

Measuring Success:

Evaluation Strategies for Distance Education

*Only by evaluating the effectiveness of
DE programs can we justify their use and
continue to develop their quality*

By **Barbara Lockee, Mike Moore, and John Burton**

Institutions that offer distance-delivered courses and programs share a compelling need to determine their quality. The many factors involved in the success of distance offerings makes the creation of a comprehensive evaluation plan a complex and daunting task. Unfortunately, what may seem the most logical approach to determining effectiveness is often theoretically unsound. For example, comparing student achievement between distance and face-to-face courses may seem a simple solution, yet the design is flawed for a number of reasons. However, theoretically sound approaches do exist for determining the effectiveness of learning systems, along with many different methods for obtaining answers to the relevant questions.

This article describes the different stages and types of evaluation for distance-delivered experiences. We also offer some guidance for developing an evaluation plan for distance education programs.

First, we need to distinguish among research, evaluation, and assessment. In cases where the learning system under development involves testing a component of a learning theory (or indeed is designed exclusively to allow testing of important variables), then the concentration is on research.

In contrast, designers of learning activities and artifacts generally focus on what works in a very applied sense. Such efforts seek to discern whether a system does what it was designed to do in an effective and efficient manner. These practitioners are evaluating their efforts.

Assessment refers to measuring learner performance (either before or after a teaching intervention, or both). Thus, assessment can be part of an evaluation, but assessment and evaluation aren't synonymous. More

importantly, comparing assessment scores from different learning systems is a serious, but common, error.¹⁻³

The Challenge of Evaluating DE Systems

Distance education systems consist of a complex array of infrastructures and personnel. A few of the factors to consider are instructional, technological, implementation, and organizational issues. Additionally, while these factors can be isolated and itemized, by no means are they independent of each other. As in any system, the separate components must work together effectively so that the whole DE system can operate holistically.

When DE delivery technologies break down, distance learners cannot engage in the planned instructional event. Without institutional policies that provide for online support services, distance learners can find it difficult or impossible to get assistance with matters necessary for their basic participation in a higher education program. Thus, a comprehensive review of DE efforts must not only scrutinize the individual system components, but also attempt to get a clear picture of how the parts work together as a whole to create positive outcomes (learning, satisfaction, matriculation, and so on).

Perhaps the most feasible manner in which to appraise the effectiveness of typically complex DE efforts is to do so incrementally. Fortunately, the tradition of educational evaluation has established stages and data collection approaches that lend themselves to the cause.

Evaluation generally breaks down into two broad categories: formative and summative. Formative evaluation serves to improve products, programs, and learning activities by providing information during planning and

development. Data collected during the design and development process provides information to the designers and developers about what works and what doesn't, early enough to improve the system while it remains malleable.

Summative evaluation determines if the products, programs, and learning activities, usually in the aggregate, worked in terms of the need addressed or system goal. Simply, formative and summative evaluations differ in terms of the audience for the information collected, the time in the development cycle when the information is collected, and the intention behind the data collection. Summative evaluation is information provided to audiences external to the design and development team (such as funding agencies, clients, or accreditation agencies) about how the entire package works in a real setting. Although this information might be used to suggest changes, additions, segmentations, and such, it's more likely that the information will be used to make fiscal and policy decisions to use, or continue funding, a learning system.

Formative Evaluation of DE

Formative evaluation is the best way to ensure quality in a unit or course before its release. Robert Stake⁴ likened the two stages of evaluation to making soup: when the chef tastes the soup, it's formative; when the diners (or a food critic) taste the soup, it's summative. In other words, formative evaluation mimics internal quality control, and summative evaluation reflects how well the final object works in the real world.

Formative Issues

The development of distance courses requires consideration of

formative evaluation issues, largely falling into the two primary categories of instructional design issues (such as teaching strategy choices and assessment methods) and interface design issues (Web site navigation, aesthetics, and so forth). As course development takes place, the designers make choices with regard to each of these categories. Formative evaluation can help identify choices that might not be the most effective, giving the developer an opportunity to revise the course before implementation.

Regarding instructional design issues, evaluators seek answers to the primary question of learning effectiveness. Did students learn what the goals and objectives intended? If not, why? Was the instruction well written? Were the objectives clearly stated and measurable? Were appropriate instructional strategies chosen? Was there enough practice and feedback? Were examples provided? Did assessment methods correlate with instructional content and approaches? If these questions can be addressed within the formative evaluation stage, then corrective measures can produce more effective learning experiences for distance students.

