

# Learning Design and Formative Evaluation

A Paper Written for Week 7 of the OLDS MOOC 2013

Thomas C. Reeves, Ph.D.  
The University of Georgia

Yishay Mor, Ph.D.  
Open University of the UK

Some aspects of “evaluation,” the focus of Week 7, have been woven into the fabric of the first six weeks of the OLDS MOOC: 1) Initiate, 2) Inquire, 3) Ideate, 4) Connect, 5) Prototype, and 6) Curate. For example, as a learning designer, you would normally conduct reviews of existing educational resources before setting off to develop your own during the Initiate and Inquire phases of a learning design project. The success of Prototype phase of a learning design initiative is heavily dependent upon evaluative activities. However, regardless of how much upfront analysis and investigation you do, your initial design concepts and prototypes will almost always be subject to improvement. This is where formative evaluation comes in.

Formative evaluation is the essential “life-force” of the learning design process. Formative evaluation is “the systematic collection of information for the purpose of informing decisions to design and improve the product.”<sup>1</sup> Virtually everything about a learning design 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 learning designers require specific information to guide decisions about how to improve the learning design as it is being designed, developed, and implemented. As described below, you can collect this information using many different methods from a variety of different people, ranging from subject matter experts to members of the intended learning population for the learning design. Within the context of the OLDS MOOC per se, peer evaluation is especially relevant, given that you have been asked to engage in reviewing the learning design artifacts of others at virtually every step of the process.

## Objectives

A careful reading of this paper will enable you to:

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

## Formative Evaluation Strategies

As illustrated in Figure 1, different types of decisions must be made when you attempt to improve a learning design, each of which is tied to one or more specific questions that can be addressed by formative evaluation activities, such as observations and interviews.

Decisions	Example Questions
Should the interface proposed for an Open Educational Resource (OER) be redesigned?	Is navigation clear to learners? Are the meanings of icons clear? Do learners get lost in navigating through the OER?
Should the number and length of activities in a learning design be decreased?	Which activities do learners engage in most frequently? To what extent do learners exhibit a lack of engagement in specific activities?
Should more practice opportunities be added to the learning design?	Do learners pass quizzes? Do learners exhibit mastery of intended objectives? Do learners rate interactive components highly?
Should the learning design scope be expanded?	Is the learning design aligned with curricular guidelines? Do content experts rate the OER as comprehensive?

**Figure 1.** Typical decisions and questions in a formative evaluation.

Some people resist evaluation. However, there is an ethical imperative to evaluate a learning design while it is being developed. After all, learning designs are intended to change people, to modify their knowledge, skills, attitudes, and intentions as well as to influence them to behave differently. The risk of misguiding learners with an untested learning design is too great, and therefore formative evaluation can be viewed a morally responsible activity.

In Barbara Flagg's classic book, *Formative Evaluation for Educational Technologies*<sup>1</sup>, she 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 learning design budgets, if there even is a budget, fail to provide sufficient funding for rigorous formative evaluation. (At least 10% of a project budget should be allocated to formative evaluation.)
- Human Nature – Many learning designers are reluctant to subject their innovations to potential criticism, especially from learners 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 learning designer. In short, you cannot just toss together a rough prototype of a design, and expect formative evaluation alone to turn it into a winner.

- **Measurement Difficulties** – Although some aspects of formative evaluation are relatively easy to determine (e.g., investigating whether learners think various parts of an learning design are engaging), there is a lack of reliable, valid, and feasible methods of evaluating certain kinds of outcomes of learning that a particular design may address, e.g., problem-solving.
- **Knowledge** – Formative evaluation expertise is still not widely available within the learning design community. Many designers lack the knowledge and skills to conduct systematic formative evaluation in an efficient and effective manner.

Engaging in formative evaluation should result in an overall reduction in development and implementation costs over the lifespan of a learning design initiative. (“Costs” encompass far more than just currency, and within the context of learning design often are best calculated in terms of the “sweat equity” and time that you put into a design effort.) Therefore, you should evaluate “early and often.” The sooner formative evaluation is conducted during a learning design initiative, the more likely that substantive improvements will be made and errors avoided.

