CHAPTER 7

Formative Evaluation

Objectives

After reading Chapter Seven, you should be able to:

- identify decisions involved in formative evaluation of interactive learning systems;
- specify questions that should be answered before making these decisions about improving interactive learning systems;
- identify the information needed to answer these questions; and
- decide how to collect and report the required information so that an interactive learning system can be improved in a timely manner.

Why should you conduct formative evaluation?

The overall purpose of formative evaluation is to provide information to guide decisions about “debugging” or enhancing an interactive learning system at various stages of its development. As illustrated in Figure 7.1, different types of decisions must be made when you attempt to improve an interactive learning system, each of which is tied to one or more specific questions that can be addressed by formative evaluation activities, such as observations and beta tests.

As described in Chapter Three, formative evaluation is the essential “life-blood” of the instructional systems design (ISD) process. According to Flagg (1990), formative evaluation is “the systematic collection of information for the purpose of informing decisions to design and improve the product” (pp. 1-2). Virtually everything about an interactive learning system can be enhanced at some stage of its development. Sometimes all that is needed for improvement is a flash of creative insight, but more often than not you will need specific information to guide your program improvement decisions. This information can be collected in many different ways from a variety of different people, ranging from subject matter experts to members of the target user population for the interactive learning product.
<table>
<thead>
<tr>
<th>Decisions</th>
<th>Example Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should the interface be redesigned?</td>
<td>Is navigation clear to users?</td>
</tr>
<tr>
<td></td>
<td>Are the meanings of icons clear?</td>
</tr>
<tr>
<td></td>
<td>Do users get lost in navigating through the program?</td>
</tr>
<tr>
<td>Should the number and length of video segments be decreased?</td>
<td>Do users select to view video segments?</td>
</tr>
<tr>
<td></td>
<td>Do users use video replay options?</td>
</tr>
<tr>
<td></td>
<td>Do users rate video highly?</td>
</tr>
<tr>
<td>Should more practice opportunities be added?</td>
<td>Do users pass module quizzes?</td>
</tr>
<tr>
<td></td>
<td>Do users achieve mastery on unit tests?</td>
</tr>
<tr>
<td></td>
<td>Do users rate practice highly?</td>
</tr>
<tr>
<td>Should the program scope be expanded?</td>
<td>Are the materials coordinated with curricular guidelines?</td>
</tr>
<tr>
<td></td>
<td>Do content experts rate the program as comprehensive?</td>
</tr>
</tbody>
</table>

**Figure 7.1.** Typical decisions and questions in a formative evaluation.

Resistance to formative evaluation

Although practices analogous to formative evaluation are common in many fields, for example advertising, there is sometimes resistance to rigorous formative evaluation among both designers and sponsors of interactive learning systems. This may partially derive from an unfortunately tendency within the software industry to develop a new program without substantial formative testing, invest heavily in packaging and marketing the program, and then to use the feedback from the early adopters of the software to fix or improve the program before Version 2 is released. A vice-president of a very large international computer corporation once admonished one of us for promoting formative evalua-
tion of a computer-based education package, exclaiming, “Why should we pay for user testing when the first ten thousand people who buy our software will tell us how to improve it for free!”

While it is hard to argue with the “economics” of this corporate attitude, we believe that there is an ethical imperative to evaluate interactive learning systems while they are being developed. After all, interactive learning systems are intended to change people, to modify their knowledge, skills, and attitudes, and to influence them to behave differently. The risk of misguiding learners is too great, and therefore we view formative evaluation as a morally responsible activity.

In her valuable book, *Formative Evaluation for Educational Technologies*, Flagg (1990) identified six reasons why people resist formative evaluation:

- **Time** – In the rush to meet project deadlines, reducing or eliminating formative evaluation activities is perceived as an easy way to save time.
- **Money** – Most development budgets fail to provide sufficient funding for rigorous formative evaluation. (The authors of this book recommend at least a 10% allocation to formative evaluation.)
- **Human Nature** – Many developers are reluctant to subject their programs to potential criticism, especially from users they may view as uninformed or from experts they may view as threatening.
- **Unrealistic Expectations** – Although formative evaluation can provide information to guide decision-making, it cannot substitute for the expertise and creativity of a qualified developer. In short, you cannot just toss together a rough prototype of an interactive learning system, and expect formative evaluation to turn it into a winner.
- **Measurement Difficulties** – Although some aspects of formative evaluation are relatively easy to determine (e.g., investigating whether users think various parts of an interactive learning system are appealing), there is a lack of reliable, valid, and feasible methods of evaluating certain kinds of outcomes of interactive learning that a particular program may address, e.g., problem-solving.
- **Knowledge** – Formative evaluation expertise is not yet widely available within the interactive learning systems development industry or within academe. Most developers lack the skills to conduct systematic formative evaluation in an efficient and effective manner.

Investments in formative evaluation should result in an overall reduction in development and implementation costs over the lifespan of interactive product systems, and hence resistance to formative evaluation should decline as formative procedures become more routine. Even within the
software industry, there is an increased emphasis on usability testing and other formative practices (Nielsen, 2000; Preece, 1994), and we predict that this trend will continue.

**When should you conduct formative evaluation?**

It can be said that conducting formative evaluation should be like voting in mayoral elections in certain large cities, that is, you should evaluate “early and often.” Duby (1988) presents a sound rationale for early formative evaluation of instructional products such as educational television. The sooner formative evaluation is conducted during a development project, the more likely that substantive improvements will be made and costly errors avoided. Avoiding expensive mistakes is especially critical with some of the technical elements of interactive learning systems, such as video, which remains a particularly costly component of most products. Producing video per se is expensive enough, but when you add in the costs of compressing video for digital movies on a DVD or transmission via the WWW, you begin to spend “real” money!

A former student relayed a story that illustrates the importance of early formative evaluation. She was involved in the production of a CD-ROM multimedia program about basketball featuring a famous college coach. A crew was sent to the coach’s university to videotape him giving tips about playing the game. It was a sunny day and the crew decided to tape him on an outside court so that the light would be good. The coach was videotaped against a background of trees with the wind rustling the leaves. The video looked great, but when they took it into the digitizing software, the compression algorithms were so overtaxed in compressing the constant movement of the foliage that the coach’s face and lip movements were distorted. By the time this was discovered, it was too late and too expensive to re-tape the coach against a neutral background. A little formative evaluation of the video compression process up front would have saved the producers a lot of grief.

**What kinds of decisions can you anticipate?**

Figure 3.1 in Chapter Three presents a typical ISD process for interactive learning systems in terms of the stages, steps required in each stage, the team-members involved in each stage of the process, and a list of the interim products resulting from each stage. Each of the draft documents and interim products represents an opportunity for making important decisions about enhancing the effectiveness and efficiency of the final interactive learning system. Should you increase the difficulty of the
program’s objectives? Should test item formats be revised? Should the fonts used in different parts of a program be changed? Should more or less humor be incorporated into video scripts? These and other decisions will be faced by you and other members of the development team.

