CHAPTER 3

The Functions of Evaluation

Objectives

After reading Chapter Three, you should be able to:

- identify the major functions of evaluation that should drive decision-making during the process of designing and implementing interactive learning systems;
- identify typical activities that are carried out to fulfill each of the major evaluation functions; and
- appreciate the importance of evaluation within the instructional design process.

Evaluation functions within the development process

If you are a designer of interactive learning systems, you probably employ a systematic approach to development, whether it is a classic instructional systems design (ISD) model (Briggs, Gustafson, & Tillman, 1991; Dick & Carey, 1990), a “rapid prototyping” approach (Tripp & Bichelmeyer, 1990), or some hybrid method (Piskurich, 2000). Although ISD models have come under attack recently (Gordon & Zemke, 2000), they remain important processes for developing education and training programs around the world. Despite new developments in automated course development tools and the widespread availability of reusable course components, predictions of the demise of ISD are probably premature (Zemke & Rossett, 2002).

Regardless of the method used, few people design interactive learning systems alone; you are likely to be a member of a team of instructional designers, content experts, programmers, videographers, graphic artists, project managers, evaluators, and others. In our experience, creating interactive learning systems has much in common with film production. Just as a team representing diverse skills and talents is required to produce motion pictures, few individuals can accomplish all the tasks involved in creating interactive learning systems by themselves.
Figure 3.1 illustrates a typical ISD model that might be used to develop an interactive learning system, especially one involving multimedia. The problem with this and many other representations of ISD is that they make the steps involved in creating an interactive learning system appear entirely too linear. In actuality, the process is much more iterative and much less mechanical than ISD models imply. In addition, creativity is as essential as the application of scientific principles within the process of designing interactive learning systems.

![Figure 3.1. Instructional design model for interactive learning systems (ILS).](image-url)
Perhaps the Walt Disney Corporation best captures the essence of producing interactive learning systems with its creative and patented term, “Imagineering!” One of the major reasons ISD is being questioned is the perception that ISD practitioners are not as creative as developers who follow less structured models (Gordon & Zemke, 2000). In addition, ISD is weakened by the failure of researchers in the instructional technology field to produce robust design principles (Clark & Estes, 1998). Despite the efforts of Gagné (1985), Merrill (2002) and others, there is no grand, unified science of instruction, and frankly we don’t think there ever will be. Human learning is simply too complex for complete prediction and control. Lacking comprehensive scientific rules, you and your design team must combine the limited scientific principles and incomplete learning theories that do exist with imagination and common sense to develop interactive learning systems that are effective, efficient, and responsive to learner needs.

Another problem with most graphical representations of the ISD process is that they make evaluation appear to be an activity that is conducted late in the development process when it is, or should be, the key function that literally informs or drives the other functions involved in the process. In Figure 3.2, we present an alternative representation of the relationships between major development functions (conceptualize, design, develop, implement, institutionalize, and re-conceptualize) and the major functions of evaluation (review, needs assessment, formative evaluation, effectiveness evaluation, impact evaluation, and maintenance evaluation).

![Figure 3.2. Relationships among evaluation and development functions.](image-url)
Admittedly, the representation in Figure 3.2 is also flawed in that few actual interactive learning projects will follow the idealized path that is represented in this model. Evaluation is rarely implemented to the degree that it should be, perhaps for fear that projects will falter or even come to a complete halt if draft deliverables are evaluated fully. Our illustration intentionally emphasizes the importance of evaluation. As described in Chapter One, our rationale for writing this book is to provide information and tools that will enable you to collect and utilize evaluation data as the basis for informed decision-making throughout the process of developing and implementing interactive learning systems. We don’t think the importance of evaluation can be overestimated, but we recognize that some may view this emphasis as impractical.

A case study: Macintosh Fundamentals

A case study from our own experience as members of a design team may provide a better portrayal of the typical process of how an interactive learning system comes to be developed and evaluated. When the Macintosh computer was first introduced by Apple Computer, Inc. in 1984, the interfaces for the operating system and the few software programs then available for the novel machine (e.g., MacWrite, MacPaint, and MacDraw) were so simple and intuitive that Apple promoted the idea that no training was necessary to use a Macintosh. However, within a few years, the complexity of the different Macintosh models that became available, accompanied by the proliferation of sophisticated peripheral devices and powerful software packages, changed the situation dramatically. In the face of these changes, the people marketing, selling, supporting, or just using Macintosh computers began to demand more and better training. In response to these needs in the United States, Apple established a network of education centers around the country where new employees, vendors, and users could experience hands-on training with the Macintosh. As Apple sales went up, the demand for basic training increased at the education centers. At the same time, the need for higher level courses was identified as even newer programs and peripherals for the Macintosh came on the scene.

