

# **Experiments in the Cost-Effective Uses of Technology in Teaching: Lessons from the Mellon Program So Far**

**Saul Fisher**

**The Andrew W. Mellon Foundation  
New York, New York USA**

**Thomas I. Nygren**

**The Andrew W. Mellon Foundation  
Princeton, New Jersey USA**

## **Introduction: In search of empirical data**

What do we really know about whether instructional technology in higher education works or is desirable? While the endeavors of instructional technologists are much-studied, most of these studies focus on such topics as the advancement of particular learning tasks, the attitudinal resonance of instructors and students with the use of such technologies, the effectiveness of various media in transmitting data (independent of content), and the merits of different strategies for streamlining and dovetailing technologies. What we learn, in short, from the research on instructional technology is whether the technology helps subjects meet fine-grained cognitive tasks, is appreciated, can be further optimized, and/or might be deployed efficiently.

We suggest, however, that extremely few studies have actually looked at this most pressing set of workability and desirability issues: what are the instructional merits of instructional technology? Where such issues have been addressed, they have not been addressed in anything like a rigorous

manner.<sup>1</sup> One such rigorous brand of study would look at whether using technology to enhance teaching at the university level is cost-effective. Literally speaking, this is the question as to what effectiveness, in this case, in teaching, one purchases for a given cost.

This sort of question—so little recognized, let alone studied—is just the kind of challenge that the Andrew W. Mellon Foundation is committed to addressing, and for which we happily have the resources to pursue. The long-standing goals of the Mellon Foundation include, broadly, the support of higher education, and a strong specialization of Mellon funding has historically been in support of academic research on higher education itself. It is fitting, then, that the Mellon Foundation has undertaken an initiative to support a series of studies that take a hard look at the promises of cost-savings and pedagogic effectiveness of university instruction with technology.

---

<sup>1</sup> A common complaint about studies that measure learning with technology is that those measures do not hold up to close scrutiny; see Phipps & Merisotis (1999). The literature asserting that there are no significant differences in outcomes across media is collected in Russell (2000).

The Mellon Foundation's program, "Cost-Effective Uses of Technology in Teaching" (CEUTT), supports the assessment of instructional technology as used in higher education, through experiments that gauge its cost and pedagogic effectiveness.<sup>2</sup> The hope is that these experiments will provide evidence to suggest whether such technology can be deployed without incurring greater institutional costs than can be recouped by later savings. The common failure to promote efficiencies through the use of technology represents a lost opportunity to help cut and lower the rate of rising costs in academe, as elsewhere. The Mellon CEUTT program seeks to help determine if, and how, there are ways that instructional technologies may be used to make good on such opportunities.<sup>3</sup> To measure cost-effectiveness, investigators are asked to measure the costs and pedagogic outcomes of teaching with and without the technology under study.<sup>4</sup> Accordingly, assessment in the CEUTT projects is generally conducted by teams of experts in pedagogic assessment, cost measurement, and instructional technology. In the optimal case, a CEUTT experiment compares two teaching situations—with and without the given technological intervention—

that are otherwise relevantly similar to the greatest degree possible.<sup>5</sup>

As a research program, CEUTT looks at the use of instructional technology with an unprejudiced eye. There is much enthusiasm for teaching with technology in certain corners of the academic community, and one may legitimately wonder about the prospects of conducting impartial studies on the use of such technologies when the enthusiasts are the investigators. On the other hand, a cynicism about the possibilities for teaching with technology is not likely any more helpful. With a truly unjaundiced eye, CEUTT projects should assign some prior probability of cost-effectiveness to the teaching situations using technology, as a working hypothesis and without presuming cost-effectiveness or the lack thereof.

The mechanics of CEUTT experiments is as follows. A given technology is introduced as an intervention in an instructional context, such as a course, major, or degree program. A control group is set up, and a method is outlined for the selection of students, in a way that is maximally random and satisfies fairness constraints and ethical ground rules. This last matter might seem to be a small problem, but there are substantive worries about simply assigning students to a class delivered in

---

<sup>2</sup> A fuller description of the CEUTT program can be found at the Foundation's website at <http://www.mellon.org/cutt.html>.

<sup>3</sup> The CEUTT program does not attempt to assess distance education *per se*. The focus is rather on illuminating how traditional higher education institutions may use instructional technology. Further, distance education does not lend itself to this sort of experimental design.

