

Evaluating (and Improving) Benefits of Educational Uses of Technology

A paper prepared for WCET through the Technology Costing Methodology
Project Initiative funded by the Fund for Improvement of Post Secondary
Education (FIPSE)

Written By:

Stephen C. Ehrmann, Ph.D., The TLT Group

December 2002

Evaluating (and Improving) Benefits of Educational Uses of Technology

In education, we sometimes study costs and we sometimes study benefits. But rarely do we study both. That's one reason why people are fearful of cost studies: they assume that the cheaper alternative will be favored over the more expensive one, because no one will know whether the more expensive alternative also has better outcomes.

This failure to study benefits and costs simultaneously is not a coincidence. It's difficult to assess benefits. Imagine that we want to study the costs *and* benefits of two types of activity (a course, major, service, or institution-wide educational use of technology) in order to decide which of the two is better. We'll refer to these competing activities as Program A and Program B. This might be a 'before and after' comparison, a comparison of two competing pilot programs, or a comparison of a real activity with a hypothetical alternative, for example.

Let's simplify the problem a bit by assuming that the educational benefits of interest are who can learn, what they learn, and the consequences of those outcomes. Here are a few of the barriers to studying such outcomes while also studying costs:

- Outcomes are sometimes more difficult to measure than are costs. How would you measure the impact of a course on a student's character? Or on the economic welfare of the region? Because some benefits are difficult to assess, analysts often surrender and measure *outputs* (for example, how many students completed the program?) rather than *benefits* (for example, how were students and the community affected in the short and long term by the fact that the student took the program?).
- A study designed to measure benefits may focus on a 'chunk' of education whose costs are unusually difficult to measure. Suppose, for example, that the key benefit is the employability of the graduate. Employability is influenced by many courses taught by different departments, as well as a variety of extra-curricular experiences. What did the set of courses and extra-curricular experiences cost? How much of those costs should be allocated to this particular benefit?
- It would be wonderful if we could reduce benefits and costs to the same quantities and calculate ratios. "Does Program A produce benefits that are worth at least 10% more than its costs? Does Program B have an even better ratio of benefits to costs?" But what's the dollar value of a 5% increase in test scores? It's difficult or impossible to translate into dollar terms such outcomes as improvements in skill or extension of minority access to education.

Even though studying benefits while studying costs is difficult, it's not impossible and it's certainly important. This chapter will explore three key questions that you would need to answer in order to design such a study.

1. Are the program's outcomes intended to be the same for all its beneficiaries? If not, how can you assess them?
2. To help design assessment procedures, how can we be more specific than merely saying that the technology is meant to cause "better educational outcomes?"
3. What kinds of data about benefits might help the people running the program to improve those benefits (paralleling the way that activity based cost data ought to be able to help policy makers control costs)?

1. Are Benefits Intended to be the Same for all Beneficiaries?

What's a typical example of the kind of outcome goal that ought to be measured? "All students should learn to think critically (though perhaps to different degrees of skill)." "All students should get jobs (perhaps at different salaries)." In other words, the goals assume that everyone is supposed to benefit in the same ways. If that were true, it would certainly make things simpler to measure – the analyst could devise one test of achievement of benefit (e.g., a test of critical thinking skill) and apply it to all the beneficiaries.

But what if some students are gaining in critical thinking while others are mainly improving their creativity and still others are gaining in interpersonal skills?

As those examples indicate, there are two ways to look at almost any educational program. One perspective focuses on program benefits that are the same for everyone (“uniform impacts”) while the other perspective focuses on benefits that are qualitatively different and somewhat unpredictable for each learner (“unique uses”) (Balestri, Ehrmann, et al., 1986; Ehrmann and Zúñiga, 1997, 2002). This section of the chapter explains these complementary perspectives on education. The following section will use these ideas to suggest ways to assess specific types of benefits.