Even an instructionally sound, online course can fail to produce learning outcomes if students encounter a poorly designed Web site. In evaluating the interface design of a Web-based course, a few simple questions can provide insight into the strengths and weaknesses of a site's look and feel. Was the Web site easy to navigate? Was it aesthetically pleasing, as well as legible? Did each page in the site download easily? If special plug-ins were needed, were links provided to acquire them?

Also, consideration of learners with special needs should be addressed at this stage. If graphics or images were used, were alternative ways provided for sight-impaired learners to get the intended information? If course information was presented using audio, could hearing-impaired learners access transcriptions? Was informa-



Even an instructionally sound, online course can fail to produce learning outcomes if students encounter a poorly designed Web site.

tion clearly available to learners with disabilities on where to get assistance if needed?

While it would be ideal to conduct a formative evaluation on an entire distance program, it's probably not feasible. Still, one could gain predictive evidence of a program's success by having the program's stakeholders, such as potential students and relevant professional organizations, evaluate its curriculum design. For example, a graduate program in engineering could seek input from relevant engineering associations or a program advisory board comprised of potential employers, requesting information regarding necessary skills and knowledge that the targeted distance program should produce.

Formative Stages and Methods

In general, formative evaluation can involve several different stages. Although these stages vary across authors of evaluation models, we will use the following six stages for purposes of illustration.

Design review encompasses the use of other designers or others familiar with the design strategy and underpinning

learning theory to evaluate the elements of the design itself. This review begins very early in the design process and continues at least until the design is "set." This stage aims to improve the project's instructional design.

Expert review usually involves a review by content experts, grade-level teachers, test creators, and others who can judge the appropriateness of the content and the learning activities for the target learners. This step avoids developing a product that simply will not work with the intended learner population or ultimately won't "sell" for classroom use because of unacceptable content, content bias, missing content, or approach to the content.

One-on-one review involves a member of the design/development team meeting with a representative of the target population to go through a piece of the project, talking concretely about it. The piece might be a mockup or prototype to test such things as design of Web pages or placement of interactive buttons. It could also be a script used to get a "feeling" for the suitability of a piece of video before expensive production begins. These data can provide the first information from real learners about everything from appeal to usability to workability. The review is crucial because it occurs when relatively little investment of time, money, and ego has occurred. Changes get much more difficult beyond this point because so much is committed to the design in progress that inertia can carry problems (even known flaws) forward once full-scale production begins.

Small group reviews try out fairly finished components in a setting that permits interactions among representatives of the target group. Data are often collected by survey instruments, group interviews, sample test performance, and the like.

Field trials try out the completed package with a sample or samples of target learners in the actual field setting for which the program was designed. Think of opening a play "off Broadway" — changes are possible,

Table 1**Summative Evaluation**

Categories of Evaluation Concerns	Potential Data
Program Inputs	<ul style="list-style-type: none"> ■ Budget information ■ Personnel information
Performance Outcomes	<ul style="list-style-type: none"> ■ Expected learning outcomes <ul style="list-style-type: none"> Knowledge Skills Attitudes
Attitude Outcomes	<ul style="list-style-type: none"> ■ Expected attitudes <ul style="list-style-type: none"> Interest Motivation Participation ■ Unexpected attitudes ■ Program or product design perception ■ Social interaction concerns
Programmatic Outcomes	<ul style="list-style-type: none"> ■ Market reach ■ Professional impacts (promotion, job change) ■ Faculty incentives and rewards ■ Faculty time ■ Organizational change
Implementation Concerns	<ul style="list-style-type: none"> ■ Technological concerns <ul style="list-style-type: none"> Stability Maintenance ■ Student support concerns ■ Faculty concerns <ul style="list-style-type: none"> Faculty preparedness Involvement in curriculum development Involvement in course development Professional development needs Incentives and rewards ■ Learner concerns <ul style="list-style-type: none"> Access to delivery system(s) Learner preparedness Communication/interaction with faculty Communication/interaction with peers ■ Organizational concerns <ul style="list-style-type: none"> Quality assurance Accreditation criteria

but not many of them and not without relatively large costs.

Ongoing reviews are required because learning systems cannot ever be considered complete — certainly not for very long, anyway. No matter how well evaluated the program, errors will

creep in. More importantly, content, production values and technologies, learning orientations, and more will change, making redesign always necessary. Reanalysis, redesign, and redevelopment all require additional information if they are to be done well.

Summative Evaluation of DE

A second approach to determining the success of a distance education course or program involves collecting evidence at the conclusion of the instructional event, a process called summative evaluation. Summative data can be collected from a variety of sources using an extensive menu of data collection methods.

