.

validity, given changes in the culture of education and in the technologies of modes such as independent study. The omission of older research may in fact improve the external validity of our findings.

Re-estimation of Costs and Prices

Perhaps the most important sensitivity analysis is to check the accuracy of the costs data, because these are imputed. There are a number of well-documented practical difficulties in estimating costs directly. Breneman (1998, pp.364–367) describes the difficulties of getting consistent and reliable cost information on US college remediation programs (because of confidentiality by colleges); Harbison and Hanushek (1992) chart similar problems in accessing data in their cost-effectiveness analysis of Brazilian education. These difficulties are compounded when the educational program draws on budgets from multiple sources. These practical difficulties may make cost imputation appear economical and justifiable as a proxy for "real" costs.[12]

The specific check on costs employed here was to apply different prices to each of the inputs. This is equivalent to estimating different unit values. Three alternative vectors for the sets of prices were used, where these were distinguished by their pricing of different inputs: one vector (#2) used high materials costs; another (#3) priced student

---

11 Comparing our sample to the population from Bligh (1998) in Table 2, there also appears to be an over-representation of the independent study mode. However, the variance in results for this mode do not differ significantly from that for other modes.

[12] Other common problems with costs analysis may be obviated as a result of the general method adopted here. One such problem is effort. Although effort within an educational setting may be estimated based on the duration of the course, less readily measurable is the (unrecorded) effort of students in response to each mode. (In their analysis of the cost-effectiveness of learning in higher education, for example, Zietz and Cochran (1990) assume that the cost of inducing extra effort from the enrollees is zero). A second problem is that, typically, costs within the public sector may not be available, because market prices for the inputs may be unavailable and the depreciation of assets involved in one mode of delivery may differ from that for another. For the method applied here, however, these problems are only significant to the extent that they vary across the groups. Using experimental literature, it may be legitimate to assume these factors are differenced away.

effort at double the expected wage. A fourth vector (#4) of prices was taken from the only paper in the sample to report costs (Harding et al., 1981). Each vector was applied: to estimate costs; to derive vote-counts of clearly cost-effective and cost-ineffective interventions; and to re-draft Figure 1.

The alternative vectors did produce some variation in the costs as represented in Table 2. Specifically, vector #4 caused each of the modes to appear more costly than the lecture mode. However, under vectors #2 and #3 the signs of the relative average cost per mode were unchanged from the principal vector. As well, the alternative vectors preserved the ranks of costs: the personalized instruction mode was the most costly, and the discussion mode the least costly. For the last two re-applications, the results are largely unaffected. Re-deriving the vote-counts to identify clearly cost-effective and cost-ineffective interventions made little impact on the results. The only notable effect was that application of vector #4 served to strengthen the argument that the discussion mode was not cost-effective relative to lectures: no interventions in this mode appeared to be definitely cost-effective using this vector. Similarly, re-drafting Figure 1 added little explanatory power. (In part this reflects the fact that the quadrants indicate relative cost-effectiveness). Overall, we can be reasonably confident of the findings reported in Tables 1 and 2 and the above discussion.

Using Cost-effectiveness Analysis

Our results suggest that, based on the experimental evidence, there is no mode of education that is more clearly cost-effective than lectures for imparting information. Although the other mode appeared to be more effective than lectures, it was in many cases more costly, with a high variance in costs. For the personalised instruction mode, this was both more costly and less effective. For the discussion/enquiry and independent study modes, costs were lower, but so was effectiveness. None of these results suggests that lectures can be replaced so as to achieve more cost-effective provision for imparting information.

In part, these nugatory results are from experiments that are relatively small-scale. The intention is not to entirely change a mode of delivery, but simply to augment a particular mode with either extra lectures or the alternative mode.[13] It is therefore debatable whether such changes to education provision can yield substantive effects. Moreover, even where there are clear indications that one mode is more costly than another, using this evidence-base may not be straightforward. Specifically, whether managers can use the information from such cost-effectiveness analysis to raise educational achievement within universities will depend on a number of issues.