A. Uniform Impacts

To some degree, all students in an educational program are supposed to learn the same things. As shown in Figure 1, such learning by two people can be represented by two parallel arrows. The length of each person’s arrow represents the amount of growth during (and sometimes after) the program. Students usually enter a program with differing levels of knowledge, grow to differing degrees, and leave with differing levels of achievement. The uniform impact perspective assumes that the desired direction of growth is the same for all students.

In an English course, for example, uniform impact assessment might measure student understanding of subject-verb agreement, or skill in writing a 5 paragraph essay, or even love of the novels of Jane Austen. The analyst picks one or more such dimensions of learning and then assesses all learners using the same test(s). I’ve labeled this perspective “uniform impact” because it assumes that the purpose of the program is to benefit all learners in the same, predesigned way.

B. Unique Uses

However, that same English course (or other educational activity) can also be assessed by asking how each learner benefited the most, *no matter what that benefit might have been*. I’ve termed this perspective “unique uses” because it assumes that each student is a user of the program and that, as unique human beings, learners each make somewhat different and somewhat unpredictable uses of the opportunities that the program provides.

In that English course, for example, one student may fall in love with poetry, while another gains clarity in persuasive writing, and a third falls in love with literature, and a fourth doesn’t benefit much at all. (See Figure 2)

Faculty members cope with this kind of diversity all the time. An instructor may give three students each an “A” but award the “A” for a different reason in each case. The only common denominator is some form of excellence or major growth that relates to the general aims of the course. There are multiple possibilities for growth and it’s likely that different students will grow in different directions.

Notice that uniform impact methods tend to miss a lot when benefits are better described in unique uses terms. In that English class for example, imagine that the instructor had decided to grade all students only on poetry skills. One student would pass and the others would fail. Or imagine that the instructor tested all students on poetry, persuasive writing, and love of literature, and only passed students who did well on all three tests: everyone would fail the course. Meanwhile, an instructor using a unique uses approach (seeking excellence in at least one dimension of learning) would pass three of the four students.

Uniform impact and unique uses are both valid, and usually are both valid for the same program. The challenge for the analyst is to make sure that the assessment approaches are in tune with the program’s goals and performance. If, for example, the program’s goals are strongly “unique uses” then it is inappropriate to employ only “uniform impact” measures, and vice versa.

How can unique uses benefits be assessed? Most unique uses assessments follow these steps:

1. Decide which students to assess. All of them? A random sample? A stratified random sample?

2. Assess the students one at a time. Ask the student what the most important benefit(s) of the program have been for him or her. (At this point, the respondent's statement should be treated as a hypothesis, not a proven fact.) This hypothesis about benefits can also be created or fine-tuned by asking the instructor(s), peers, or job supervisors about the program's benefits for that student.
3. Gather data bearing on this hypothesis. If the student said that the program helped her get a job, what data might help you decide whether to believe the assertion? (For example, did the student really get a job? If the student said that certain skills learned in the program were important in getting the job, did the interviewer notice those skills?) If appropriate, assess the benefit for the student (for example, if the benefit is a skill, assess how skilled the student is).
4. If appropriate, quantify the benefit for that student. Panels of expert judges are sometimes useful for this purpose. Their expertise may come from their experience with programs of this type. (This is exactly what teachers do when they grade essays.)
5. Identify patterns of benefits. Was each student completely unique? Or, more likely, did certain types of students seem to benefit in similar ways? These findings about patterns of benefit may suggest ways in which the program can be improved. For example, suppose program faculty consider "learning how to learn" to be only a minor goal of the program. But 50% of their graduates report that "learning how to learn" was the single most important benefit of taking the program. In that case, the faculty might want to put more resources into "learning how to learn" in the future.
6. Synthesize data from the sample of students in order to evaluate the program's success.

2. Additional Defining Questions about Benefits

Here are some additional questions to ask yourself before you begin assessing benefits.

