Guidelines for Designing Evaluations of Web-Based Instructional Materials

M.K. Iding  
*Education Psychology*  
University of Hawaii, Manoa  
miding@hawaii.edu

B. Auernheimer  
*Computer Science*  
California State University, Fresno  
brent@CSUFresno.edu

M.E. Crosby  
*Information and Computer Sciences*  
University of Hawaii, Manoa  
crosby@hawaii.edu

E.B. Klemm  
*Teacher Education and Curriculum Studies*  
University of Hawaii, Manoa  
klemm@hawaii.edu

Abstract

This paper is an overview of central issues for effective evaluation of web-based instructional materials. It describes relevant issues from the perspectives of practitioners in the fields of computer science, educational psychology, and teacher education. In addition, it describes educator concerns, reviews relevant aspects of HCI, usability, and educational research, and provides an overview of usability evaluation.

1. Introduction

User-centered design and usability evaluations have become common practices in many organizations, but they are still novel and the development of many web-based instructional materials do not accommodate these practices. Widespread inclusion of usability engineering methods for web-based instructional applications should be fostered by empirical studies validating these methods and case studies while addressing cost/benefit issues. This paper is an overview of central issues for effective evaluation of web-based instructional materials. It describes relevant issues from the perspectives of practitioners in the fields of computer science, educational psychology, and teacher education.

2. Teachers, Students, Classrooms

Designing Web-based tools for distance education is a challenge for educators, curriculum designers, and computer scientists. The development of viable, effective educational tools requires interdisciplinary collaboration. Since teachers and students are stakeholders in the design of these systems, their considerations and interests are input both for the formative evaluation and iterative design processes. Specific concerns of teachers include: effort required for implementation and deployment in the classroom, clear and accessible support, simplicity and transparency of the educational materials, and likelihood of students’ misconceptions as a result of interacting with the material. Furthermore, conformance to educational standards is important to teachers.

When evaluating Web-based educational materials, teachers consider planning, teaching, and assessment. Questions that educators ask include “what's being taught?”, “what is the standard?”, “what do I want the student to be able to do?”, “how will the learners show me what they have learned?”, and “how will it be taught?” [1].

Teachers and learners need tools in order to research, communicate with others, and to produce a product or performance [2]. Among other things, educators need to know whether Web-based lessons require direct teaching, whether the lesson’s organization is static or dynamic, and are links provided for additional information, review, extension, and supportive scaffolding.

If the teaching is to be done as distance learning via the web, educators need to determine the framework of the instruction. Will the student work independently, or collaboratively? For either independent or collaborative instruction, teachers need to know what tools are needed for communications, interactions with audiences and which media and formats are to be used.

If students are to collaborate, will they work synchronously or asynchronously? For synchronous instruction, teachers need to know if the students will work with others at same location or remotely. Furthermore, how will collaborators interact, and who is responsible for what-and-when? If the students are independent, teachers still need to know what
technological skills are necessary or expected and what are the parameters for carrying out the task.

For students, a critical characteristic of Web-based material is whether the performance expectation is clear. Are the lesson’s tasks mapped out, or are students using the Web-based system to state a problem, develop, and carry out their own solution?

Finally, how big is the Web-based material? Is it a comprehensive system spanning a topic through an entire term of instruction? Alternatively, Web-based material can be very specific simulations or demonstrations.

Our first guideline reflects the need to ground evaluation of Web-based educational materials in the experience of teachers and students:

**Guideline:** The language used in the specification, design, implementation, and use of Web-based education materials must be authentic to educators, yet capable of meeting the needs of programmers who must work within implementation resource constraints.

In the next sections we describe aspects of HCI and web research that have implications for the evaluation of educational courseware. Then we briefly summarize relevant educational research from multimedia. We describe principles or guidelines that emerge from this research.

### 3. Implications of HCI Research

Human-computer interaction (HCI) researchers study the way the Web is used for communication, research, and commerce. This results in guidelines for interfaces that span low-level web page layout to abstract models of tasks and interactions. Some of these findings and guidelines may be helpful in the evaluation of visualizations and courseware used in distance education.

Shneiderman’s “eight golden rules of interface design” [3], pp. 74-75, are widely known and include recommendations about interface consistency, shortcuts, dialogs, error prevention, reversal (“undo”), and loci of control. Basic texts like Shneiderman’s [3] or Preece’s [4] books can provide education evaluators with a useful introduction to basic aspects or characteristics that are used in effective user interface design.