### **Decisions and Questions**

As a learning designer, you will produce draft design documents and prototypes of various types. Each of these represents an opportunity for making important decisions about enhancing the effectiveness of the final learning design. Should you increase the level of the program’s objectives? How can the design be more engaging for learners? Should more assessment be incorporated into the design? These and other decisions will be faced by you and other members of your design team.

The impetus to make decisions about improving a learning design will come from many directions. You may see an interactive program developed by someone else that inspires a new interface idea. You may find out new information about the interests and knowledge of your prospective learners. Your budget, if any, may be cut, thereby requiring you to reduce the more expensive elements of an interactive program such as animation. These and other factors will be signals that there is room (and often a need) for improvements in a prototype design.

Of course, 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 learning design process. What’s more, a formative evaluation perspective is no less important for those involved in implementing a learning design. The bottom line is that all of us are human and our first efforts to create or implement learning designs are bound to be somewhat flawed. Formative evaluation is essential to detecting and reducing these flaws and eventually attaining the high quality we all desire.

Each decision can inspire different types of questions about improving the learning design. Do learners understand the structure of the design and what their options are at any given moment? Does the learning design maintain the learners’ attention? Do they accomplish the learning objectives? Is it feasible to implement the program as designed? It is too late to wait

until you have completed a learning design to ask these types of questions. Instead, they must be addressed throughout the creation and development of the design.

There are no universal criteria established for formative evaluation of learning designs, but some of the most commonly considered factors are functionality (To what extent does the program work as designed?), usability (To what degree is the program user-friendly for the intended learners?), appeal (How much do learners like it and why?), accessibility (To what degree is the learning design able to be used fully by all potential learners?), and effectiveness (What do they learn?). Different criteria entail many different types of questions. For example, usability implies criteria that can be broken down into smaller issues such as the quality of the user interface or the meaningfulness of icons. 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 learners navigate through a computer-based learning design? How does their navigation relate to the underlying pedagogy of the design? What parts of the interactive learning resource are underutilized? Where would learners like to go, but don't seem to know how? Answering these and other questions will provide you with the information you need to enhance the navigational aspects of an interactive learning resource and ultimately improve its usability.

### **Specific Formative Evaluation Methods**

The key to sound formative evaluation is to collect data systematically at critical stages of the learning design's development and to utilize the findings of each formative evaluation strategy as much as your time and resources allow. The following formative evaluation methods are essential within the context of most learning design initiatives:

- peer review,
- expert review,
- learner review,
- usability testing, and
- alpha, beta, and field tests of prototype learning designs.

**Peer review** is most often done during the earliest stages of learning design, but it can continue through the entire lifespan of a project. It can range from informal activities (asking a colleague to react to your design ideas over coffee) to more formal activities (providing formative feedback via a tool like Cloudworks<sup>2</sup> to the prototypes developed by others).

**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 a learning design's content. Learning theory experts can assist by critiquing the potential effectiveness of the pedagogical dimensions of a design. Graphic designers can suggest how to enhance the aesthetics of the look and feel of an interactive program. Expert teachers can help you anticipate the logistical requirements for successful implementation of a learning design in schools or universities.

With respect to formative evaluation, an expert is anyone with specialized knowledge that is relevant to the design and implementation of your learning design. 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 seek to utilize both internal (members of the design team) and external (non-members of the design team) 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 2 presents an expert review form to guide design experts with experience in interactive multimedia when they critique a prototype OER.