Outcomes or Value-Added? When studying benefits, are you interested in outcomes (the state of things after the student completes the program) or in value-added (how much did their math understanding improve from the beginning of the course to the end)? Outcomes can often be improved simply by recruiting more skilled incoming students, while value-added is more a result of the education.

When is "after"? Imagine two programs about literature: A and B. Program A teaches a thousand facts about novels that can be easily memorized but that are quickly forgotten soon after taking the final exam. In contrast, Program B teaches students to love novels so that they continue reading and rereading books after the course ends. Program B also encourages students to join or organize book clubs so that they can talk with friends about the books they've been reading. Program B's students finish with less factual knowledge than students from Program A but, over the years, Program B graduates become increasingly knowledgeable about literature. An exam taken immediately after the completion of the two programs might show higher scores for graduates of Program A. But in another exam, given three months later, Program B's students might outscore Program A's. Two years later, the advantage of Program A over Program B might be even larger. There are many factors to consider in deciding when to assess benefits. The purpose of the program is one of those considerations.

Same Outcomes, or Just Similar? When comparing learning outcomes of Programs A and B, ask whether the two programs are trying to teach exactly the same things. If they are, comparing benefits is easier: use the same assessment measure for both programs. That's the assumption that many people make about assessment: the most fair and appropriate approach is to use the same test of outcomes on the two competing programs.

But that equivalence of goals is rare, especially when technologies are used differently. Instead the two programs usually have goals that only overlap, as shown in Figure 3.

Imagine that Program A is taught mainly via lecture in a classroom. The competition, Program B, uses videotapes of that faculty member's lectures supported by an online seminar that is led by an adjunct staff member. Goals distinctive to Program A include benefits of face-to-face contact with a tenured faculty member. Goals distinctive to Program B might include benefits of greater student freedom to explore topics

of individual interest, greater in-depth exploration of certain topics in the online seminar, and learning how to collaborate online with other students. A study of benefits that only attended to the common goals (learning of course content, for example) would miss some of the major reasons for choosing one program over the other. In cases such as these it's important to assess all the important goals, not just those that are common to the competing programs.

3. Categories of Benefit and How to Assess Them

There are many categories of benefit from technology use for education, including:

- A. Enrollment and attrition (access to education)
- B. Better outcomes on traditional goals (teaching-learning effectiveness)
- C. New outcomes not previously sought or emphasized (e.g., computer-dependent aspects of disciplines such as geographic information systems in geography)
- D. Variety of offerings available to each learner
- E. Consequences of A, B, and C for the graduate (e.g., employment)
- F. Consequences of A, B, and C for the community in its economic, social, spiritual, and political life.
- G. Consequences of gains in personal and program efficiency (e.g., writing more because it's easier to use a word processor than a typewriter)
- H. Cost savings and revenue increases
- I. Helping the institution attract and retain students and staff who expect a certain degree of technology access.
- J. Helping the institution attract and retain support from outside constituencies who expect to see a certain level of technological infrastructure.

This chapter will focus on methods for analyzing benefits A, B, C and D. The rest of this volume focuses mainly on benefit H: cost saving and revenue gains.

A. Access benefits

Some programs are designed to produce gains in access to education: people who couldn't otherwise have taken courses of this type; people who can now take more courses; people who would have been less likely to pass such courses.

The uniform impact perspective usually invites attention to changes in total enrollment and retention either for all learners (total enrollment) or a particular target group (e.g., students of color). To assess changes in enrollment obviously requires counting students (not as easy as it sounds) and, sometimes, getting data to indicate why they are enrolled. For example, evaluators of distance learning programs need to know not only how many students are enrolled but also how many of those course enrollments would have occurred even without the distance learning program.

The unique uses perspective raises the question of whether particular types of students are especially aided or impeded by program features. For example, do online programs tend to attract and retain students who are more comfortable in that environment than in a face-to-face class?