<sup>4</sup> For an overview of cost-effectiveness studies for policy and research ends, see Levin (1983).

---

<sup>5</sup> When the CEUTT program started in late 1996, there was an additional focus on supporting development of those new technologies that would be suitable for assessment. Over time, however, that focus has waned because such development is widespread with or without Mellon assistance—either with institutional funding or the support of other parties. This frees up the Mellon program to concentrate on pure experimental projects that focus on assessment and measurement.

one particular medium or another. They might actually want to take the other version of the course! However, allowing this to transpire poses the problem of self-selection. Fortunately, there are any number of ways around this problem, generally resulting from there almost always being a scarcity of resources—as with respect to classroom seats or server capacity.

One aspect of the experimental hypothesis is to say how the application of the technology might be expected to enhance teaching, and so augment learning outcomes. A second aspect of the hypothesis would be to outline how costs might be reduced or slowed in growth: Would the number of students served be increased? Would faculty time be reduced and freed up for alternative uses? Would students be able to cover more material in a shorter time? Would budgeted resources not have to be deployed? Would costs savings increase with class size?

To gauge whether these hypotheses are correct, investigators seek a longitudinal picture—a view that starts with baseline data (to gauge performance before the technology was ever deployed) and ends with comparative data drawn from the traditional and technology-enhanced classroom. The pedagogic outcomes are most typically measured by course outcomes, but a range of other instruments are available, for capturing results of various sorts. These instruments include pre-test and post-test comparisons, tests for retention across semesters, and gateway exams to measure rates of progress. Any of these instruments can be referenced to the norm of performance by other students, or to the criteria of a fixed mark of achievement. The cost measurements

are typically captured by taking apart the constituent activities in teaching a class (with and without the technological intervention), costing out the goods and services entailed, then reconstituting the whole activity in terms of the dollars expended.

By putting these two figures together—pedagogic outcomes and costs—we can generate cost-effectiveness ratios. Such ratios tell us what degree of change in pedagogic outcomes come at what cost. We foresee a range of cost-effectiveness results that reflect a spectrum of successes and failures. Sometimes it will be cheaper to teach through traditional means, other times with the use of technologies; sometimes it will be more effective pedagogically with the technologies, other times, without. The critical question is why particular results are attained—in each case, and—we hope to establish—more generally. After collecting sufficient data on the various projects, we expect to be able to plot the results and determine the sorts of circumstances under which instructional technology turns out to be cost-effective.

It is reasonable to assume that some experiments will raise or lower one side of the cost-effectiveness ratio but not the other, or keep one side unchanged though not the other, and that this will happen to varying degrees across types of instructional technology, units of instruction, disciplines, and institutions. The important task, in our view, is to find out how and why outcomes are maximized at minimal cost (or at maximal savings)—that is, if indeed the two sides of the ratio might be optimized at once. What would those optimal states look like? Perhaps we would only come to hope for minimal pedagogic

gains at maximal savings, or maximal gains at minimum increased costs, or yet some other scenario. At the same time we also expect to pick up data on the relative cost-effectiveness in each data set—each experiment—concerning the value of that particular technological intervention under the specific circumstances. This experiment-specific data will be of great value to the individual institutions, given the same local conditions (or, minimally, local culture) in which instructional technologies are deployed. This data will also be of interest to other schools—as case studies—the lessons from uses of instructional technology at one institution may have great precedent value for like institutions.

### **Review of selected projects**

To date, the CEUTT program has made twelve grants and we expect another eight to ten this year. The projects are studying diverse technologies at a wide spectrum of institutions, in courses of varying sizes, and in a multitude of disciplines, ranging from astronomy to writing. A number of the projects address some common issues, including instructional contexts (large introductory classes or laboratories), technologies (web-based instructional modules), and pedagogical features (self-paced instruction, interactivity, and simulation). Since most of these are multi-year projects, it is too early to report hard cost-effectiveness data. However, we can report some preliminary news from a couple of ongoing projects.

Thus, at Michigan State University (MSU), archival and experimental studies are measuring the costs and pedagogic effectiveness of courses

taught in the natural and social sciences at MSU and Texas A&M University (TAMU) with the use of an application that personalizes problem sets, quizzes, and exams, and records students' participation and performance. The investigators hypothesize that efficiencies may be achieved through increases in the quantity and quality of feedback to students, and reduced time in generating such feedback. They also are exploring whether the software raises or lowers classroom anonymity and students' conceptual learning.