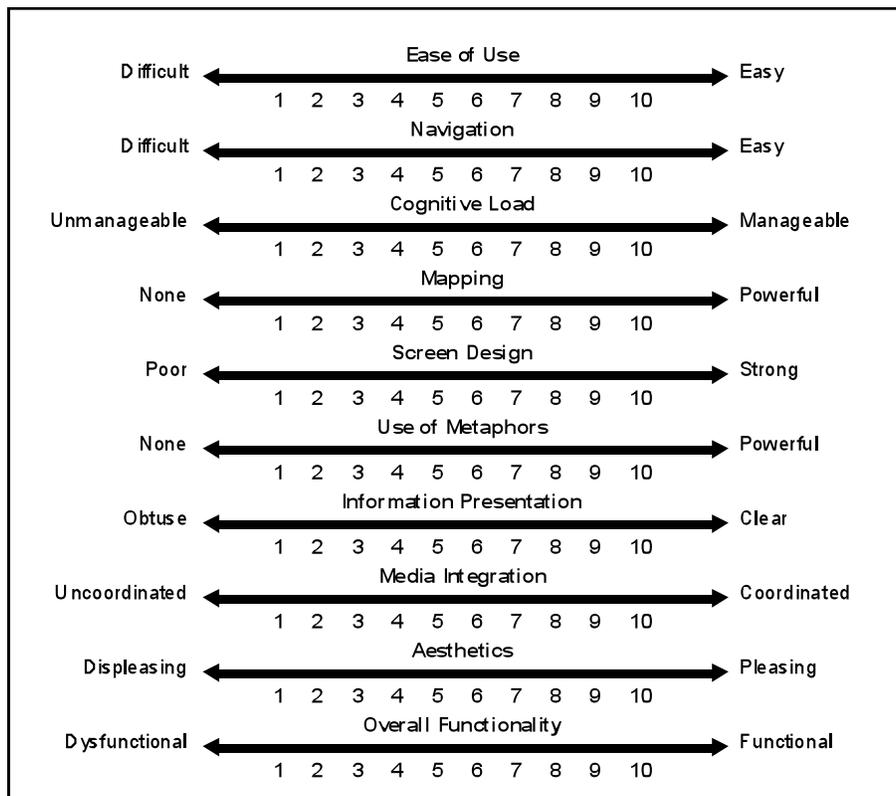
If you must limit expert review, content experts are probably the most important expert sources of formative information in education settings. After all, if you do not get the content right, the eventual learners will be misled. One of the problems with many learning designs is that they have poor subject matter integrity because of a lack of expert review. This is a major challenge because so much material can be incorporated into a single learning design.

<b>Reviewer:</b> Dr. Ima Gladtohelp		<b>Due Date:</b> June 10				
Please circle your rating and write comments on each aspect of the Open Educational Resource (OER).						
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 OER. Use additional sheets to write comments.						
NA=Not applicable 1=Strongly disagree 2=Disagree 3=Neither agree/nor disagree 4=Agree 5=Strongly agree						
<b>AREA 1 – LEARNING DESIGN REVIEW</b>						
1. This OER design provides learners with a clear knowledge of the program objectives.	N/A	1	2	3	4	5
2. The learning interactions in this prototype OER design are appropriate for the objectives.	N/A	1	2	3	4	5
3. The pedagogical design of the OER is based on sound learning theory and principles.	N/A	1	2	3	4	5
4. The feedback in the OER learning design is clear.	N/A	1	2	3	4	5
5. The pace of the learning interactions is appropriate.	N/A	1	2	3	4	5
6. The difficulty level of the learning design is appropriate.	N/A	1	2	3	4	5
<b>Comments:</b>						
<b>AREA 2 – COSMETIC DESIGN REVIEW</b>						
7. The screen design of the OER follows sound principles.	N/A	1	2	3	4	5
8. Media is appropriately integrated in this OER.	N/A	1	2	3	4	5
9. The screen displays are easy to understand.	N/A	1	2	3	4	5
<b>Comments:</b>						
<b>AREA 3 – FUNCTIONALITY REVIEW</b>						
10. This OER operated flawlessly.	N/A	1	2	3	4	5
<b>Comments:</b>						

**Figure 2.** Sample expert review form for a prototype learning design or functional OER.

In addition to learning design and content experts, people with special expertise in human-computer interface (HCI) design and the aesthetics of online learning programs can provide useful expert reviews for interactive learning resources. For example, if your team doesn't include an actual art director who is responsible for the look and feel of the design, it may be useful to ask other graphic artists to critique your prototype. Of course, you would not decide to make major changes in the design elements of a learning program based on the opinions of just one or two experts because aesthetic appeal is more subjective than many other criteria to be reviewed.