It's important to look at these unique uses issues in enrollment and retention. Historically, changes in educational structures have opened access for some groups while restricting access for others (Ehrmann, 1999a). The analyst and the policy maker need to deal with whether the net change is positive, whether the groups who benefit especially need that benefit, and whether the groups that are impeded are groups that have been excluded by past arrangements as well.

B. Better Outcomes on Traditional Goals

In this situation, the goals of the two competing programs are the same.

In a uniform impact assessment, it's appropriate to use objective tests of student performance students from Program A and B. A high degree of skill is often needed to design objective tests, but only a low amount

of skill is needed to “grade” the results: how much time did the student take to finish the task? Did the project designed by the engineering student actually function? How many questions were answered correctly?

One sign that a unique uses perspective is important for assessment is that there is more than one way to define “successful learning.” Then a high degree of expertise is usually needed to assess and grade student work, e.g., evaluating an essay or term paper, judging a student project.

C New Outcomes, Better Outcomes?

Computers are often used in order to change the goals of instruction: a new course of study in e-business or computer music; education in how to solve problems in a virtual team, an increased emphasis on complex problem solving and abstract thinking in a course where computers can now handle the skills that once required memorization of rote problem-solving methods. So part of the value comes from outcomes that are unique to one program or the other. This brings us back to the challenge of comparing programs whose goals are at least somewhat different (figure 3) or even wholly different.

In these cases, program A and program B use different projects and tests to assess student learning. Even if we discover that students in program A scored 5 points higher on test A than students in program B did on test B, that tells us nothing about which program is more valuable. What about giving students in both programs a test that includes everything in both program A and B? Testing students on something they weren’t taught often leads to rebellion.

There are at least two feasible ways to assess learning outcomes in programs with different goals.

Criterion-based assessment: It is sometimes possible to assess learning against a standard. Program A is teaching pilots to fly airplanes while program B is teaching students to ride bicycles. Program A’s students also learn to fly, while program B teaches only half its students to ride a bicycle without falling over. In that sense Program A is more successful than Program B, even though different tests have been used.

But that kind of comparison doesn’t deal with the value of teaching people to be pilots versus bicycle riders, and that’s a tough question. But suppose advocates of Program A and Program B could agree on a panel of expert judges to assess their programs. Those judges would be given materials describing the programs’ goals and teaching methods, the tests and projects used to assess student learning, and the results of the assessments (test scores, student projects). Using these materials, the judges could then compare the two programs. For example, suppose a disciplinary association in graphic arts was considering two ways of teaching, one of which was more technology-intensive than the other. A panel of employers and graduate school representatives might examine data about entering students, the curricula, tests, and artwork from seniors. The panel would then report on which Program they preferred, and why.

D. Variety of Offerings Available to Learners

Education is being transformed by our uses of technology (e.g., Ehrmann, 1999a). One benefit of that change is the variety of offerings, learning resources, experts and peers that are potentially available to each learner. How might the analyst assess the value of this variety – both what’s offered, and what’s actually used?

The uniform impact perspective treats all learners and potential learners as equal. For example, in comparing Program A and B, the analyst might ask how many sources of information are used by students doing research papers. In comparing a virtual university to a campus-based institution, the analyst might compare the ways and places where faculty members were educated: does the virtual institution offer a more varied set of teachers than the campus?

The unique uses perspective focuses on the different experiences of each learner. It tends to direct attention toward the ways in which different types of students exploit the available resources. Perhaps a unique uses evaluation would conclude that Virtual University A fostered a greater variety of student learning, due to its

flexibility and ability to reach out for resources than did Campus B, whose students learned more in lock step, using similar academic resources for similar purposes.

4. Assessing Activities

The previous section focused on four categories of outcomes and how (using uniform impact and unique uses methods) each might be assessed.

But assessing outcomes alone doesn't tell us much about how to *improve* those outcomes (e.g., Ehrmann, 1999b) We need, at minimum, to look at activities, also: what people are actually doing in order to produce those outcomes. For example, knowing that mathematics scores are higher in Program A than in Program B doesn't tell us how to improve math scores unless we also know whether and how students learned math in each program. That's as difficult as finding out how people spend their time in cost studies: what students are supposed to do in programs is not always the same as what they really do do. But it's people's actual activities that determine educational outcomes, not pedagogical theories.