More specifically, several areas of HCI research and their possible application are discussed below. Possible guidelines based on each area are given.

#### 3.1. Modes and Modelessness

Users typically interact with computers through keystrokes, clicking, and pointing. Raskin [5] calls these actions *gestures*. Although real life is naturally moded - a gesture results in different actions depending on circumstance - designers generally minimize human-computer interface modes. Raskin [5], pp. 42, explains

A human-machine interface is *modal* with respect to a given gesture when (1) the current state of the interface is not the user's locus of attention and (2) the interface will execute one among several different possible responses to the gesture, depending on the system's current state.

For example, typing the 'a' key usually causes an 'a' to appear on the screen. Users expect a what-you-see-is-what-you-get (WYSIWIG) word processor to behave that way. But sometimes typing an 'a' might be the 'append text' command to the word processor instead of the literal character 'a'. For such a word processor, in WYSIWIG mode a keystroke means one thing, but in command mode it means something very different. It's through feedback that users know the mode (more derisively, "mood") of the system. Multi-moded systems, or those with little feedback, confuse users and increase their errors. For users of safety-critical systems, mode errors can be life threatening.

In the distributed classroom, mode errors can frustrate students and slow their progress. Gestures used by students interacting with visualizations or examples should evoke responses consistent with the same gesture used in related parts of the distributed classroom system. A common design goal for user interfaces is *modelessness*. Thus, a recommendation for evaluation of materials and systems for distributed learning is

**Guideline:** The number of modes encountered by students should be small with obvious feedback.

#### 3.2. Approaches to Evaluation

Given a guideline, how are Web-based education materials evaluated? Again looking to HCI research, Shneiderman [3], pp. 15, posits five factors as focal points of evaluation. These include: learning time, performance, error rate, “retention”, and subjective satisfaction. These are similar to Preece’s presentation of Shackel’s [6] five factors. Preece, however, includes flexibility, which she describes as “the extent to which users can adapt a system to new ways of interaction as they become more experienced” [4], pp. 47.

Preece [4] provides useful descriptions of types of evaluation that may be conducted and their advantages and disadvantages. For example, analytic evaluation consists of designers and others predicting performance
based upon specifications of interfaces. This is described
as more theoretical than expert evaluation which involves
experts trying to predict potential problems and possible
solutions, based upon their prior knowledge and interface
design or human factors expertise. By contrast,
“observational evaluation involves collecting data that
provide information about what users do when they
interact with an interface” [4], pp. 112. This may
include observing, recording, employing think-aloud
protocols, or “software logging” (the software logs user
responses and corresponding response times). Survey
evaluation is described as consisting of interviews
(structured or flexible) and/or questionnaires, with open
or closed questions. A useful discussion of these
descriptive methodologies is provided by Knupfer and
McLellan [7]. Finally, experimental evaluation enables
examination of specific variables in controlled ways.

Related to Preece’s expert evaluation method, Nielsen
[8] uses a “discount engineering” approach called
heuristic evaluation. A small group of experts inspects an
interface or system noting problems. A checklist of
common issues and a scale of severity guide the expert’s
evaluation. Nielsen shows that a small number of
evaluators (“about five evaluators, but certainly at least
three”) is a cost-effective way of capturing most usability
problems. The ten general usability heuristics Nielsen
suggests for the experts’ checklist can be easily modified
to for evaluation of distance education materials. For
example, some of Nielsen's heuristics and our comments
are:

Visibility of system status.
"The system should always keep users informed about
what is going on, through appropriate feedback within
reasonable time.” Students need both awareness of the
status of their interaction with the Web-based system,
and to see the results of their actions.

Consistency and standards.
"Users should not have to wonder whether different
words, situations, or actions mean the same thing." The
first two of Nielsen’s heuristics address mode errors as
well as ensuring that students receive appropriate
feedback.

User control and freedom.
"Users often choose system functions by mistake and
will need a clearly marked "emergency exit" to leave the
unwanted state without having to go through an extended
dialogue. Support undo and redo." As part of their
interaction and exploration, students will make mistakes.
Explorations should be encouraged, and students
reassured that they cannot break the system.

Recognition rather than recall.
"Make objects, actions, and options visible. The user
should not have to remember information from one part
of the dialogue to another." Students should spend
cognitive energy meeting the learning objectives, not
learning a web-based tool and its idiosyncrasies.