The impetus to make decisions about improving an interactive learning system will come from many directions. You may see a program developed by a competitor that inspires a new interface idea. Your clients may cut your budget, thereby requiring you to cut down on the more expensive elements of a program such as 3-D animation. The colors that looked great on your high-end development machine with millions of colors may look awful on the consumer level machines with reduced color graphic capabilities. These and many other factors will be signals that there is room (and often a need) for improvements in your prototype interactive learning system.

Of course, under the best circumstances, formative evaluation is not something that is initiated when there is a crisis such as a budget cut. Instead, it is a professional practice integral to the overall instructional development process. What’s more, a formative evaluation perspective is no less important for those involved in implementing interactive learning systems. The bottom line is that all of us are human and our first efforts to create or do anything are bound to be somewhat flawed. Formative evaluation is the key to detecting and reducing these flaws and eventually attaining the high quality we all desire.

**What questions should be answered before making decisions?**

Each possible decision will inspire many different types of questions about improving your interactive learning system. Do learners understand what their options are at any given moment? Does the program maintain the learners’ attention? Do they accomplish the objectives of the program? Is it feasible to implement the program as designed? It is too late to wait until you have completed an interactive learning system to ask these questions. Instead, they must be addressed throughout the development of the interactive product. As noted above, the earlier these questions can be asked and enhancements made based upon the responses, the more efficient your overall development effort will be.

There are no universal criteria established for formative evaluation of interactive learning systems, but some of the most commonly considered factors are functionality (Does the product work as designed?), usability (Can the intended learners actually use the program?), appeal (Do they like it?), and effectiveness (Did they learn anything?). Different criteria entail many different types of questions. For example, usability implies
criteria that can be broken down into small issues such as the quality of the user interface. User interface can be further divided into factors such as navigation, mapping, aesthetics, and control. Finally, a factor like navigation can be examined via several different questions: How do users navigate through the interactive learning system? How does their navigation relate to the underlying pedagogy of the program? What parts of the program are underutilized? Where would users like to go, but don’t seem to know how? Answering these and other questions provides the development team with the information they need to enhance the navigational aspects of the program and ultimately improve its usability.

How should formative evaluation be conducted?

The key to sound formative evaluation is to collect data systematically at critical stages of the interactive learning system’s development and to utilize the findings of each formative evaluation strategy as much as your personnel, time, and financial resources allow. Numerous articles and books have been written about formative evaluation (Beyer, 1995; Flagg, 1990; Kinzie, 1991; Maslowski & Visscher, 1999; Tessmer, 1994), and there are comprehensive volumes covering individual aspects of formative evaluation (Branaghan, 2001; Hix & Hartson, 1993). As described below, we recommend the following formative evaluation activities as essential within the context of most development projects:

- expert review,
- user review,
- usability testing, and
- alpha, beta, and field tests of prototype program.

Two main classes of usability evaluation methods can be differentiated (Ziegler & Burmester, 1995). One class focuses on users of a particular product and aims to determine usability by studying users while they interact with a product. This approach is referred to as user review (or sometimes user testing). The other method is designed to identify specific human factors issues of a product and is referred to as usability testing. No matter how much analysis has been done in designing a product, experience has shown that there will be problems that only appear when the design is tested with users, i.e., people as much like those who will eventually interact with the interactive learning systems as genuine learners. The learner’s experience of an instructional product’s usability is an important indicator of its quality.

Another level of formative evaluation involves consideration of user acceptance. In user acceptance testing, it is recommended that users test
not just the product, but all parts of the package that the users will receive, such as training, written procedures, forms, manuals, computer-based training, and online help (McManus & Hammond, 1991, p. 101). This integrated approach ensures that there is no mismatch between the different components and highlights the users’ perspective of the whole product rather than a number of the parts. For this testing, a prediction is needed of the organizational and task changes that will occur as a result of the introduction of the new product. For example, if employees are expected to complete interactive training programs on their own computers at home rather than in the workplace, then usability testing should be conducted in homes rather than in corporate usability labs (Mitropoulos-Rundus & Muszak, 2001). Once the new product has been implemented, it is important to follow up with effectiveness evaluation in order to understand the actual learning process, usability issues, and use of the product by novices and experts in realistic work or education contexts.

**Expert review**

Expert review may be the most frequently used formative evaluation strategy. It is important to remember that there are several different kinds of “experts,” and that each type of expert can add unique kinds of information to the review and enhancement process. Content or subject matter experts can help you improve the scope, sequence, and accuracy of an interactive program’s content. Instructional experts can assist by critiquing the potential effectiveness of the pedagogical dimensions of an interactive program. Graphic designers can suggest how to enhance the aesthetics of a program’s look and feel. Teaching and training experts can help you anticipate the logistical requirements for successful implementation of an interactive learning system in schools or businesses.

With respect to formative evaluation, an expert is anyone with specialized knowledge that is relevant to the design of your interactive learning system. Experts can provide different perspectives on the critical aspects of your program, e.g., its accuracy, completeness, user-friendliness, motivational strategies, aesthetics, instructional validity, effectiveness, efficiency, and feasibility. You should utilize both internal and external experts to the degree that your resources allow.

It is often useful to structure an expert’s review so that you are assured of getting the types and depth of information you desire. Figure 7.2 presents an expert review form to guide instructional design experts with experience in interactive multimedia when they critique a prototype interactive instructional program.
Review: Dr. Ima Knowitall  

Due Date: June 10

Please circle your rating and write comments on each aspect of the interactive multimedia (IMM) package.

1 represents the lowest and most negative impression on the scale, 3 represents an adequate impression, and 5 represents the highest and most positive impression. Choose N/A if the item is not appropriate or not applicable to this package. Use additional sheets to write comments.

NA=Not applicable  1=Strongly disagree   2=Disagree  3=Neither agree/nor disagree  4=Agree  5=Strongly agree

AREA 1 – INSTRUCTIONAL DESIGN REVIEW
1. This IMM program provides learners with a clear knowledge of the program objectives. N/A 1 2 3 4 5
2. The instructional interactions in this IMM program are appropriate for the objectives. N/A 1 2 3 4 5
3. The instructional design of this IMM program is based on sound learning theory and principles. N/A 1 2 3 4 5
4. The feedback in this IMM program is clear. N/A 1 2 3 4 5
5. The pace of this IMM program is appropriate. N/A 1 2 3 4 5
6. The difficulty level of this IMM program is appropriate. N/A 1 2 3 4 5

Comments:

AREA 2 – COSMETIC DESIGN REVIEW
7. The screen design of this IMM program follows sound principles. N/A 1 2 3 4 5
8. Color is appropriately used in this IMM program. N/A 1 2 3 4 5
9. The screen displays are easy to understand. N/A 1 2 3 4 5

Comments:

AREA 3 – PROGRAM FUNCTIONALITY REVIEW
10. This IMM program operated flawlessly. N/A 1 2 3 4 5

Comments:

Figure 7.2. Sample expert review form for a multimedia program.
If you must limit expert review, content experts are probably the most important expert sources of formative information for education and training products. Why? Because if you do not get the content right, the eventual users will be misled. One of the problems with many interactive learning systems is that these programs lack subject matter integrity because of a lack of expert review. This is a major challenge because so much material can be incorporated into a single interactive product. The integrity issue is especially challenging when interactive learning is delivered via the World Wide Web. Many Web sites include links to other sites, and guaranteeing the accuracy and currency of all the related links is beyond the powers of most developers. At the very least, the content of the first two or three levels of links should be examined. Unless the accuracy and validity of information and its organization are carefully reviewed, the level of integrity necessary for educational materials may be lacking.