In the face of unrealistic demands on the live trainers at the Apple Education Centers, the training development group at Apple headquarters in Cupertino, California began to explore the idea of developing an interactive learning system to deliver basic Macintosh training. Eventually, a large-scale training package called Macintosh Fundamentals was developed under the leadership of Jim Laffey, a multimedia pioneer at Apple Training Support (and now a professor at the University of Missouri). Although it may seem like ancient history to today’s e-
learning designers, the development of the award-winning *Macintosh Fundamentals* program provides a rare and noteworthy example of the integration of evaluation throughout the ISD process. The initial conceptualization of the interactive learning system included thorough reviews of existing interactive training materials as well as a comprehensive literature review. In addition, external evaluators were contracted to conduct a needs assessment at Apple Training Centers to identify the key goals and objectives that should be included in this first comprehensive interactive learning system for the Macintosh computer.

Actual design of the training system commenced once the specifications for the delivery platform were identified. The design and development phases of the process were informed by a sequence of alpha, beta, and field tests conducted both in-house at Apple headquarters in California and at three Apple Education Centers in different states. Alpha tests normally involve assemblage of a working version of the interactive learning system so that it can be tested in-house. During beta tests, a relatively complete version of the system is provided to all internal clients as well as to selected external clients. During field tests, well-tested versions of the training system are released to most customers with the understanding that changes may still be made in the program. Numerous other formative evaluation activities drove the development effort.

Although it took almost an entire year to complete, and cost nearly a million dollars, *Macintosh Fundamentals* was used by more than 10,000 trainees in the first two years of its use. During the first year of use, effectiveness evaluations were conducted by external evaluators at several training centers. After effectiveness was demonstrated, limited impact evaluations were conducted. Overall, these evaluations indicated that *Macintosh Fundamentals* was as effective as Apple’s leader-led training had been, yielded more consistent results, and had significant long-term impact on the knowledge, skills, and attitudes of the trainees who had used the interactive learning system. Ongoing maintenance evaluation activities eventually indicated the need for a major reconceptualization of introductory training for the Macintosh, and a new learning system was developed. The subsequent program did not have to be delivered at Apple Training Centers because it could be distributed via CD-ROM to trainees at their own local sites. (Apple now delivers much of its training over the Web.)

Of course, not everything about the development of *Macintosh Fundamentals* proceeded as smoothly as described above, nor was every decision made solely on the basis of evaluation data. Internal politics as well as design prejudices and other factors influenced some of the decisions made by the project managers and team members, but most of the critical decisions were informed by timely and accurate evaluation.
data. All six functions of evaluation that we have identified (review, needs assessment, formative evaluation, effectiveness evaluation, impact evaluation, and maintenance evaluation) were represented in this case study. The details about how these functions should be carried out during the development of interactive learning systems are presented in chapters five to ten of this book.

Defining the functions

A chapter is devoted to each of the six functions of evaluation identified in Figure 3.2. However, to give you an overview of those chapters, we define each of the functions in a bit more detail below by identifying its purposes and presenting a hypothetical example of its application.

Review

The function of review is most important during the initial conceptualization for an interactive learning system. There are two primary review activities that are presented in this book, the review of professional literature related to the project and the review of existing interactive learning systems. The overall purpose of the review function is to ensure that you and your team members are as well-informed as possible about the primary options for interactive learning related to a project during its earliest stages of conceptualization.

Suppose that you are the head of a small entrepreneurial company trying to identify a product line that will set your company apart from your competitors. With your background in elementary education, you believe there is a market for remedial multimedia programs for students ages 6-9 in the area of science education. There are several review strategies that you should undertake before preparing a business plan or funding proposal for the development of this product. You would want to review the science education literature with the following questions in mind:

- What national, state, or local standards or guidelines have been established for science education by groups such as the National Science Foundation or the Department of Education?
- What do critiques of existing educational software reveal about the quality and effectiveness of these products?
- What do projections indicate about the home school market for educational software?