First, cost-effectiveness results are contingent on the scale of provision. Where study X, for example, refers to an intervention on t students, its relevance to an intervention applied to 2t students is not certain. The importance of scale-contingency becomes particularly clear when cost-effectiveness analysis is applied, because the marginal cost for an enrollee is not typically the same as the average cost. Costs data show that the cost function is typically U-shaped in enrollments (see Koshal and Koshal, 1995). Discussion groups, say, may be more cost-effective with t-x students than lectures, but less cost-effective with t+x students. Given the differences across modes in the proportions of fixed costs (and because modes will differentially impinge on the

---

13 In a catalogue of 27 interventions to boost Science teaching, for example, Anderson (1990) itemizes few that require a substantial increase in resources. From the perspective of the provider, only four out of the 27 interventions cost more than 10% of the cost of lengthening the school day by one hour: these involve

capacities of education providers, as well as on their capacity utilization), scale effects may be non-trivial.[14] To make general inferences from the evidence base, we need to know how both the effectiveness function and the costs function vary with scale or enrollment levels. As yet, the evidence base on these functions is not available.

Second, there may be costs of transition between one mode and another. Although one mode is more cost-effective, it may not be possible to transfer provision to that mode, because the resale value of the current resource provision is too low. These adjustment costs may be high, particularly for universities where a substantial proportion of education is delivered through lectures. The cost-effectiveness of the alternative mode may be offset by the liquidation costs incurred in disposal of the existing lecture-related physical assets. Consequently, cost savings from re-organization may not be that substantial: providers may be constrained by resource deployment rules. However, Johnes (1997) estimates subject-based re-organization of higher education and finds potentially large cost reductions; Mayer and Peterson (1999) present economic evaluations of various school-based policies and find substantial variation in cost-effectiveness.

Third, the most cost-effective mode for imparting information may not be the same as that for, say, generating increases in graduates' earnings or other social outcomes (for a substantial list of the effects of education, see McMahon, 2000; Herrnstein and Murray, 1994). As well, modes may be differentially effective: lectures, say, may be equally

---

giving teachers a reduced workload, lowering the teacher–student ratio, rotating in alternative professionals (from business) into the classroom, and moving toward a system of full-year institutes.

14 Also, the rate of research obsolescence reported above appears — to the authors at least — to be non-trivial. Much evaluative research is of limited usefulness beyond the immediate and proximate circumstances in which the research takes place. This raises concerns about the cost-effectiveness of undertaking cost-effectiveness analysis itself (particularly when evaluations must themselves be justified, Datta, 1999).

effective across all student abilities, whereas discussion groups may favor high ability students. Thus the results may not be generalized to school pupils or different schooling systems. Plus, effectiveness may be contingent on a portfolio of modes, rather than only one mode (see the estimates of the benefits of mixing vocational and academic courses in Kang and Bishop, 1989; Mane, 1999). Further, to predict the most cost-effective mode of education, it may be necessary to understand the enrolment decisions and time allocation decisions of students (Bacdayan, 1997). More generally, the concept of effectiveness itself may be subject to debate, perhaps to include multiple measures rather than single point estimates (see Coe and Fitz-Gibbon, 1998; Reynolds and Teddlie, 2000, 324).

Fourth, for reviews of cost-effectiveness to be generalisable, the interventions themselves must have construct validity, i.e. it must be possible to specify the intervention as a reasonably stable "technology" (Levin, 1988). Unfortunately, some of these alternative modes may not easily conform to a standard. One mode for classification was "other"; this includes interventions as diverse as laboratory-style learning; modeling with role-play; correspondence study; and audio-visual instruction. 15 This criticism of scientism, however, has ramifications for general research methods in the social sciences.