In addition to instructional design and content experts, we have found that people with special expertise in human-computer interface (HCI) design and the aesthetics of interactive learning systems can provide useful expert reviews. For example, if your team doesn’t include an art director who is responsible for the look and feel of the product, it is wise to ask graphic artists to critique your prototype. Of course, you would not decide to make major changes in the design elements of an interactive learning system based on the opinions of just one expert because aesthetic appeal is much more subjective than some of the other criteria to be reviewed. On the other hand, if several graphic artists retch at the sight of your interface, you may have a problem worth fixing!

Experienced designers of interactive multimedia are often the best experts for reviewing user interface issues, but there are people who specialize in HCI issues per se. Figure 7.3 presents a user interface instrument that can be used to guide reviews provided by expert instructional designers and very experienced users of interactive learning systems (Reeves & Harmon, 1994).

Contracting with the “right” experts for review services is a crucial step in setting up a formative evaluation. If subject matter experts (commonly called SMEs) are already part of your team, one of their primary responsibilities will be checking the accuracy and currency of your content. However, even when working with qualified SMEs, it is a good idea to have the content reviewed by other content experts. In 1984, one of us learned this lesson the hard way when we collaborated with a team of nurses on the design of an interactive videodisc program about shock treatment. The video components were taped at a hospital in Philadelphia where nurses still wore traditional nursing hats. Little did we know that these were practically the last nurses on earth still wearing this type of
cap! When we showed the program at a medical education conference, many of the nurses in the audience criticized the video for being out of date, even though it had been recorded only weeks beforehand. If we had hired nurses from other hospitals to review the videotaping early on, we would have detected this problem and included nurses with more up-to-date apparel. As it was, the “face validity” of this program was seriously damaged in the eyes of many members of the target audience.

The costs for SMEs to review interactive multimedia programs can vary widely depending on the field. Some instructional design, graphics, and HCI experts will charge hundreds of dollars per hour to review interactive prototypes, but we have found that using graduate students and university faculty can be a much less expensive source of expert review. Of course, your clients may insist on using the recognized “gurus” in any given area, and they are often worth every cent they are paid because of their insight and creativity. Using the Web for interactive learning delivery is becoming prevalent (Kahn, 1997, 2001), but procedures for expert review of instructional Web sites are not as well defined as they need to be (Reeves & Carter, 2001).

**Figure 7.3.** Sample user interface review form for ILS.
User review

The perspectives of “experts” are valuable, but the opinions of the target audience for your interactive learning system are of equal importance. User review is based on the analysis of user behavior during the use of the product to be evaluated. Therefore, user review requires an understanding of the actual user profiles, their tasks, and the contexts in which the tasks are performed. As one important objective is to ensure that user differences are accommodated, it is important that user review be done with a sample of people whose background knowledge and expectations approximate those of the final intended users. During user review, users should be allowed to work in realistic conditions, without interruption from an observer, in order to accurately replicate the intended context of use (Bevan, 1995).

Suppose you are designing an interactive learning system for use in schools. In that case, the most valuable information for making decisions about improving the program can be derived from systematic observations of learners while they use the program. Observations can be conducted in the development lab or on-site at a school. Learner opinions, actions, responses, and suggestions will provide you with practical guidance in making decisions about improving your product. Of course, you would also want the teachers who must implement the system to review it, especially with respect to seeking their ideas about how they could integrate it into their existing practices. Widening the review process to include parents, administrators, and school media specialists is also advised in this context. Few interactive learning systems have been successfully integrated into schools (Cuban, 2001), a problem that might have been reduced by more inclusive formative evaluation.

Observations of learners engaging with your interactive learning system at various stages of its development can be a valuable, if somewhat humbling, experience. You may be surprised at how frequently what seemed to be the most user-friendly aspects of your program befuddle would-be learners. Alternatively, what you view as motivating may bore the intended audience. Fortunately, you will often find that your creative design ideas are validated by learners. A sample protocol from the Apple Interface Laboratory is given in Figure 7.4.

During your observations, you will see users doing things you never expected. When you see participants making mistakes, your first instinct may be to blame the mistakes on the participant’s inexperience or lack of intelligence. This is the wrong position to take. The purpose of observing users is to see what parts of your product might be difficult or ineffective. Therefore, if you see a participant struggling or making mistakes, attribute the difficulties to faulty design, not to the participant. If you are an evaluator, you may be more tolerant of bad news from users than the
instructional designers on your team, but they will learn to be more accepting as they see that formative evaluation enhances their products.

**User Observation (Based upon Apple HCI Group Protocol)**

The following instructions guide you through a simple user observation. With this protocol, you will see where people have difficulty using your product, and you will be able to use that information to improve it.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Introduce yourself.</td>
<td>Make the session and task as welcoming as possible. Remember, this test is not designed as a controlled experiment, so keep the environment friendly.</td>
</tr>
</tbody>
</table>
| 2 – Describe the general purpose of the observation. | Set the participant at ease by stressing that you’re trying to find problems in the product. For example, you could say:  
   *You’re helping us by trying out this product. We’re looking for places where the product may be difficult to use. If you have trouble with some tasks, it’s the product’s fault, not yours. Don’t feel bad; that’s exactly what we’re looking for. If we can locate the trouble spots, then we can go back and improve the product. Remember, we’re testing the product, not you.* |
| 3 – Tell the participant that it’s OK to quit at any time. | Make sure you inform participants that they can quit at any time if they find themselves becoming uncomfortable. Participants shouldn’t feel like they’re locked into completing tasks. Say something like this:  
   *Although I don’t know of any reason for this to happen, if you should become uncomfortable in any way, you are free to quit at any time.* |
| 4 – Talk about the equipment. | Explain the purpose of each piece of equipment and how it will be used in the test (hardware, software, video camera, microphones, etc.). |
| 5 – Explain how to “think aloud.” | Ask participants to think aloud during the observation, saying what comes to mind as they work. You’ll find that listening to users as they work provides you with useful information that you can get in no other way. Unfortunately, most people feel awkward or self-conscious about thinking aloud. Explain why you want participants to think aloud, and demonstrate how to do it. You could say:  
   *We have found that we get lots of information from these informal tests if we ask people to think aloud as they work through the exercises. It may be a bit awkward at first, but it’s really very easy once you get used to it. All you have to do is speak your thoughts as you work. If you forget to think aloud, I’ll remind you to keep talking. Would you like me to demonstrate?* |
| 6 – Describe why you will not be able to help. | It is very important that you allow participants to work with your product without any interference or extra help. If a participant begins having difficulty and you immediately provide help, you may lose the most valuable information you can gain from user observation: where users have trouble, and how they figure out what to do. Of course, there may be situations where you must step in and provide assistance, but you should decide what those situations will be before you begin testing. You may decide that you will allow someone to flounder for at least 3 minutes before providing assistance. Or you may identify distinct problems you will provide help on. As a rule of thumb, try not to give your test participants any more information than the true users of your product will have. Here are some things you can say to the participant:  
   *As you’re working through the exercises, I won’t be able to provide help or answer questions. This is because we want to create the most realistic situation possible. Even though I won’t be able to answer your questions, please ask them anyway. It’s very important that I capture all your questions and comments on tape. When you’ve finished all the exercises, I’ll answer any questions you still have.* |