If you discover that there are interactive products already on the market that claim to meet the needs you have identified, you will want to conduct a systematic review of these materials addressing questions such as:
• How well do the objectives of these products align with the national, state, or local standards or guidelines you identified during your literature review?
• What pedagogical strategies are employed in these materials?
• What evidence exists that these materials are effective at a reasonable price?

Chapter Five presents activities, procedures, and tools for carrying out systematic reviews of the literature and existing products.

Needs assessment

Needs assessment is one of the most widely underutilized evaluation functions in the process of developing interactive learning systems. In our experience, most projects are undertaken because of some political decision or in response to a request for proposals (RFP) that is uninformed by sufficient up-front analysis. A case in point is Western Governors University, an online university conceived in 1995 with great fanfare by then Governor Roy Romer of Colorado and Governor Mike Leavitt of Utah. Six years later, the virtual university still had more courses (developed at the cost of millions of dollars) than it had students. A cover story in The Industry Standard (Boynton, 2000) about Governor Leavitt describes the initiative:

But the jewel in Leavitt's crown – as well as the thorn in his side – is the virtual Western Governors University, which he and Colorado Governor Roy Romer concocted in 1995. The idea was for 13 Western states to collaborate on class offerings to ease overcrowding and minimize the expense of educating the children of the baby boomers – which in Utah’s case is estimated will add 130,000 students to an already overburdened school system by 2010. The university would also part with certain academic standards as criteria for a degree, in favor of competency in a specific field. Launched with lots of hype in 1998, the university has been a bust. The goal was to have 500 degree-seeking students and 7,000 students in corporate training programs by this year; it currently has only 200 degree-seeking students and hasn’t met its other enrollment targets, lacks accreditation and is running at a deficit. (p. 151)

There is little evidence that even a rudimentary needs assessment or market analysis was conducted before launching Western Governors University (Berg, 1998). There was and is, of course, plenty of hype promoting the future of online education and the end of schools and campuses as we know them (Jones, 2000; Perelman, 1992; Pittinsky, 2003), and therefore, the Western Governors and their backers may have
felt compelled to get to the market first with virtual degrees. In hindsight, a relatively small expenditure on needs analysis and other evaluation strategies would have been a good investment.

The primary activities carried out under the rubric of needs assessment are task analysis, job analysis, learner analysis, and answering the question, “Why is this project uniquely needed?” (Rossett, 1987, 1998). The overall purpose of needs assessment is to identify the critical needs that the proposed interactive learning system is intended to meet. The results of needs assessment activities are often stated in terms of goals and objectives. Content, delivery system, and authoring system specifications may also be derived from these activities.

Suppose that you work in the training department of a mining company where the number of accidents resulting in injury and death has increased steadily over a five-year period to an alarming level. Management has ordered your department head to develop an interactive multimedia kiosk that will be located at the entrance of each work site to remind workers of safety requirements and the procedures for reporting hazards. You have been appointed to lead the multimedia development team. There is considerable pressure on you to get these kiosks up and running within four months, but you insist on a thorough needs assessment first. You contract with an outside firm to conduct this needs assessment, and after two weeks of interviews, focus groups, observations, and document analysis, the needs assessment report reveals that workers are quite adept at reporting hazards in the workplace. The source of the injuries appears to be the introduction of new high-speed drilling equipment on which the workers have been inadequately trained. It is recommended that, instead of an interactive multimedia kiosk emphasizing the reporting of hazards, what is really needed is better training in the operation of the new drills. Management negotiates with the equipment manufacturer to provide this training, and the manufacturer subsequently awards your department a contract to produce the new training program.

Chapter Six presents activities, procedures, and tools for carrying out job analysis (Hartley, 1999), task analysis (Schraagen, 2000), and other components of needs assessment.

**Formative evaluation**

The function of formative evaluation is perhaps the most important one in the entire process of developing interactive learning systems. Budget constraints often mean that some evaluation activities must be abbreviated or even eliminated, but formative evaluation should be the least diminished because it often yields the greatest payoff of any function of evaluation. The overall purpose of formative evaluation is to provide
information to guide decisions about creating, debugging, and enhancing an interactive learning system at various stages of its development. Some of the primary activities carried out during formative evaluation include expert review, user observations, and usability testing (Flagg, 1990; Hix & Hartson, 1993; Shneiderman, 1997).