Fifth, there is the issue of choosing a cost-effective re-organization of provision. The evidence here makes clear that some discussion mode interventions were more cost-effective than lectures, and some were less cost-effective. The same applies to the other mode and to independent study. An unambiguous preference for these modes, say, over

15 An additional consideration is the reporting of information in the investigations under review. Our investigation has highlighted the terse reporting of effectiveness from most interventions and the almost non-existent costs analysis. More generally, many of the papers failed to specify in detail the nature of the intervention, even as this is a critical element of the research.

lectures cannot therefore be derived: cost-effectiveness depends on how a particular mode is implemented. However, in no case was personalized instruction more cost-effective than lectures. For this mode, an unambiguous preference for lectures can be identified.

A final issue arises where the control over the education is split across agencies. A mode that is cost-effective when viewed from the provider perspective may not be so when viewed from the enrollees' perspective. The most obvious example of this occurs with comparisons between lectures and independent study. Promotion of the latter will be lower cost from the perspective of the provider. Relatedly, the agents responding to a cost-effectiveness analysis may not be able to control some of the costs: from the perspective of enrollees, for example, lectures may be free at point-of-attendance where education is subsidized by government.

Notwithstanding these caveats to simple inference from basic cost-effectiveness analysis, our evidence base suggests that any changes to the mode of provision in universities are unlikely to yield substantial gains in learning outcomes.

References

Altonji, J.G. and T.A. Dunn (1996), 'The effects of family background on the return to education', *Review of Economics and Statistics*, **42**, 692–704.

Anderson, R.D. (1990), 'Policy decisions on improving science education: a cost-effectiveness analysis', *Journal of Research in Science Teaching*, **27**, 553–574.

Audit Commission and Office for Standards in Education (1993), *Unfinished Business: Full–Time Educational Courses for 16–19 Year Olds*, London: HMSO.

Bacdayan, A.W. (1997), 'A mathematical analysis of the learning production process and a model for determining what matters in education', *Economics of Education Review*, **16**, 25–37.

Bassi, L. (1984), 'Estimating the effects of training programs with non-random selection', *Review of Economics and Statistics*, **66**, 36–43.

Bligh, D.A. (2000). *What's The Use of Lectures?* Jossey-Bass.

Bradley, S. and J. Taylor (1998a), 'The effect of school size on exam performance in secondary schools', *Oxford Bulletin of Economics and Statistics*, **60**, 291–327.

Bradley, S. and J. Taylor (1998b), 'Cost functions in secondary schools', University of Lancaster, mimeo.

Breneman, D.W. (1998), 'Remediation in higher education: its extent and cost', in D. Ravitch (ed.) *Brookings Papers on Education Policy*, Washington, D.C.: Brookings Institution.

Brown, CA, Belfield, CR and S Field. (2001). A review of the cost-effectiveness of continuing professional development for the health professions. *British Medical Journal*, forthcoming.

Bullock, A.D. and H.R. Thomas (1997), *Schools at the Centre: A Study of Decentralisation*, London: Routledge.

Coe, R. and C.T. Fitz-Gibbon (1998), 'School effectiveness research: criticisms and recommendations', *Oxford Review of Education*, **24**, 421–438.

Cooper, H and LV Hedges (2000) *The Handbook of Research Synthesis.* Russell Sage: New York.

Datta, L-E. (2000), Seriously seeking fairness: strategies for crafting non-partisan evaluations in a partisan world. *American Journal of Evaluation*, **21**, 1-14.

Davis, DA, Thomson O'Brien, MA, Freemantle, N, Wolf, FM, Mazmanian, P and A Taylor-Vaisey. 1999. Do conferences, rounds and other traditional continuing education activities change physician behaviour or health care outcomes? *Journal of the American Medical Association*, **282**, 867-874.

Duncombe, W., J. Miner and J. Ruggiero (1995), 'Potential cost savings from school district consolidation: a case study of New York', *Economics of Education Review*, **14**, 265–284.

Feinstein, L. and R. Symons (1999), 'Attainment in secondary schools', *Oxford Economic Papers*, **51**, 300–321.

Fielding, A., C.R. Belfield and H.R. Thomas (1998), 'The consequences of drop-outs on the cost-effectiveness of 16–19 colleges', *Oxford Review of Education*, **24**, 487–511.