So, if one purpose of a benefits study is to guide future action, the study must look not only at outcomes but how people actually used the technology to behave differently in program A than in program B. For example, if program A was spending money on an advanced e-mail system, did faculty use it to communicate more frequently with students? If so, is there evidence linking that change in faculty-student contact to better learning outcomes?

The study can go even deeper in looking for data to understand and improve benefits: *why* did faculty and students choose to use the advanced e-mail system as they did. Why did others fail to use it at all? If, for example, some students didn't use the system because they didn't know how, a modest investment in technical support might improve use of the system, faculty-student contact, and learning outcomes. If other students didn't use the system because they thought the faculty member didn't want to be bothered, the faculty member could take steps to correct that impression, which would also ultimately help improve learning outcomes.

Years of research indicate that improvements in activities such as faculty-student interaction, student-student collaboration, time on task, and active learning usually lead to gains in benefits. So some studies treat the changes in those activities as the benefits of interest. The Flashlight Program (<http://www.tltgroup.org/programs/flashlight.html>) has developed survey and interview questions (the Current Student Inventory and the Faculty Inventory) to help carry out such studies.

5. Summary

It's not surprising that cost studies often ignore benefits: there are many reasons why benefits are difficult to study at the same time as costs. But failing to analyze benefits creates the risk that the cheaper program option will automatically be considered better.

Before designing the particular instruments for studying benefits, one needs to consider some challenging questions first:

- a) Is the program mainly trying to attain the same benefits for all learners (uniform impacts)? Or is the program also designed to help each learner make unique use of its opportunities? Most college and university programs have both goals, and each set of outcomes needs to be assessed differently. In particular, when studying unique uses, one needs to assess each student in the sample separately and then afterward synthesize these assessments in order to evaluate the program.
- b) Is the study going to consider educational value-added (students at the end of the program contrasted with students at the beginning) or only outcomes? If value-added is to be evaluated, then some kind of pre-test is necessary.
- c) Is the study going to measure benefits as the program is concluding (e.g., final examination), and/or some time after the program ends (e.g., at a time when students would actually be making

- use of what they learned in the program)? During this waiting time, some knowledge and skill will diminish while other educational outcomes may improve (if the student continues to use them).
- d) Different categories of benefits (e.g., access outcomes; traditional learning outcomes; technology-related learning outcomes; variety of offerings) need to be assessed differently. The uniform impact/unique uses distinction also suggests alternative ways of assessing each of these types of outcome.
 - e) If one of the goals of the study is to improve program effectiveness, it's important to gather data on what people are actually doing in the program ("activities") as well as about outcomes. It's even more useful to gather data on why people are behaving as they are. For example, study factors affecting their choices about whether and how to use technology; those insights can be used to foster more appropriate and successful use of technology to improve learning outcomes.

6. References

Ehrmann, Stephen C. (1999a) "Access and/or Quality: Redefining Choices in the Third Revolution," *Educom Review*, September, pp.24-27, 50-51. On the Web at <http://www.tltgroup.org/resources/or%20quality.htm>

Ehrmann, Stephen C. (1999b), "What Outcomes Assessment Misses," in *Architecture for Change: Information as Foundation*. Washington, DC: American Association for Higher Education. On the Web at <http://www.tltgroup.org/programs/outcomes.html>

About the Author

Stephen C. Ehrmann is Director of The TLT Group's Flashlight Program and co-author of The Flashlight Cost Analysis Handbook. The Teaching, Learning, and Technology Group is a non-profit whose mission is to help institutions improve education by making more successful use of technology. Dr. Ehrmann has written four books and dozens of articles over the last 25 years on technology and innovation in education.

Figures