Suppose that you are a curriculum director at a state department of education. You have won a federal award to develop an interactive multimedia Web site that will train teachers to recognize and report signs of child abuse. You and your team have prepared a draft treatment and illustrated it with story-boards. Before further development, you conduct a series of focus groups in which you present the treatment to several groups for review, including teachers, school administrators, parents, and health professionals. You also present the treatment to a team of attorneys from the state attorney’s office. The reviews indicate that your treatment is generally acceptable, but that you have not sufficiently included opportunities to learn the legal responsibilities and liabilities that teachers have with respect to recognizing and reporting child abuse. Based on this review, the legal aspects of the treatment are clarified, and script development can begin.

Chapter Seven presents activities, procedures, and tools for carrying out observations, expert reviews, alpha and beta tests, usability studies, and other components of formative evaluation (George & Cowan, 1999).

**Effectiveness evaluation**

The function of effectiveness evaluation is to drive decisions about the marketing and implementation of an interactive learning system. The overall purpose of effectiveness evaluation is to determine whether the interactive learning system accomplishes its objectives within the immediate or short-term context of its implementation. Some of the primary effectiveness evaluation activities include field tests, observations, interviews, and performance assessment.

Suppose that you are a professor of educational technology at a large state university, sometimes serving as an evaluation consultant for the defense industry. A new virtual reality (VR) system has been developed by a military contractor to train soldiers in land navigation, and you have been commissioned to evaluate the effectiveness of this training. Traditionally, soldiers have been trained to navigate with compasses via a combination of leader-led classroom training and field exercises. The VR program puts soldiers into simulated battlefields where they must traverse mine fields and other hazards using a virtual compass.

The military command has requested an experimental study of the effectiveness of the VR program wherein a random sample of 50 soldiers...
will undergo the traditional training and another random sample of 50
will experience VR training. You request a delay to study the actual use
of the VR system in more detail, but the officers and contractors insist
that the study proceed immediately. After the training, all soldiers
complete a written examination and a performance test wherein they use
compasses to negotiate their way across several types of terrain on the
south end of the military base. The results for the written tests are
comparable for both groups, but the performance tests indicate that the
traditionally trained soldiers successfully navigated the course 63% of
the time whereas the VR group had a success rate of only 47%. These
results are extremely disappointing to both the military command and the
outside contractor. Convinced that the VR system would be effective,
some of the contractors even question the validity of your evaluation.

Fortunately, you conducted a series of observations during the training
implementation that help to explain the poor results. First, you observed
that the sergeants conducting the traditional training used slides of the
site where the performance tests were conducted to highlight specific
navigation strategies, and in doing so inadvertently familiarized their
trainees with the test course. Second, you discovered in focus group
interviews that 50% of the VR trainees experienced severe nausea while
using the VR system, and thus ended up closing their eyes during much
of the training to overcome vertigo. A subsequent analysis of the data
indicated that of the soldiers who did not suffer nausea, 78% were
successful in the field exercise. Your analysis of the training implementa-
tion problems indicated the need for an investigation of the physical
and psychological effects of the VR training on soldiers.

Chapter Eight presents activities, procedures, and tools for carrying out
implementation studies, quasi-experimental studies, participant observa-
tions, and other components of effectiveness evaluation.

Impact evaluation

The impact evaluation function is extremely challenging. Not surpris-
ingly, few interactive learning systems have been subjected to a rigorous
impact evaluation. The overall purpose of impact evaluation is to deter-
mine whether the knowledge, skills, and attitudes learned in the context
of instruction transfer to the intended context of use, e.g., the work place.
It is not unusual to find that interactive instructional programs that
possess immediate effectiveness fail to affect practice “on the job.” For
example, workers may be able to recite safety procedure rules such as
wearing safety goggles, but neglect to follow these rules when perform-
ing their jobs. Some of the primary activities carried out during impact
evaluations are document analysis, interviews, and observations.
Suppose that you work in the educational services department of a medical school where an innovative interactive multimedia program has been purchased to train senior medical students in the practice of cost-benefit analysis regarding the prescription of pharmaceuticals for flu and cold sufferers. Studies indicate that past graduates of the medical school frequently over-prescribe antibiotics and other drugs for patients at considerable added expense. The program is designed to train physicians to be more aware of the cost-benefit ratios involved in drug regimens, to consider less expensive, off-the-shelf remedies, and to recognize the long-term effects of antibiotics and other pharmaceuticals on their patients and on public health.