Freudenstein, U and A Howe. 1999. Recommendations for future studies: A systematic review of educational interventions in primary care settings. *British Journal of General Practice*, **49,** 995-1001.

Gomel MK, Wutzke SE, Hardcastle, DM, Lapsley, H and RB Reznik. 1998. Cost-effectiveness of strategies to market and train primary health care physicians in brief intervention techniques for hazardous alcohol use. Social Science Medicine, 47, 203-211.

Harbison, R.W. and E.A. Hanushek (1992), *Educational Performance of the Poor: Lessons from Northeast Brazil*, New York: Oxford University Press.

Harding, CM, Riley, IS and DA Bligh. 1981. A comparison of two teaching methods in Mathematical Statistics. *Studies in Higher Education*, **6**, 139-146.

Heckman, J.J. and J.A. Smith (1999), 'The pre-programme earnings dip and the determinants of participation in a social programme: implications for simple programme evaluation strategies', *Economic Journal*, **109**, 313–348.

Herrnstein, R.J. and C. Murray (1994), The Bell Curve: Intelligence and Class Structure in American Life, New York: Free Press.

Inouye, D.K., H.L. Miller and J.D. Fletcher (1997), 'The evaluation of system cost and replicability', *International Journal of Educational Research*, **27**, 137–152.

Johnes, G. (1997), 'Costs and industrial structure in British higher education', *Economic Journal*, **107**, 727–737.

Kang, S. and J.H. Bishop (1989), 'Vocational and academic education in high school: complements or substitutes?', *Economics of Education Review*, **8**, 133–148.

Koshal, R.K. and M. Koshal (1995), 'Quality and economies of scale in higher education', *Applied Economics*, **27**, 773–778.

Koshal, R.K. and M. Koshal (1999a), ‘Demand and supply of educational service: a case of liberal arts colleges’, *Education Economics*, **7**, 121–130.

Koshal, R.K. and M. Koshal (1999b), ‘Economies of scale and scope in higher education: a case of comprehensive universities’, *Economics of Education Review*, **18**, 269–277.

Krueger, AB. (2000), ‘Economic considerations and class size’, Princeton University Working Paper, mimeo.

Levin, HM and PJ McEwan (2001), *Cost-effectiveness: Methods and Applications.* (2nd Ed.). Beverly Hills: Sage Publications.

Levin, HM, Glass, GV and GR Meister. (1985). ‘Cost-effectiveness of computer-assisted-instruction’, *Evaluation Review*, **11**, 50–72.

Levin, HM. 1988. Cost-effectiveness and educational policy. *Educational Evaluation and Policy Analysis*, **10**, 51–70.

Levin, HM. 2001. Waiting for Godot. Cost-effectiveness Analysis in Education. In RJ Light (Ed.). *New Directions for Evaluation*, **90**, Jossey-Bass.

Lewis, DR, Dalgaard, BR and CM Boyer (1985), ‘Cost-effectiveness of computer-aided economics instruction’, *American Economic Review*, **75**, 91–96.

Light, A. (1999), ‘High school employment, high school curriculum and post-school wages’, *Economics of Education Review*, **18**, 291–309.

Mane, F. (1999), ‘Trends in the payoff to academic and occupation-specific skills: the short and medium run returns to academic and vocational high school courses for non-college-bound students’, *Economics of Education Review*, **18**, 417–438.

Mayer, SE and PE Peterson. (1999). ‘The costs and benefits of school reform.’ In Mayer SE and PE Peterson (eds.) *Earning and Learning. How Schools Matter*. Brookings/Russell Sage Foundation: Washington, D.C.

McMahon, W.W. (2000), *Education and Development: Measuring the Social Benefits*, Oxford: Oxford University Press.

Prais, S.J. (1996), ‘Class-size and learning: the Tennessee experiment – what follows?’, *Oxford Review of Education*, **22**, 399–414.

Ragan, J.F., Warren, J.T. and B. Bratsberg (1999), ‘How similar are pay structures in similar departments of economics’, *Economics of Education Review*, **18**, 347–360.

Russell, T.L. (1999). *The No Significant Difference Phenomenon.* North Carolina State University: North Carolina.