In an MSU calculus-based introductory physics course taught with network technology this Fall to 480 students, investigators have measured an eleven percent increase in performance over students in past (traditional format) courses. In particular, seventy percent of students attained grades over 2.5 (on a 4.0 scale) in the technology-enhanced course, whereas only fifty-nine percent attained the same grades in the standard course taught earlier (controlling for difficulty of exams). These figures are not overly dramatic but they suggest that there is no loss in performance, and that there may be stronger performance at the bottom of the grade distribution. In addition, for the same course performance on the higher end of the grade distribution also increased.<sup>6</sup>

The MSU/TAMU investigators have reported a growth of costs due to the preparation and development of educational content accompanying the development of new technologies. As instructors move to technology-enhanced environments, the sheer

---

<sup>6</sup> Thirty-two percent achieved grades higher than 3.5, as against eighteen percent in the traditional course. Data courtesy of Ed Kashy (personal communication, February 2, 2000).

amount of teaching and testing material increases because of a new need for randomized variation which is introduced with automated exercises. This growth is not matched in traditional courses because they lack such variation in performance assessment. This does not appear to reflect a significant difference between the curriculum, though one might question whether the pedagogic approach has been altered.

Finally, the MSU/TAMU project investigators report a marked reduction in the number of faculty needed to teach a psychology statistics course (PSY 203) to 200 students at TAMU. Previously, some five faculty taught the course; with the use of the self-paced instructional software, only one faculty member is needed. The number of graduate assistants has climbed from five to six, but their services are needed for only ¼ time, as opposed to full-time or ½ time.

At Drexel University, an entire online Master of Science in Information Sciences (MSIS) program is being assessed in terms of costs and pedagogic effectiveness. For three years, Drexel has offered an online MSIS program and collected data on its use and effectiveness. Investigators are examining data from courses taught in the online and traditional classroom-based programs, and on this basis are testing the hypothesis that the online courses are more cost-effective than the traditional ones.

In two comparisons of final grades—one holding the course constant, and the other holding the instructor constant—investigators have found insignificant differences between students in the online and traditional versions of courses

in the MSIS program.<sup>7</sup> The Drexel investigators also compared completion rates for individual courses in the program, and found that those rates also do not differ significantly between cohorts for the online and traditional courses.

Some of the other projects we have supported have not yet reported results but promise a range of outcomes for different technologies, distinctive contexts, and varying scales of intervention.

For example, at the University of Pennsylvania, web-based pre-laboratory instructional applications are being studied in five engineering courses. Investigators are gauging whether these applications may reduce (a) valuable time spent in class on preparing for laboratory work, and (b) costly laboratory mishaps. This study is almost complete and the initial results seem to suggest potential cost savings.

At Dartmouth College, investigators are looking at a calculus course where a commercially-available software is being used to more fully integrate case studies, facilitate practice of calculus skills, and automate management of the course. The hypothesis is that this technology can help cut costs of instruction while sustaining improvements in teaching and learning associated with the case studies approach. Investigators are gauging the cost-effectiveness of teaching calculus with the technology-enhanced curriculum, as compared with both the case studies curriculum without

---

<sup>7</sup> In the first case,  $\chi^2 = 0.228$ ,  $p = 0.63$ ,  $n = 131$ ; in the second case,  $\chi^2 = 1.49$ ,  $p = 0.22$ ,  $n = 211$ . Data courtesy of Greg Hislop (personal communication, January 30, 2000).

technological enhancement and the traditional curriculum.

At the Georgia Institute of Technology, researchers have developed an instructional technology that attempts to mimic in an online environment those elements of a traditional classroom that can promote collaborative learning. Called “CoWeb”, the software provides an easy-to-use forum for dialogue, contribution of classwork, and feedback or elaboration from other class participants. CoWeb allows students to instantly develop and edit webpages, and thereby easily build and enhance collaborative websites. Investigators are examining the use of CoWebs in composition and computer science courses. They are testing whether this technology may produce efficiencies by promoting collaborative work among students (thereby increasing learning) at lower costs than in the traditional counterparts of those courses.