Experienced learning designers are often the best experts for reviewing user interface issues, but there are people who specialize in HCI issues per se. Figure 3 presents a user interface instrument that can be used to guide reviews provided by expert designers or experienced users of OERs.



**Figure 3. Sample user interface review form for an OER.**

Enlisting the help of 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 external content experts. The costs for SMEs to review learning designs can vary widely depending on the field. Recruiting graduate students and academic staff can be a much less expensive source of expert review.

**Learner review** is based on the analysis of learner behavior during the use of a learning design. The perspectives of “experts” are valuable, but the opinions of the target audience for your learning design are equally important. It is critical to ensure that learner differences are accommodated to the extent possible, and thus it is important that learner review be done with a sample of people whose background knowledge and expectations approximate those of the final intended learners. During learner review, learners should be allowed to work in realistic conditions, with minimal interruption from an observer, in order to accurately replicate the intended context of use.

Suppose you are designing an interactive educational resource such as a game for use in schools. In that case, valuable information for making decisions about improving the game can be derived from systematic observations of learners while they use it. Observations can be conducted at your development site or on-site at a school. Learner opinions, actions, responses, and suggestions will provide you with practical guidance in making decisions about improving the game. Of course, you would also want the teachers who must implement the learning game 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 other school specialists is also advised in this context. Few innovative games designs have been successfully integrated into schools, a problem that might have been reduced by more inclusive formative evaluation.

Observations of learners engaging with your prototype learning design 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 4. During your observations, you will see learners doing things you never expected. When you see learners making mistakes, your first instinct may be to blame the mistakes on their inexperience or lack of intelligence. This is wrong. The purpose of observing learners is to see what parts of your program might be difficult or ineffective. Therefore, if you observe someone struggling or making mistakes, attribute the difficulties to faulty design, *not* to the learner.

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 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.

## 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 interactive program, and you will be able to use that information to improve it.

1 – Introduce yourself	Make the session and task as welcoming as possible. Remember, this user observation is not designed as a controlled experiment, so keep the environment friendly.
2 – Describe the general purpose of the observation	<p>Try to make the participant feel at ease by stressing that you're trying to find problems in the program. For example, you could say:</p> <p><i>You're helping us by trying out this e-learning program. We're looking for places where the program may be difficult to use. If you have trouble with some tasks, it's the program'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 program. Remember, we're testing the program, not you.</i></p>
3 – Tell the participant that it's OK to quit at any time	<p>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:</p> <p><i>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.</i></p>
4 – Talk about the equipment	Explain the purpose of each piece of equipment in the observation room and how it will be used in the test (hardware, software, video camera, microphones, etc.).
5 – Explain how to "think aloud"	<p>Ask participants to think aloud during the observation, saying what comes to mind as they work. You'll find that listening to learners as they engage with your e-learning program 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:</p> <p><i>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 e-learning program. 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 go through the program. If you forget to think aloud, I'll remind you to keep talking. Would you like me to demonstrate?</i></p>
6 – Describe why you will not be able to help.	<p>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 in the e-learning program 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:</p> <p><i>As you're working through the e-learning program, 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 the recording equipment. When you've finished all the program, I'll answer any questions you still have.</i></p>

(Continued on next page)