Schiefelbein, E., Wolff, L. and P. Schiefelbein (1999), ‘Cost-effectiveness of primary education policies in Latin America: a survey of expert opinion’, *UNESCO Bulletin*, **49**, 53–76.

Shadish, W.R. and C.K. Haddock. 1994. Combining estimates of effect size. In H Cooper and LV Hedges. (ed.) *The Handbook of Research Synthesis*. Russell Sage: New York.

Slavin, R.E., N.A. Madden, L.J. Dolan and B.A. Wasik (1996), *Every Child, Every School*, California: Corwin Press Inc.

Teddlie, C. and D. Reynolds (2000), The International Handbook on School Effectiveness Research. London: Falmer.

Thomas, H.R. (1990), Education, Costs and Performance: A Cost-effectiveness Analysis, London: Cassell.

Udharvelyi, I.S., Colditz, G.A., Rai, A.M. and A.M. Epstein. (1992). Cost-effectiveness and cost-benefit in the medical literature – are the methods being used correctly? *Annals of Internal Medicine*, **116**, 238-244.

Van der Drift, K.D.J.M. (1980), ‘Cost-effectiveness of audiovisual media in higher education’, *Instructional Science,* **9**, 355–364.

Zietz, J. and H.H. Cochran (1997), ‘Containing cost without sacrificing achievement: some evidence from college-level economics classes’, *Journal of Education Finance*, **23**, 177–192.

Table 1.1: Weighted Effect Size Relative to Lectures For Imparting Information

	Weighted Effect Size Relative to Lectures		
	Mean	S.D.	N
Treatment mode:			
Other modes	0.0970	0.0653	15
Independent study	-0.0379	0.0445	17
Discussion and enquiry modes	-0.0382	0.0504	18
Personalized systems of instruction	-0.8736	0.0658	4
All treatment modes:			
Mean effect size	-0.1563	0.0271	54
Effect size per hour	-0.0002	$3.8*10^{-5}$	54

Table 1.2: Total Per Participant Incremental Costs Relative to Lecture Provision

	Total Per Participant Incremental Costs Relative to Lecture Provision		
Treatment mode:	Mean	S.D.	N
Treatment mode:			
Other modes	1127.69	1768.36	15
Independent study	-411.57	968.43	17
Discussion and enquiry modes	-387.74	702.06	18
Personalized systems of instruction	2449.51	658.02	4
All treatment modes:			
Total	52.38	1137.66	54