The summative evaluation process breaks down into a few definitive steps. First, the general areas of concern must be delineated; from that, questions regarding those areas of focus can be developed. Experience and review of the distance education literature suggest that most categories of interest fall under the three primary headings of inputs (resources, personnel, and so on), outcomes (performance, attitude, and programmatic results), and implementation issues. Table 1 provides a tabular representation of the three general areas, their subcomponents, and some potential data that could be collected to address each area.

After determining specific evaluation questions, appropriate information gathering techniques can be chosen. Data is collected, analyzed, and reported to constituent audiences.⁵ The following sections address each general area of concern, proposing related questions and suggesting beneficial data collection strategies for each.

Course or Program Inputs

Of great interest to those responsible for funding distance course or program development is the identification of resources required to do so. Budgetary information can be helpful in terms of delineating monetary allocations for equipment, personnel costs for instructional development and distance support services, training and development costs, and so on. Combined with programmatic data, such as the number of students served, input data can demonstrate cost efficiency and sustainability to university-level personnel, as well as to

accrediting agencies interested in those factors.

Outcomes

Analysis of course or program outcomes — changes that occur as a result of the instructional experience — is another common evaluation target. This category encompasses performance, attitudes, and programmatic outcomes.

Performance Outcomes. Performance measurements generally represent the most common type of data collected for summative distance course and program evaluations. A posttest or project usually measures expected learning outcomes, with assessment items referenced to specific course objectives. Unfortunately, a popular strategy for assessing distance student performance is to compare the learning outcomes of distance students to those learners in the same class on campus. Such comparisons ignore the many factors that influence learning and falsely attribute success (or failure) to the distance delivery medium.^{6,7} Instead of comparing groups on and off campus, student achievement analysis can go to the heart of the issue (and be more effective) by determining whether or not distance students learn what the course is designed to teach. The Council of Regional Accrediting Commissions agrees, requesting that

... as a component of the institution's overall assessment activities, documented assessment of student achievement is conducted in each course and at the completion of the program, by comparing student performance to the intended learning outcomes.⁸

Attitudes. In addition to positive learning outcomes from a distance course or program, educational providers likely hope that their instructional efforts produce positive attitudes among participants about the courses, the distance experience, and the host institution. Therefore, learner attitudes constitute another

type of data generally collected for summative DE evaluations. Interest, motivation, and attitudes toward participating in the learning experience are useful measurements. Data on expected and unexpected attitudes of the learners could be collected through self-administered or interviewer-administered questionnaires, open-ended interviews, observations, and focus groups. Sample items might include, "I always knew where to go for help when I needed it during the course" or "It was always clear to me what I was supposed to be learning throughout this program." Measuring unexpected attitudinal effects can provide insightful information. Surveys could include items like "Compared to other online courses, this was one of my favorites." A positive response in this case is gratifying, but not necessarily expected.⁹

Attitudes about social interaction concerns within the course or program can help guide the design of the course. Evidence suggests that alleviating the common feeling of disconnectedness among distance learners is important for creating positive attitudes about the distance learning environment. Effective communication strategies within such courses can facilitate bridging this psychological distance.¹⁰

Data acquisition regarding student perceptions of social interactions can use the same tools as for the other attitudinal information types: self-administered or interviewer-administered questionnaires, open-ended interviews, observations, and focus groups. Such data can also be collected at any time, including at the conclusion of the program. Sample items might include "Being forced to work with a partner was good for me" or "I had to help other people in my group learn some of the skills for this course." Responses to these items can help inform instructors' communication decisions within the context of the distance course.

Programmatic Outcomes. Programmatic outcomes or effects are important to program administrators who

must demonstrate the impact of the program to others. Factors such as enrollment and attrition rates are basic types of information for program justification and continuation.¹¹ Another important impact to consider is the course or program's market reach. Related questions are "Is the program reaching its target audience?" and "Do enrollees represent non-targeted groups or regions?" An analysis of enrollment data by demographic and geographic information should provide insight to answer these questions. Also, focus groups within professional organizations can supply informative data regarding program reach.

Professional impacts resulting from the program also provide indicators of success. While distance program goals may not be related to this issue, if program participants receive a promotion or accept a new position based on their newly acquired skills and knowledge, this offers a definite indicator of program quality. Follow-up surveys or interviews from those who have completed the distance program can be an effective method for collecting such data.