*Figure 7.4. Sample User Observation Protocol.* (Continued)
7 – Describe the tasks and introduce the product.  
Explain what the participant should do first, second, third, etc. 
Give the participant written instructions for the tasks.  
**Important:** If you need to demonstrate your product before the user observation begins, be sure you don’t demonstrate something you’re trying to test. (For example, if you want to know whether users can figure out how to use certain tools, don’t show them how to use the tools before the test.)

8 – Ask if there are questions.  
Before you start, make sure the respondent knows your expectations, then begin the observation.

9 – Conclude the observation.  
When the test is over:  
- Explain what you were trying to find out during the test.  
- Answer any remaining questions the participant may have.  
- Discuss any interesting behaviors you would like the participant to explain.

10 – Use the results.  
To get the most out of your test results, review all your data carefully and thoroughly (your notes, the videotape or cassette tape, the tasks, etc). Look for places where participants had trouble, and see if you can determine how your product could be changed to alleviate the problems. Look for patterns in the participants’ behavior that might tell you whether the product is understood correctly.

It’s a good idea to keep a record of what you found out during the test. That way, you’ll have documentation to support your design decisions and you’ll be able to see trends in users’ behavior. After you’ve examined the results and summarized the important findings, fix the problems you found and test the product again. By testing your product more than once, you’ll see how your changes affect users’ performance.

**Figure 7.4.** Sample User Observation Protocol (continued).

Observing learners can be a time-intensive and exhausting process. It can range from a very simple one-on-one observation protocol to a complex arrangement wherein several observers, video cameras, and computers are used to record learners’ reactions. Whatever type of procedure is followed, it is important that you record information carefully and that you later deal with each issue that arises during the observations.

Figure 7.5 presents a simple formative evaluation review form with three columns, one for indicating what section of an interactive program is being reviewed, one for recording observations, and the last for recording the actions taken in response to the issues raised by the observations. The last column is very important because it provides the evidence that the evaluation data collected has actually had an impact on design decisions.
**Formative Review Log**

**Program:** Learn or else!  **Reviewer:** Smith  **Date:** May 15

<table>
<thead>
<tr>
<th>Screen</th>
<th>Comments, Questions, Suggestions</th>
<th>Action Taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-17</td>
<td>The forward and back navigation arrows are so small that users seem to have trouble placing the mouse cursor on them.</td>
<td>Enlarge the navigation arrows by 50% and repeat observations.</td>
</tr>
<tr>
<td>C-23</td>
<td>The users think that the “Question Mark” icon will take them to help, but it takes them to a list of frequently asked questions instead.</td>
<td>Use the “Question Mark” icon for help, and find a different icon for the frequently asked questions.</td>
</tr>
</tbody>
</table>

**Figure 7.5.** Formative evaluation review log for e-learning program.

**Usability testing**

Usability is an important issue in the design of any software, including interactive learning systems. According to Shneiderman (1987), usability is a combination of the following user-oriented characteristics:

1. ease of learning,
2. high speed of user task performance,
3. low user error rate,
4. subjective user satisfaction, and
5. user retention over time.

There are several books that provide excellent guidance to evaluating user interface issues, a process known as usability testing (Hix & Hartson, 1993; Nielsen, 1993). Usability testing is especially critical in the design, dissemination, and implementation of interactive multimedia for education, training, performance support, and information access (cf., Blattner & Dannenberg, 1992; Laurel, 1990; Polson, 1988, Preece, 1994; Shneiderman, 1987). Too many formative evaluation studies are only focused on whether users like a program or not, but usability is a much deeper subject.

There are instances when you might evaluate usability without users. Time with users is often limited; it is not a free resource. In addition, users can find it difficult to visualize how a product could behave differently and they therefore tend to evaluate according to what already exists, rather than to what is possible. Some usability criteria will only be reliably identified or articulated by trained human factors evaluators using protocols such as heuristic evaluation or usability inspection.
At least three evaluators with a mix of experience and expertise are required for heuristic evaluation or usability inspection because fewer will not identify all the usability problems (Nielsen & Mack, 1994).

**Usability in product development.** Designing for usability requires early specification of usability goals (Nielsen, 2000). Usability is built into a new product from the analysis phase of a project by identifying and analyzing the critical features of and interactions between users, their tasks, and the product. Specific contexts in which usability is to be measured should also be identified. These usability goals can then be used to interpret the findings from the user analysis and to identify the goals and constraints that will direct the design and set criteria against which a design can be tested once it is built.

As part of the design phase, prototypes are commonly developed and tested. The outcome of this early evaluation results in iterations of the design. It is important to begin usability evaluation at the earliest phases of design because, if left until just before a product is released, there will be little chance to make any significant design changes. Testing formally for compliance with usability specifications takes place in the testing phase. This is usually at the alpha, beta, or final gold master versions of the product. An alpha version is used in the earliest tests, conducted when the product is still at the prototype stage. Alpha testing is generally done with in-house personnel. Beta versions are released to selected samples of users for testing. The product is more refined at this point, nearing completion, but still needing to be debugged. The gold master version of a product is supposed to be flawless, although anyone who has bought version 1.0 of a new piece of software realizes that this is rarely the case. The persistence of problems (bugs) is why testing even gold versions of a product is important.

There are a number of benefits for usability evaluation if it is iterative and considered in all phases of the development life cycle. Iterative design helps with the management of product development and so reduces the risk of projects going off track. Early testing can detect unclear or unreasonable usability goals. Usability objectives can help to facilitate communication and decision-making between human factors evaluators and product designers. Also, usability testing allows developers to obtain and appreciate a user perspective of their product.