**Figure 4.** Sample User Observation Protocol.

7 – Describe the tasks and introduce the program	<p>Explain what the participant should do first, second, third, etc.</p> <p>Give the participant written instructions for the tasks.</p> <p><b>Important:</b> If you need to demonstrate your program 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 interactions, don't show them how to engage with these particular interactions before the test.)</p>
8 – Ask if there are questions	<p>Before you start, make sure the respondent knows your expectations, then begin the observation.</p>
9 – Conclude the observation	<p>When the test is over:</p> <ul style="list-style-type: none"> <li>• Explain what you were trying to find out during the test.</li> <li>• Answer any remaining questions the participant may have.</li> <li>• Discuss any interesting behaviors you would like the participant to explain.</li> </ul>
10 – Use the results	<p>To get the most out of your test results, review all your data carefully and thoroughly (your notes, the video recording, the tasks, etc). Look for places where participants had trouble, and see if you can determine how your program could be changed to alleviate the problems. Look for patterns in the participants' behavior that might tell you whether the program is understood correctly.</p> <p>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.</p>

**Figure 4.** Sample User Observation Protocol (continued).

### Formative Review Log

**Program:** Learn or else!

**Reviewer:** Smyth

**Date:** May 15

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

**Figure 5.** Formative evaluation review log for OER.

**Usability testing** is an important method in the design of interactive programs. User interface expert Ben Shneiderman<sup>3</sup> maintains that usability is a combination of the following characteristics:

- ease of learning,
- high speed of task performance,

- low error rate,
- subjective satisfaction, and
- learner retention over time.

Too many formative evaluations 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 learners. Time with learners is often limited; it is not a free resource. In addition, learners may find it difficult to visualize how a prototype learning design such as an OER could be designed differently and they therefore may 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 evaluators using protocols such as heuristic evaluation. At least three evaluators with a mix of experience and expertise are required for heuristic evaluation because fewer will not be able to identify all the usability problems. Other types of usability evaluation will require more participants.<sup>4</sup>

As part of the learning design process, prototypes are commonly developed and tested. When developing any type of digital learning resource, it is important to begin usability evaluation at the earliest phases of design because, if left until just before an interactive program is released, there will be little chance to make any significant design changes.

An alpha version of a design is used in the earliest tests, conducted when the product is still at the early prototype stage. Alpha testing is generally done with in-house personnel. Beta versions are released to selected samples of learners for testing. The product is more refined at this point, nearing completion, but still needing to be tested and refined. The field version of a program 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 field versions of an interactive program is important.

Finally, although this brief paper is focused on formative evaluation in the context of learning design, it is worth noting that formative evaluation is also a major component of the more formal genre of R&D activity known as educational design research<sup>5</sup> (also referred to as design-based research or simply design research). Educational design research (EDR) entails the investigation of learning in practical educational settings by addressing a significant learning problem through the close collaboration of teachers and researchers. EDR projects usually last two or more years. EDR has two primary goals: first to create an effective intervention that directly addresses an important problem, e.g., a serious learning game that is intended to increase the interest of young minority students in careers in science and engineering, and second to extend theoretical knowledge related to learning design, e.g., the pedagogical principles underlying the design and use of serious games. The second outcome is usually represented in the form of design principles that others can use when they engage in learning design and/or EDR. It is this second goal of knowledge production that distinguishes EDR from learning design.

## Conclusion

This paper has introduced you to the concept of formative evaluation. There is a lot more to it than can be included in this brief paper. You may wish to consult the references<sup>6</sup> listed below.

## References

- <sup>1</sup>Flagg, B. N. (1990). *Formative evaluation for educational technologies*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- <sup>2</sup>Conole, G., Culver, J., Williams, P., Cross, S., Clark, P., & Brasher, A. (2009). Cloudworks: Social networking for learning design. *Australasian Journal of Educational Technology*, 25(5), 763-782.
- <sup>3</sup>Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (5th Ed.). Reading, MA: Addison-Wesley.
- <sup>4</sup>Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: The 10±2 rule. *Communications of the ACM*, 53(5), 130-133.
- <sup>5</sup>McKenney, S. E., & Reeves, T. C. (2012). *Conducting educational design research*. New York: Routledge.
- <sup>6</sup>Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Englewood Cliffs, NJ: Educational Technology Publications.