Another factor affected by the development and implementation of distance courses or programs is the use of faculty time. Many critics have raised the issue of the time it takes to deliver a course online, given the increased amount of direct communication with students, plus the frequent increase in student numbers. Good questions to ask relate to the balance of workload and efficient use of time by faculty who teach at a distance. Is time spent on DE courses significantly detracting from research and scholarship? Are DE faculty designing efficient strategies to implement their courses? This valuable data can be collected both quantitatively through anonymous surveys and qualitatively through interviews.

Implementation Concerns

Perhaps the broadest area of summative evaluation concerns evaluating the implementation of distance

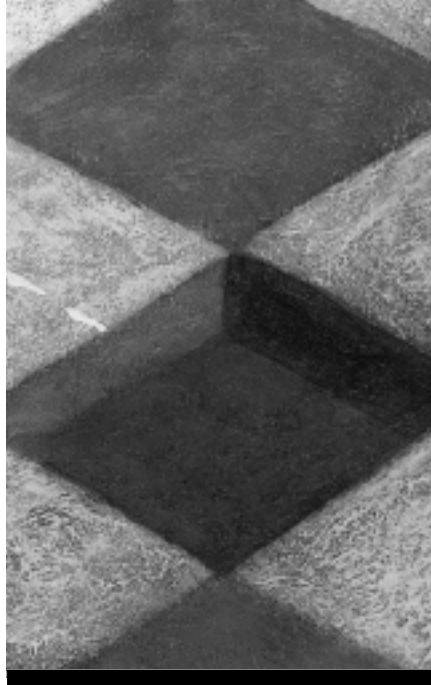
courses and programs. The process of DE has a variety of stakeholders, from students to faculty to support personnel to the host institution itself. Some implementation concerns are shared by all stakeholders, such as the reliability of the delivery technology and the accessibility and effectiveness of the student support services. Other concerns are specific to individual stakeholders. For example, distance learners must understand the distance environment and be prepared to engage in self-directed learning.¹² Also, distance learners should clearly understand faculty expectations and know who to contact for technological and instructional needs.

Regarding faculty concerns, their preparedness to teach in distance settings is important, while accessibility to appropriate professional development activities is essential. These factors, as well as incentives and rewards for teaching at a distance, are very real issues that faculty face, hence worthy of evaluation.

Finally, educational providers, such as institutions of higher education, are concerned with quality assurance. Are our distance courses and programs of strong quality and rigor? Do they meet our professional accreditation criteria? These questions can also be answered within a summative evaluation effort.

Sample Evaluation Scenario

To demonstrate the proposed procedures for evaluating a distance education course or program, we present our personal experiences with these strategies. For the past three years, Virginia Tech has offered a master's degree in instructional technology specifically for public school teachers in the state of Virginia, and now on a national level. We are three of the seven faculty who developed and implemented the original program, now on its third iteration. Through both formative and summative evaluation mechanisms, we have collected data to inform the program's continuous improvement.



Our analysis of learning outcomes relies on assessment of our distance learners' performance at two levels.

For the formative evaluation of our course and program development, we relied primarily on design review, expert review, and one-on-one review to determine the effectiveness of the Web courses we developed. Our time to production was typically short (a few months), so it wasn't feasible in most cases to conduct small-group reviews or field trials. Having instructional design expertise was an added advantage in our case, but we strongly recommend having online courses reviewed by colleagues and/or instructional support personnel before implementation. At this stage of course development we remedied many instructional and interface design issues that could have caused problems.

Additionally, we identified possible course management challenges in our formative evaluation efforts. For example, it became clear to us upon review of initial assignments in the first online course that no mechanism existed to receive and manage student work. (This preceded the advent of learning management systems like Blackboard or WebCT.) Since the student projects consisted of Web pages

that we wanted to access collectively and review online, we had to establish a server system and accounts for all 70 students. The resource intensiveness of that solution led us to seek a more efficient system for the management of our student Web sites. Fortunately, in a parallel effort, the university established a filebox server system for all students that we leveraged for our distance learners. Dealing with this issue during the formative stage of evaluation led to a workable solution before it became a crisis.

Our summative evaluation efforts have focused on the three general categories presented earlier: program inputs, program outcomes, and implementation issues. Incidentally, since the program continues to be offered to new groups of students, what's typically seen as summative information is truly formative, as results and lessons learned have been used for the program's continuing evolution.

First, we tracked our distance program inputs through our budget system, a process that clearly identifies the resources necessary for implementing the program. Our records are used primarily for administrative purposes; however, the information demonstrates the program's cost-efficiency to a variety of constituents.