Speed of using software is often the focus for usability studies. It was reported in *Information Week* (Leibs, 1994) that studies conducted at Carnegie Mellon University’s Software Engineering Institute estimated that a savings of $2.5 million could be realized by a large company if they developed “an interface that could trim eight-tenths of a second off the time a user needs to perform a repetitive computer-based task, such
as order entry or customer service” (p. 28). Although the economics of this particular example may seem inflated, there is no question that big business is increasingly concerned about the computer interfaces that “knowledge workers” must use and the degree to which these user interfaces support effective, efficient performance. Time is money. A second saved may be a penny saved, and when enough pennies are saved, significant economic benefits can result.

Time saved in business and industrial training is also highly prized because corporations generally want to get their employees back on the job as quickly as possible. In fact, increasing the efficiency of training, rather than improving effectiveness, is sometimes used as the primary criteria for summative evaluations of interactive learning systems (Reeves, 1988).

But what about education? Is time important in schools? At first glance, time might appear to be a relatively unimportant resource. To be sure, time is a factor in planning and managing most schools, but it is usually thought of in terms of 180 school days, six and a half hour days, 50 minutes class periods, and so forth. (The days, hours, schedule, etc., devoted to schooling vary considerably from country to country.) But when it comes down to what teachers and students do with their time on an hour by hour, minute to minute basis, there appears to be little concern for accountability.

Hence, speed may not appear to be important in examining educational software, but we will argue that it is an issue. The ease and speed with which a learner is able to engage in meaningful cognitive interactions with an interactive learning system is an indicator of how soon he or she will be able to devote his or her cognitive powers to the content and learning dimensions of the program rather than to figuring out how to navigate through and control the system. After all, human information processing power is limited, and interactive learning systems should be designed so that meaningful engagement with a program is enabled without unnecessary stress. Therefore, the usability of educational software is just as relevant as it is for other types of software.

**Formally assessing usability.** At the Learning and Performance Support Laboratory at The University of Georgia, we acquired a portable multimedia usability lab (the one depicted in Figure 7.6) to help in formative evaluation, especially usability studies. This lab is transportable to any site where an interactive learning system or any other type of software is being used for education, training, information, or performance support purposes.

A portable lab includes a remote-controlled video camera that can be focused on the user’s face, the user’s computer, keyboard and mouse,
any other aspect of the user environment considered important in the evaluation. The system simultaneously records whatever appears on the user's screen. Evaluators sit at a control panel that allows them to observe the user directly or on any of the video screens displaying selected aspects of the context. Evaluators can control what is recorded, e.g., most of the user’s screen along with a small insert image of the user’s facial expressions or body language.

**Figure 7.6.** Portable usability lab.

Commercial software developers have employed usability labs for formative evaluation of software applications for many years (Branaghan, 2001; Gomoll, 1990). Fixed usability labs generally consist of two rooms separated by a one-way glass window (see Figure 7.7). In one room, a computer user sits at a desk and interacts with the application being evaluated, e.g., a new spreadsheet program. Several video cameras mounted in the room are focused on various aspects in the room. In the other room, evaluators and designers sit at control panels where they can simultaneously observe the user in the room through the one-way glass or any of the video screens displaying selected aspects. The user may be instructed to “think aloud” as he/she uses the program, e.g., talk about why certain choices are made or describe any confusion about the

---

**WEB LINK**

For more information about the portable lab pictured here, go to: www.usabilitysystems.com
program’s interface. Alternatively, the evaluators may question the user via headsets or speakers about why he/she has done certain actions. Users are informed that they will be observed, and they have the right to discontinue a test at any time for any reason. Typically, these sessions are videotaped for later analysis and documentation. Some fixed usability labs feature a third room where clients can observe the usability testing as it is being conducted.

**Figure 7.7.** Fixed usability laboratory.

A portable software usability lab is patterned after these commercial labs, but rather than forcing users to come to a lab and test software in an artificial environment, the portable lab allows the users to stay in their own environment. We believe that this increases the validity of many evaluation studies. Regardless of what type of lab is used (fixed or portable), usability testing enables evaluators to collect both quantitative and qualitative data related to issues such as user interface, mental models, navigation, documentation utility, effectiveness, and efficiency.

You can do some simple usability testing with only a single video camera, especially if you cannot afford to rent a formal laboratory or buy a portable usability lab. We have used a single video camera to record two users working their way through a prototype package. This has some unique advantages. The fact that there are two people with one system means that they have to talk about their interpretations of screen images and negotiate their actions. Reviewing the video record of these two-person interactions can be very informative. It also circumvents the difficulty of getting users to participate fully in a think-aloud protocol.

**Usability protocols.** A variety of evaluation protocols are possible to assess usability (Hix & Hartson, 1993; Hughes & Burke, 2001; Nielsen, 1993, 2000). Nielsen (1993) identified the following methods for gathering usability data: observation, think aloud, questionnaires, interviews, focus groups, logging actual use, and user feedback. Evaluators
select the appropriate methods to collect data to address different usability issues and questions. Each of these methods has different strengths and weaknesses, and combining different methods is often necessary to improve overall usability testing.

As noted in the User Review section, you might want to ask questions to users during observations, but asking questions during an observation can change what the user would naturally do. An alternative is a delayed think-aloud approach whereby you record the user with the portable usability lab, and later play the tape back to the user. During the playback, you can ask the user to state what he/she was thinking while interacting with the program or ask specific questions. The tape assists the user in recalling the recorded session. In addition the same tape can be shown to human factors experts for their advice and interpretations. Alternatively, a focus group of designers can review videotapes of users in their actual working conditions to stimulate new ideas about the program.

Formative evaluations sometimes involve the use of experts to judge the performance of learners on various types of tasks. Reliability is an important issue whenever human judges are used. Having videotaped data scored by multiple experts can provide reliability information about the data collection process. Collecting data about benchmark tasks is another use of the portable usability lab. A benchmark task is a common activity the learner performs with the system. These benchmarks are selected by the developer to measure quantitatively the interface design (Hix & Hartson, 1993). The usability system can record the learner’s performance on benchmarks for later analysis.

Working with one of the authors of this book, Conyer (1995) provided an excellent summary of a range of literature on the topic of usability testing. Conyer described six alternative methods that can be employed to determine the usability of an e-learning product: 1) heuristic evaluation, 2) pluralistic walkthroughs, 3) formal usability testing, 4) empirical methods, 5) cognitive walkthroughs, and 6) formal design analysis.

**Heuristic evaluation.** Originally conceived by Nielsen (1993), this method employs a set of principles (termed heuristics) which have been defined prior to the evaluation. Although usually done with experts, heuristic evaluation can also be done with a sample of intended users. Evaluators (experts or users) independently examine the product and judge its compliance with the set of heuristic principles. Conyer (1995) elaborated:

> Each evaluator works through the interface at least twice, the first time to get a feel for the flow of the interaction and the second time to focus on the specific interface elements within the con-
text of the larger whole. Observers can offer help to evaluators when they are clearly having difficulty and after they have commented on the usability problem they are experiencing (p.40).