Figure 1.1

Cost-effectiveness Proportions

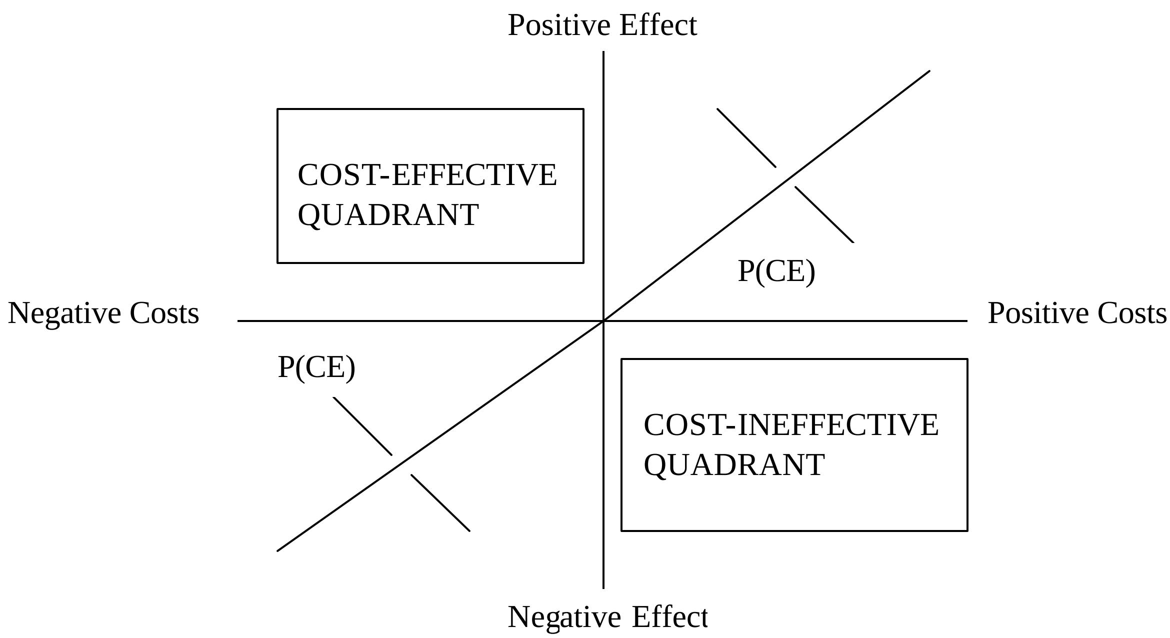
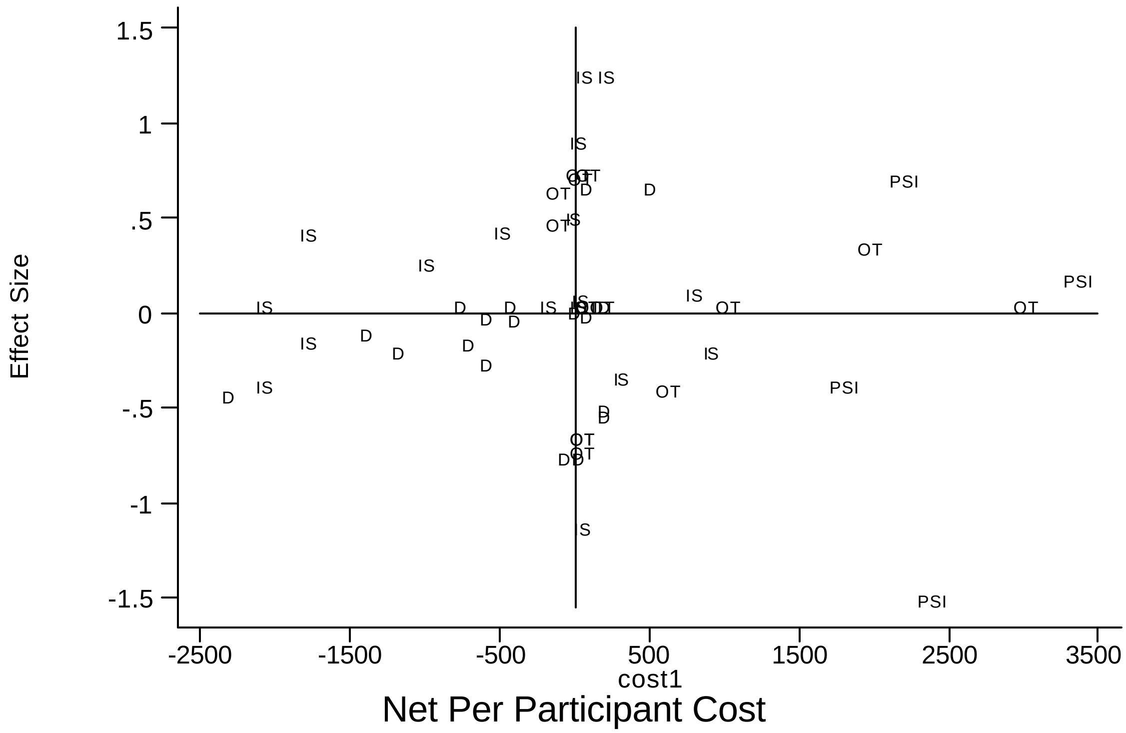


Figure 1.2

Cost-effectiveness Relative to Lectures by Mode



Key:

IS – Independent Study; D – Discussion/enquiry mode; PSE – Personalized System of Instruction; OT – Other.

Costs are in 2000 year dollars.

Figure 1.3

Funnel Plot