Our analysis of learning outcomes relies on assessment of our distance learners' performance at two levels. Each module has assignments, many of which are projects graded not only at the module level, but that also contributed to an electronic portfolio of student work. At the end of the program, the capstone event is a faculty committee review of each student's electronic portfolio to determine if they have acquired the necessary IT skills and knowledge as delineated by the professional standards of our field. Student grades from individual modules and portfolio review results provide helpful measures demonstrating that the program effectively produces the intended learning outcomes.

Attitudinal outcome data was collected at the end of each module and also upon completion of the degree.

Perhaps the most revealing information came from reflective statements written by students at the end of the IT program. While the program's intended learning outcomes clearly were being achieved, since students demonstrated the targeted skills, the program's faculty found it especially rewarding to know that our learners felt that their experience and efforts were worthwhile and professionally beneficial.

Programmatic effects were also collected at the end of our first cohort cycle. Data such as number of students served, their geographic locations, program attrition rates, and professional impacts of their experience, help define our reach and justify the program's continuation.

As we continue to compile information about our graduates, we have found an interesting (and unintended) effect of career transition among many of our participants. While some teachers have taken more technology-focused jobs within their own school systems, several have left education altogether for corporate IT positions. This trend (though possibly unpopular with school administrators) would have gone unnoticed without a concerted effort to collect follow-up programmatic data.

Finally, evaluation of the implementation of our online program occurred both during its execution and upon its completion. Documentation of implementation concerns was made on a weekly basis and summatively through faculty and student reflections on lessons learned. Implementation challenges often resulted in organizational change (a programmatic effect), as our efforts to leverage the flexibility of asynchronous courses tended to bump head-on into policies and procedures designed for synchronous, campus-based systems. The need to accommodate our learners, as well as those in other distance programs, led to the development of an online registration system, the ability to pay student fees by credit card, and a change in the minimum course load requirement (from three credit hours down to one).

Making a Plan

To create an evaluation plan that determines the success of a distance-delivered course or program, begin by defining what you mean by "success." Is it increased enrollments, broader geographic reach, or high retention rates? Or is it more academic in perspective, meaning outstanding learner performance, increased job potential of graduates, or higher satisfaction from current employers? Is it a distance education system that works smoothly and accommodates the needs of all stakeholders, including students, instructors, and support personnel? Success is most likely defined as a combina-

tion of these factors, prompting the need for an evaluation strategy that is comprehensive and carefully planned. We hope this review offers some insights and guidance to answering the question of quality in distance education. *e*

Endnotes

1. B. B. Lockee, J. K. Burton, and L. H. Cross, "No Comparison: Distance Education Finds a New Use for No Significant Difference," *Educational Technology Research and Development*, 47 (3) (1999), 33-42.
2. B. B. Lockee, D. M. Moore, and J. K. Burton, "Old Concerns with New Distance Education Research," *EDUCAUSE Quarterly*, 24 (2) (2001), 60-62.
3. B. B. Lockee et al., Framing the Questions of Distance Education Research," *Conference Proceedings of the Distance Education Conference 2000* (College Station, Tex.: Center for Distance Learning Research, 2000), 115-125.
4. R. Stake, *Evaluating Educational Programmes: The Need and the Response: A Collection of Resource Materials* (Paris: Organisation for Economic Cooperation and Development, 1976).
5. F. Stevens, F. Lawrenz, and L. Sharp, *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering, and Technology Education* (Arlington, Va.: National Science Foundation, 1997).
6. Lockee, Burton, and Cross.
7. Lockee, Moore, and Burton.
8. Council of Regional Accrediting Commissions, "Best Practices for Electronically Offered Degree and Certificate Programs" (2000), 12, accessed from <<http://www.wiche.edu/Telecom/Article1.htm>>, September 26, 2001.
9. Lockee et al.
10. L. Wolcott, "Distant, but Not Distanced," *TechTrends*, 41 (7) (1996), 23-27.
11. Lockee et al.
12. D. W. Hardy and M. Boaz, "Learner Development: Beyond the Technology," *Teaching and Learning at a Distance: What It Takes to Effectively Design, Deliver, and Evaluate Programs: New Directions For Teaching and Learning*, T. Cyr, ed. (San Francisco: Jossey-Bass, 1997), 41-48.

Barbara Lockee (lockeebb@vt.edu) is an associate professor of instructional technology at Virginia Tech in Blacksburg, Virginia. Mike Moore (moorem@vt.edu) and John Burton (jburt@vt.edu) are professors of instructional technology at Virginia Tech.