The evaluators’ comments can be recorded either by themselves or by an observer. Then they are interpreted and summarized for an overall evaluation. Conyer (1995) then recommended that:

a debriefing session is then held with all evaluators, observers, and representatives of the design team to brainstorm possible ideas to address the major usability problems, as well as to discuss the positive elements of the interface design. A priority list is then drawn up of all usability problems with reference to the heuristics that were not followed in the design, and with a time and cost estimate to correct each problem. Priority is determined according to the frequency and impact of the problem, and if the problem can be overcome in another way, e.g., with training (p.41).

The heuristic evaluation method has advantages and disadvantages. It is less time-consuming that other approaches and it does reveal many important weaknesses in a product. The debriefing session is especially important in determining how the problems identified can be fixed or ameliorated. This method is relatively easy and economic to undertake and can be prepared quickly to provide efficient feedback to a team of designers. However, recruiting experts can be difficult because most experts are busy people. If end-users are involved, getting a representative cross section of the target audience is yet another challenge. According to Nielsen (1993), a further weakness of heuristic evaluation is that it generally does not find as many problems as formal usability testing. The actual heuristics used in this method depend on the nature of the product. Nielsen has defined a set of ten general heuristics that can be applied to any type of software product. Appendix A presents a set of heuristics designed specifically for e-learning programs.

**Pluralistic walkthroughs.** This method can be used with paper prototypes as well as with fully functioning versions of an interactive learning system. The pluralistic aspect comes from the mix of users, designers, and experts involved in the walkthroughs. Conyer (1995) explained:

The goal of this method is to systematically review the usability of an interface and its flow from a task-based, user-centered perspective while at the same time considering the design constraints. In the context of task-based scenarios, end-users, product developers, and human factors experts evaluate a prod-
uct from the perspective of the end-user. The evaluators sequentially write down each action they would take when pursuing a designated task. A group discussion then follows, with end-users presenting their information first. Subject matter experts are available at all stages for domain-specific questions. (pp. 41-42)

The walkthrough method is task-based, and thus it is more narrowly focused than other methods. Therefore, it identifies more specific problems than general problems. For example, the walkthrough method might be used to examine whether users correctly understand how to log-in and log-off using a course management system.

**Formal usability inspection.** This approach can be used to examine both cognitive processing and behavioral tasks involved in the usability of a program. However, end-users are generally not involved in this type of evaluation, and it is not conducted in the context in which the e-learning product would normally be used. Instead, members of a design team and external experts participate in this type of inspection, as described by Conyer (1995):

Usability issues are reviewed within the context of specific user profiles and defined goal-oriented scenarios by applying a task performance model and heuristics. This method captures how evaluators perceive the information, plan to use the information, decide how to proceed, and perform the selected action. A six-step process is normally used, namely (1) Planning; (2) Kick-off Meeting, when the team comes together for the first time; (3) Preparation, when the evaluators review the program independently; (4) Review, to discuss the aggregated usability issues; (5) Rework, when solutions are found and implemented; and (6) Follow-up, to determine the effectiveness of the evaluation process. There are clearly defined participant responsibilities, namely: Moderator, who manages the process; Design Owner, who is responsible for representing and then upgrading the product being inspected; Evaluators, who find and report usability problems (such as designers, documentation specialists, and human factors evaluators); and Scribe, who records all identified problems and decisions. (pp. 42-43)

The focus of formal usability inspections is more general than in walkthrough methods. Ideally, this type of formal review is conducted as soon as a reasonably complete beta version of the interactive learning system is available. The inclusion of design team members in this process can be especially useful, provided they can keep an open mind with respect to problems that might be identified. However, there should be other participants who are not part of the design team. Just as an author of a book needs an external proofreader to spot errors in a text,
design teams need external experts to see the inevitable flaws in a user interface design.

**Empirical methods.** This approach is effective for establishing cause and effect, or for addressing a specific question or problem through focused testing. However, it can be very time consuming and it does require an evaluator trained in empirical methods. In practice, empirical methods should not be undertaken until a formal prototype is working and is robust enough to test. Conyer (1995) summarized empirical methods as follows:

Data can be collected in an experimental test to prove or disprove a hypothesis, e.g., the number of correct responses and errors made by a user under controlled conditions. A hypothesis is posed based on a set of objective measures for the evaluation. A plan for how the measures are to be collected is then determined. Subjects are found for the test, data is collected and analyzed to determine if the proposed hypothesis has been proven. (p.43)

Empirical methods should be reserved for resolving significant disagreements among design team members because of the time and expense involved. Suppose some team members want to use realistic icons and “roll-overs” as a major feature of the e-learning interface whereas more conservative members of the team want to employ simple icons with text labels. An empirical test could be set up to compare the effectiveness of the two interface designs, using representative samples of the target population. It would be especially important to examine such a major design issue in light of requirements for accessibility for disabled learners for whom features such as roll-overs may present unnecessary barriers.

**Cognitive Walkthroughs.** This approach is an effective method for revealing problems that affect users’ overall performance, and it can capture cognitive processes of both novice and expert users. This method is especially useful in revealing whether an e-learning product presents a “cognitive overload” problem. Human mental processing has limits, and if the interface of a program demands too much of a learner’s mental capacity, then there may not be enough left over to engage in meaningful learning. Conyer (1995) provided an overview of this method:

Cognitive walkthroughs are used to evaluate the ease of learning to use a product, particularly by exploration. The method is a formalized way of imagining people’s thoughts and actions when they use a product interface for the first time (Lewis & Riemann, 1994). Cognitive walkthroughs focus most clearly on problems that users will have when they first use an interface.
without training. The method uses an explicitly detailed procedure to simulate a user’s problem-solving process at each step, checking to see if the user’s goals and memory for actions can be assumed to lead to the next correct action (Nielsen & Mack, 1994). There are three phases in the procedure, namely (1) Preparatory, when the analysts agree on the input conditions for the walkthrough, such as type of users, tasks and action sequence for each task; (2) Walkthroughs, which can be an individual or group process; and (3) Analysis.” (p.44)

In designing a protocol for this method, designers are forced to consider the user’s background knowledge, the user’s goal structure, and the cognitive complexity required for a user to use the product. However, the method continually interferes with the interaction, and people not trained in cognitive psychology may find it difficult to decompose tasks into a collection of sub-tasks. Other constraints are the lack of comparable measures of task time and the primary focus on one attribute of usability, namely ease of learning.