At Rutgers, investigators are examining whether eight courses (in subjects including biology, geography, communications, Spanish, and engineering) taught with web-based technologies may lower costs by substituting graduate students or junior faculty for senior faculty in tasks not critical to course design or delivery. Further, they are investigating whether the frequency of large lectures can be reduced through the use of elements such as self-directed lessons, quizzes, simulated labs (for biology), and 3D visualization applications (for engineering). They are also testing the assumption that these courses may improve pedagogic outcomes by providing more interactive and collaborative learning.

Finally, at the University of California, Davis, investigators are testing the sustainability and scalability of using online lectures to offset the ineffectiveness of large lecture courses in ten courses in the natural sciences, including agriculture, biology, computer science, food science, physics, and oenology. Investigators are evaluating students’ knowledge and skills in traditional and web-enhanced versions of the courses, and measuring the capacity of the web-enhanced courses to reduce costs by off-loading some of the instructors’ tasks to the computer and eliminating the need for large lecture halls.

### **What have we learned so far?**

While we have not yet gathered much in the way of hard data, our work with the CEUTT projects has produced at least these lessons regarding method, measurement, and assessment:

1. The timeframe for CEUTT experiments is generally a short period in higher education terms. This partly reflects a feasibility constraint typical to social science and education studies experiments. But this short timeframe also helps guard against other factors that might otherwise creep into the experimental or quasi-experimental set-up if experiments are extended for too long a period. For example, would costing or pedagogic measures need to account for further curricular development? As the curriculum changes, we lose reference to the baseline data. Also, what would happen to costing data if the technology becomes antiquated? In the current regime of rapid change, this is an eminently plausible possibility after as little as three years.

2. There is a paucity of relevant literature concerning measurement of costs or pedagogic outcomes with the use of instructional technology, so it is all the more important that these studies are done carefully. This gap in the literature has provided us with an opportunity to help by furthering the development of satisfactory measurement instruments. Our support here has focused on the costs side. In 1999, the Foundation funded work by the Technology, Learning, and Teaching Group of the AAHE to support the enhancement of a guidebook to assist colleges and universities in costing the development and deployment of instructional technologies.<sup>8</sup> And this year, we have supported work by the Western Cooperative for Educational Telecommunications to advance their Technology Costing Methodology project, to assist universities in projecting the costs and savings of developing and deploying technology-mediated courses across campuses. This tool will help judge the implications of large-scale adoption of instructional technologies.<sup>9</sup>

3. What literature does exist in the field is not univocal regarding the best ways to measure costs and pedagogic outcomes. Hence the CEUTT projects must be inventive and creative, and either synthesize or judge among the varying approaches to costs and outcomes. The successes and failures of the ongoing research may help address some knotty questions concerning measurement in this field. Other

foundational issues may be resolved by careful consideration of the alternatives.

So, for example, there are several different approaches to costing, including the bottom-up varieties of process-based costing and activities-based costing (ABC), and the more traditional, top-down varieties of macro-costing.<sup>10</sup> In this last, probably more familiar variety, costing is divided in a large institution by apportioning responsibilities for expendable or expended funds to cost centers in the organization, as a matter of the strategic and budgeting priorities of the institution. In the bottom-up variety, costs are analyzed by conducting a detailed analysis and costing of individual processes or activities.

The literature in the field generally comes down in favor of ABC but this largely reflects untested assumptions about the ways universities work and so should have their costing done. Some universities do not operate on this model, though, and it is plausible that data generated through ABC is less easily compared with standard numbers from university business offices because courses are not typically costed in this way. Given different approaches and very different assumptions, there may be considerable extra difficulties in extracting the requisite costing data from the institution's records. If this is right, then ABC could be more trouble than it is otherwise worth.

Another major issue in the costing sphere is what should be counted as a legitimate cost. Thus, it may seem fair game to count the costs of using physical

---

<sup>8</sup> For information, see

<http://www.tltgroup.org/programs/fcai.html>.

<sup>9</sup> For information on the predecessor BRIDGE model developed at CSU, see

[http://www.calstate.edu/special\\_projects/mediated\\_instr/Bridge/index.html](http://www.calstate.edu/special_projects/mediated_instr/Bridge/index.html).

---

<sup>10</sup> For an overview of the various costing methods pursued in higher education contexts, see Jenny (1996).

space, employing instructional and support staff, capital expenditures on traditional equipment and computers, and equipment depreciation. But what about counting costs of developing the technologies used to teach or the curriculum? Our thinking begins with the notion that, while the primary focus of the cost assessment model should be tracking the costs of implementing and deploying the instructional technologies, development costs *are* significant.