An earlier effectiveness evaluation revealed that the medical students accomplish the stated objectives of the multimedia program with a 97% achievement rate. However, you insist that an impact study be carried out to detect whether the program has any impact on the actual performance of the graduating physicians. Subsequently, an analysis of the prescription rates between the previous graduating class and the first class of medical students who experienced the cost-effectiveness training indicates no statistically significant differences in behavior. Interviews with the students in both classes reveal that although the second class was more aware of the cost-effectiveness issues, the pharmaceutical advice of the supervising physicians during their practicums outweighed the recommendations presented in the multimedia program.

Chapter Nine presents activities, procedures, and tools for carrying out observations, interviews, and other components of impact evaluation.

**Maintenance evaluation**

The function of maintenance evaluation is an increasingly important one in the world of interactive learning systems as more and more software is developed and distributed on wider and wider networks. The enormity of the problem can be sensed if one surfs the Web searching for interactive learning opportunities concerning HTML programming. There are numerous sites available, but their effectiveness and impact remain unexamined, and the currency of their content is often questionable. The overall purpose of maintenance evaluation is to examine the viability of an interactive learning system over time. Some of the primary activities carried out in the name of maintenance evaluation are document analysis, interviews, observations, and automated data collection.

Suppose that you are the Webmaster for an electronics corporation, responsible for developing and maintaining an “Intranet” that provides training and performance support for sales personnel. The Web site provides sales personnel with the latest marketing information about a
wide range of products, recommended sales strategies, and tools for calculating sales quotes for customers. The Web site has been operational for nine months, and statistics indicate a steadily declining rate of usage of the site by the sales staff. You decide to investigate, and via a series of telephone interviews with sales personnel, you find that a major missing component of the Web site is performance and price information about products sold by your corporation’s major competitors. Subsequently, you add information and tools to the site that arm your sales staff with critical competitive information. Not only do the number of hits on the Intranet increase over the next six months, but there is a rise in sales that management attributes to the enhanced Web site.

Chapter Ten presents activities, procedures, and tools for maintenance evaluation.

Activities, procedures, and tools

For each of the six major evaluation functions in our model, we have identified activities, procedures, and tools to assist you in carrying out the functions. For example, for the formative evaluation function, one activity that is frequently used is expert review, i.e., recruiting experts to review the content, instructional design, and other aspects of your interactive learning system as it is being developed. Expert review is also a frequently used strategy in the development of the user interface for interactive learning systems (Nielsen, 1993; Preece, 1994). When you ask experts to review the human-computer interface for your interactive program, you will usually want to structure this procedure by using a tool such as a user interface rating form. This form is one of the evaluation instruments that we introduce in Chapter Seven.

The tools we provide in this book should be regarded as templates which can be modified as needed. Electronic versions of these tools can be found on the official Web site for this book. The tools have been developed in common software applications such as word-processing programs (Microsoft Word) and spreadsheets (Microsoft Excel).

Figure 3.3 illustrates the relationships among evaluation functions, activities, procedures, and tools that are presented throughout the following chapters. For each function, we identify a set of critical activities that must be carried out to fulfill the function. For each activity, we describe the specific procedures recommended for carrying it out. Finally, for each procedure, we provide one or more tools to assist in following the procedure.
Summary

This chapter has presented an overview of the six functions of evaluation that we believe are necessary to drive the instructional design (ID) and implementation processes for interactive learning systems. There are, of course, other evaluation approaches that are not as relevant to the evaluation of interactive learning systems. We encourage you to explore the broader evaluation literature (Patton, 1997; Rossi & Freeman, 1999; Scriven, 1993; Shadish, Cook, & Leviton, 1991). We have also highlighted the importance of these evaluation functions within the ID process by presenting scenarios that illustrate the roles evaluation functions fulfill at various stages of the process. In the next chapter, we present guidelines for planning and managing evaluation efforts within the context of the design and use of interactive learning systems.

References