**Formal design analysis.** This method provides assistance in identifying problems early in the design process. It is less expensive, as it can be performed by a single person. In addition, the approach enables the comparison of different design options. Conyer (1995) explained:

> Formal design analysis techniques aim at improving the design process. An example is the “Goals, Operators, Methods and Selection Rules” (GOMS) model developed by Card (Eberts, 1994).... Formal design analysis is based on the premise that understanding of the requirements of the task to be performed is the key to understanding behavior. Tasks to be performed by an expert user are decomposed into goals (a series of cognitive and motor components), operators (actions that a user executes), methods (sequences of steps), and selection rules (needed if more than one method is available to accomplish a goal). Algorithms are then applied and each design is rated with a single number. Alternative design possibilities are then compared based on the numerical result. (p.45)

Unfortunately, the formal design analysis method misses many key components of behavior that must be considered in interface design, such as learning the task, error behavior, and transfer of learning to other products. This method is somewhat difficult to learn, and few instructional designers are trained to do it. Another weakness is the method is the assumption that all cognitive operations are of equal difficulty.

So which methods do we recommend? This depends on your purpose. As illustrated in Figure 7.8, Conyer (1995) suggested different methods
and data collection tools that can be considered for different evaluation purposes.

<table>
<thead>
<tr>
<th>If the purpose of the usability evaluation is to evaluate ______</th>
<th>then consider the _____ methodology</th>
<th>using the _____ recording method</th>
</tr>
</thead>
<tbody>
<tr>
<td>the ability of the user to carry out a task using a product in a particular context</td>
<td>Formal Usability Inspection</td>
<td>Verbal Reports, Concurrent Think-Aloud, Video Analysis, Software Support</td>
</tr>
<tr>
<td>how easily users can carry out a task</td>
<td>Pluralistic Walkthrough, Formal Usability Inspection, Cognitive Walkthrough, Formal Design Analysis</td>
<td>Verbal Reports, Concurrent Think-Aloud, Questionnaire, Video Analysis, Auto-Logging Programs and Audit Trails</td>
</tr>
<tr>
<td>how quickly users can carry out a task</td>
<td>Empirical Studies, Formal Design Analysis</td>
<td>Video Analysis, Auto-Logging Programs and Audit Trails</td>
</tr>
<tr>
<td>the overall quality and acceptance of a product</td>
<td>Heuristic Evaluation</td>
<td>Verbal Reports, Questionnaire, Software Support</td>
</tr>
<tr>
<td>problems with using a product</td>
<td>Pluralistic Walkthrough, Formal Usability Testing, Cognitive Walkthrough, Heuristic Evaluation</td>
<td>Verbal Reports, Concurrent Think-Aloud, Video Analysis, Questionnaire, Auto-Logging Programs and Audit Trails, Software Support</td>
</tr>
<tr>
<td>how easy it is for a novice to learn to use a product</td>
<td>Cognitive Walkthrough, Formal Design Analysis</td>
<td>Concurrent Think-Aloud, Video Analysis</td>
</tr>
</tbody>
</table>

**Figure 7.8.** Usability testing methods. (Conyer, 1995)

**Recording methods.** When employing usability evaluation methods, there are a variety of recording methods that can be used to capture data. Conyer (1995) summarized the options as:

**Verbal Reports**

Users provide a verbal report soon after completing their evaluation. This information can then be informally reviewed or formally classified into categories for evaluation (Karat, 1988).

**Concurrent Think-Aloud Method**

Evaluators verbalize their thoughts while interacting with a product. The purpose of this method is to show what the users...
are doing and why they are doing it *while* they are doing it, in order to avoid later rationalizations. However, thinking aloud is not something that people are used to doing, and thus subjects rarely give quality think-aloud reports without prompting. Often we have undertaken this approach but using a pair of evaluators who talk to each other about assumptions and choices and these are recorded. This approach is a little more natural and produces more effective outcomes.

**Questionnaires**

Questionnaires can be composed of items that address information and attitudes. It is important to keep questions specific rather than general and to ask questions about actual product experience rather than hypothetical questions about possible product changes.

**Video Analysis**

Video recordings can be used to capture data about user interactions. See the section previously on the different ways in which video recordings can capture user performance. Video can be combined into a composite tape with screen and an insert shot of the user looking at the screen. Tools have been developed to link multimedia video recordings to data analysis techniques, such as QSR *NVivo* which enables the classification of qualitative reports (Richards, 1999).

**Auto-logging Programs and Audit Trails**

Auto-logging programs can be used to track user actions with respect to duration and frequency of use, e.g., number of keystrokes, requests for help, duration, and path through a piece of software. While the evaluator is freed from data collection, the volume of data that can be produced and the statistical analysis of such data can be very complex. Frequently, it is necessary to combine this type of evaluation with Video Analysis to obtain a complete picture of what the user was doing.

In addition to these planned usability methods, you may observe users with a portable usability system over time in their natural environment doing what they decide to do on their own. This is more like naturalistic evaluation than the types of usability evaluation methods delineated by Conyer (1995) and Nielsen (1993). Whereas the formal methods are usually more focused and efficient, natural observation is time-consuming and less directed. However, the generalizability of findings from the naturalistic approach may be greater. Both can be valuable approaches.
Many times in designing an interface, you may have multiple options for designing an interaction with the computer. “Rapid prototyping” usually requires the creation of multiple designs for small components of your interactive learning system (Hamblen & Furman, 1999; Hix & Hartson, 1993; Tripp & Bichelmeyer, 1990). By comparing tasks performed with each option for interaction, you can feel more confident about which design to use. Alternative metaphors for program designs, various types of icons, and different treatments (e.g., humor or drama) can be examined via usability studies.

Finding appropriate users who will allow you to videotape them interacting with a prototype learning system is sometimes difficult. Some authorities suggest only using a few participants that match your target users (Hix & Hartson, 1993). It is not uncommon to reach a point of diminishing returns on the feedback you receive after four to six learners have completed the usability protocol. As noted above, you may wish to supplement these by using one or two experts in human computer interface (HCI) design to review your interactive learning system.

The importance of the “usability” approach to formative evaluation is considerable. Currently, instructional designers have an inadequate base of knowledge about how users react to and learn with interactive learning systems and other types of computer programs, such as electronic performance support systems (EPSS) (Gery, 1991). The data revealed through usability testing provides an improved basis for guiding the design and implementation of interactive learning systems. We believe that the enhancement of our understanding of interactive multimedia user interfaces can improve the dissemination, implementation, and effects of using interactive learning systems at all levels of education and training. Initial marketing of educational multimedia has succeeded primarily on the basis of selling the “bells and whistles” of technology, but now school boards, superintendents, parents, and taxpayers are beginning to demand research-based evidence that multimedia enhances learning (Cuban, 2001). This demand for accountability is even greater in the corporate world (Rosenberg, 2000). Fundamental understanding of how the interfaces for interactive learning systems are understood by students and trainees is an essential part of that evidence. Usability testing can provide you with precisely that kind of evidence.

**Alpha, beta, and field testing**

Alpha, beta, and field tests are terms borrowed from the software engineering process followed by commercial software development companies. This structure of evaluation events is useful in organizing small group and field trials of an interactive learning system. Suppose that you are developing an interactive learning system with relatively expensive...
media elements, such as high-quality computer graphics, digital video, and animation. Before spending project resources to produce these materials, you should conduct an alpha test using storyboards or screen mock-ups with a small sample of typical users (Martin, 2000). If instructors will play a significant role in the eventual implementation of the interactive learning system, they should also be involved in the alpha test.