One key to grasping the extent to which development costs should count is understanding the role of the technology developed: Is the technology a device or tool for delivering instruction? Is the technology a curricular delivery medium? Does the technology provide a “space” for learning?

Consider a traditional space for learning: the classroom. The use of a classroom is a direct cost of teaching a given course, and this cost is a function of the development or construction of the classroom building and related infrastructure. Hence one should count some portion of the costs of designing, building, loading, or maintaining web-based curricular components as part of the cost of course delivery, just as one would count some portion of classroom usage time (along the lines of paying per square foot multiplied by time used, or amortizing the cost of the physical infrastructure). In the high-tech case, one would measure some fraction of the total technology infrastructure costs, per individual class and per course. If one could imagine that the total costs of designing, building, and maintaining the technology were amortized over  $n$  courses, then one would have a straightforward way of factoring in those costs of delivery.

Curricular-development costs (that is, costs of developing case studies, homework, exam problems, or class notes) are another matter. Some fraction of these creation and presentation costs should be attached to the course each time it is taught—that is, the costs should be amortized—but deciding exactly how much is *not* a function of when the curriculum is initially created. Compare the traditional case: If I teach a course for the first time, I need to prepare my lecture notes extensively, perhaps over the duration of a semester or summer prior to teaching. The next time out, although I ought to have significantly less work to do, I still want to count some portion of my earlier work towards the costs of teaching the course the  $n^{\text{th}}$  time. The costing scheme should be indifferent to whether the course is taught for the first time.

For experimental studies in the CEUTT model, curricular development costs should be measured in both settings (that is, with and without the technological intervention) to preserve the controlled nature of the experiment.

The literature concerning measurement of pedagogic outcomes is richer than in the costing field but leaves many questions unanswered. These questions concern, for example, the separation of program effect from media effect—are the outcomes measured the result of changes in curriculum, pedagogy, or the medium?<sup>11</sup> Are these factors non-

---

<sup>11</sup> The Dartmouth CEUTT study is a good illustration of this question. It would have been difficult to disentangle media and program effects if their Mathematics Department had simultaneously adopted the new case studies approach and the web-based medium; instead these occurred in stages with appropriate control groups at each stage.

separable, as some scholars maintain?<sup>12</sup> Would we want to separate them, even if we could? Our view is that they must be separable for the CEUTT projects to contribute to warranted judgments about cost-effectiveness.

Other questions center on the relative value of outcomes assessment instruments across media. Does the same exam measure the same learning in traditional and technology-enhanced classes? One reason to think it might not is that different skills may be enhanced by distinctive media, and an exam that captures content-wise outcomes would not help us identify the relevant differences. If however, one could pick out all such problems in advance, then it would be possible to design the appropriate instruments.

The answers to all of these questions may not be forthcoming, but progress can only be made through rigorous measurement and assessment in carefully-conducted studies. It is our hope that the Mellon CEUTT program will contribute to addressing these matters and advancing research on instructional technology as considered for its instructional merits and corresponding utility for institutions of higher education.

*Saul Fisher ([sf@mellon.org](mailto:sf@mellon.org)) and Thomas I. Nygren ([tin@mellon.org](mailto:tin@mellon.org)) are on the program staff of the Andrew W. Mellon Foundation, where they administer the Cost-Effective Uses of Technology in Teaching program.*

## References

- Clark, Richard E., "Media Will Never Influence Learning", Educational Technology Research & Development (42) 2 (1994): 21-29.
- Jenny, Hans, Cost Accounting In Higher Education: Simplified Macro- and Micro-Costing Techniques. Washington, DC: NACUBO, 1996.
- Levin, Henry M., Cost-Effectiveness: A Primer. Newbury Park, CA: SAGE Publications, Inc., 1983.
- Phipps, R. & Merisotis, J., What's the Difference? A Review of Contemporary Research on the Effectiveness of Distance Learning in Higher Education. Washington, DC: American Federation of Teachers, 1999.
- Russell, Thomas L., The "No Significant Difference Phenomenon", <http://teleeducation.nb.ca/nosignificantdifference>. (February, 2000).

---

<sup>12</sup> For example, see Clark (1994).