The materials used in an alpha test may be storyboards, but today they are more likely to be a number of prototype screens presented on either the intended delivery system or some easy to program alternative. An Apple Macintosh computer and a software construction program such as Macromedia Dreamweaver or even Microsoft PowerPoint can provide excellent rapid prototyping environments for interactive learning systems regardless of how they will be delivered. Some elements of the eventual program may be represented by paper scripts, story boards, sketches, and other draft documents, but even during an alpha test you should create and evaluate as much of the “look and feel” of the final product as possible.

Critical aspects to be evaluated during the alpha test include the program’s interactivity, comprehension, and appeal. You will want to look for evidence that learners know what to do at critical junctions of the program. You will assess their attention levels and level of active responding. You will look for signs that they are enjoying the experience of using the program. If you rely on learners’ responses to direct questions about these factors, you must encourage them to be frank, realizing that their human inclination will be to give you answers that please you rather than their real reactions.

Alpha tests are conducted as soon as a reasonable prototype version of the program can be assembled. By contrast, beta tests are conducted with more-or-less complete versions of an interactive learning system in settings as much like the context of final implementation as possible. While internal staff are usually the primary data collectors during an alpha test, you should consider using external evaluators for the beta test if resources permit. As an interactive learning system nears completion, internal staff are naturally less receptive to modifications. External evaluators can provide a fresh perspective on the formative process, seeing things that internal staff may not be able to perceive because of their intensive familiarity with a program, and their commitment to the present version.

Conducting a beta test at multiple sites is also highly recommended. This can be accomplished by offering beta versions of a program to typical users at no cost with the expectation that they will report their reactions and any bugs they find back to you. Multiple sites are important because the idiosyncrasies of an individual test site may unduly
influence the findings of a beta test. Although it is a great idea for development team members to visit some of the beta sites to see the program being used in an authentic context, they obviously won’t be able to be at all sites at the same time. In many cases, they will have to rely upon secondhand reports gathered from questionnaires, telephone interviews, or focus groups.

Questionnaires, interviews, and focus groups are three of the most frequently used strategies for collecting formative evaluation data. Each of these strategies is a form of survey. Figure 7.9 presents the important steps involved whenever you undertake to use a survey method to collect data. If time, money, and personnel resources allow, consider collecting formative evaluation data with more than one method during alpha, beta, and field tests. For example, a questionnaire might be used to collect information about user reactions to screen designs. Later, interviews or focus groups might be used to collect more detailed information about which aspects of screens can be improved and how. Alternatively, if you are planning to distribute a survey to many people, you might want to use interviews or focus groups first to identify the issues that should be included on the questionnaire.

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Organize a team to assemble and review the instrument.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>Determine the purposes of the survey (e.g., collecting user satisfaction data).</td>
</tr>
<tr>
<td>Step 3</td>
<td>Identify a representative sample from whom to collect the data.</td>
</tr>
<tr>
<td>Step 4</td>
<td>Generate a list of draft questions.</td>
</tr>
<tr>
<td>Step 5</td>
<td>Construct a draft instrument (questionnaire, interview protocol, or focus group protocol).</td>
</tr>
<tr>
<td>Step 6</td>
<td>Test the instrument with a small sub-sample of your representative sample.</td>
</tr>
<tr>
<td>Step 7</td>
<td>Revise the instrument and retest if necessary.</td>
</tr>
<tr>
<td>Step 8</td>
<td>Administer the instrument.</td>
</tr>
<tr>
<td>Step 9</td>
<td>Process and analyze the data.</td>
</tr>
<tr>
<td>Step 10</td>
<td>Report and use the results.</td>
</tr>
</tbody>
</table>

**Figure 7.9.** Ten steps to using survey methods.

A field test can be conducted with an interactive learning system that has been improved through alpha and beta tests, especially if there is still
time or money available for additional enhancements, perhaps in terms of packaging or implementation strategies. Generally, field tests are conducted without the direct involvement of the development staff. Instead, the emphasis is on evaluating the interactive learning system under conditions virtually identical to more widespread implementation.

Using external evaluators to oversee the process increases the credibility of field tests. Most likely, you will find that field tests and product implementation overlap because the pressure to market and/or disseminate interactive programs will be considerable. In fact, don’t be surprised if your team is pressured to release programs that have not been completely validated! This is a common (and we would argue unethical) practice in the commercial software arena where programs are expected to be “debugged” by the clients who buy version 1.0 of a program. Admittedly, field tests are expensive, but the product recalls or lawsuits that could result from a poorly tested system should warrant their application, especially when large scale products, such as integrated learning systems, are involved.

**How should the information be reported?**

Formative evaluation results are often reported to other team members and/or clients in a less formal manner than other types of evaluation reports. The emphasis in reporting data should be on its timeliness and specificity so that necessary and desirable modifications in the interactive learning system can be made as efficiently as possible. Just as motion picture film crews gather together with the producers, director, and actors at the end of each shooting day to review the “rushes,” you may want to establish a regular meeting time to review the results of formative evaluation with the members of the development team, especially the instructional designers. Of course, it is advisable to keep a detailed log of the program revision process so that the decision-making rationale can be re-examined at later dates if necessary.

The key to reporting formative evaluation data is to establish good rapport with the other members of the development team. The ideal is that the development team members will eagerly seek out the findings of your formative evaluation efforts, but the reality is that developers may sometimes feel resentful or stressed about the need to make further changes in a program into which they have poured their hearts and souls. Programmers, graphic artists, videographers, and anyone else directly involved in the production aspects of a project may be especially wary of formative feedback when they are working under difficult deadlines. The
sensitivity and communications skills demanded of the evaluator should not be underestimated.

Summary

Formative evaluation consists of a wide variety of strategies designed to help you improve the usability, effectiveness, and appeal of an interactive learning system. Generally, the earlier you begin the formative evaluation process, the better off you will be. You should also plan to continue formative evaluation as long as it is economically feasible. Frankly, it is difficult to know when to stop formative evaluation. Anything as complex as an interactive learning system can be constantly “tweaked” to make it better and better, but at some point the costs of collecting more formative evaluation data will exceed its benefits. In other words, at some point, you have to let things go.

Finally, remember to cultivate a spirit of creativity throughout the formative evaluation process. Most people have good ideas, but they will only share these ideas if they know that they are genuinely valued. If you give people credit for their recommendations, they are very likely to work all that much harder to improve the quality of an interactive learning system. Although most people agree that quality is everyone’s business, the “formative” attitude must be constantly nurtured and reinforced.

A major goal of formative evaluation is to “optimize” your interactive program before it is subjected to any type of summative evaluation, such
as effectiveness or impact evaluation. In the next chapter, we present strategies for evaluating the effectiveness of interactive learning environments.

References