Aesthetic and minimalist design.
"Dialogues should not contain information which is
irrelevant or rarely needed. Every extra unit of
information in a dialogue competes with the relevant
units of information and diminishes their relative
visibility." Novices can be distracted by extraneous
information that educators and implementers think is
"cool." Designers must take care that the interface does
not overwhelm the content.

Guideline: A small group of educators, armed with a
heuristic checklist tailored to the subject and system at
hand, can be a cost-effective way of evaluating
curricular materials.

3.3. Active Web Users

Web users are goal-driven. Fifty-one percent users in a
recent study user the web to "evaluate multiple products
or pieces of information in order to help... make a
decision", 24 percent used the Web to "understand some
topic" [9]. Finding products, facts, documents, and
downloading were the remaining 24 percent. Further, 71
percent of users collected information via the Web by
"searching for multiple pieces of information".

Studies also show that users tend not to read page
after page of linear text, neither do they concentrate on
tracking study by the Pointer Institute [11] showing that
users looked at text before graphics, read shallowly, and
interlaced their browsing by switching from site to site.
Nielsen summarizes:

In other words, the most common behavior is to hunt
for information and be ruthless in ignoring details.
But once the prey has been caught, users will
sometimes dive in more deeply. Thus, Web content
needs to support both aspects of information access:
foraging and consumption. Text needs to be
scannable, but it also needs to provide the answers
users seek.

Nielsen recommends that Web sites designed
explicitly to "accommodate people who leave and return"
frequently. This includes incorporating re-orientation and navigation aids in a standard way to minimize confusion across Web pages.

The recent retrospective study of seventeen Web users for four months by Cockburn and McKenzie [12] provides further insights. The participants had over 80,000 visits to approximately 17,000 unique sites. Subjects averaged 42 page visits per day.

The most startling result was how many times users revisited web pages:

Previous studies have shown that revisitation (navigating to a previously visited page) accounts for 58% and 61% of all page visits. Our study show that page revisitation is now even more prevalent, accounting for 81% of page visits when calculated across all users.

Their results show that browsing is rapidly interactive:

Users often visit several pages within very short periods of time, implying that many (or most) pages are only displayed in the browser for a short period of time... the most frequently occurring time gap between subsequent page visits was approximately one second, and that gaps of more than ten seconds were relatively rare.

**Guideline:** Curricular materials should be "re-entrant", facilitating multiple repeat visits as a student's understanding and goals change.

### 3.4. The Web and college-level instruction

A recent study of near-parallel Web and face-to-face university lectures exposed a common aspect of human nature. For four months, the face-to-face lectures for a large, first year, university computing courses were captured digitally and quickly made available in the university library as CD-ROMs. Ninety-four of the course's 746 students expressed an interest in using the digital lectures. Only four students actually used the system during the first two months, and only two students used it the third month. The authors[13] explain:

As a result, the course lecturer sent an email message to the class asking whether it was worthwhile continuing to make digital lectures available. Much to our surprise (given the extremely low use of the system until then), fifty-three students replied, strongly urging us to continue... Despite the appeals for us to continue, only seven students subsequently accessed the CD-ROMS.

It's not necessarily the case of "build the curriculum and the students will come".

**Guideline:** Systems for the "flexible delivery" of education should play to students strengths, not to the weaknesses of human nature.

### 4. Principles of Multimedia Applied to Evaluation of Web-Based Material

In addition to HCI principles for the effective design of Web-based instruction, research from educational psychology has examined aspects of multimedia instruction that contributes to its efficacy. For example, Iding [14] reviewed aspects of multimedia that contributed to effective learning in science. Briefly, it appeared that much of this research could be subsumed under a “cognitive load” theoretical framework. That is, aspects of multimedia displays that reduced learners’ cognitive load were associated with increased learning. The theoretical concept of cognitive load is associated with Chandler and Sweller’s work [15, 16]. Specifically, they proposed that unnecessary integrations between information sources, and redundancies among information presented in different modes are detrimental to learning.

Iding summarized research that supported this framework. These principles can be used to guide the design of effective instructional devices for content other than science [14], pp. 409-410:

The first major point is that working memory is limited in the amount of information it can hold at any given time. Furthermore, less expertise, less spatial ability and limited reading skills can impede the process of holding large chunks, processing them quickly, and continuing on in the multimedia presentation to develop complete and workable situation models or mental models. Some aspects of textual and multimedia design that can reduce cognitive load on learners include:

- presenting smaller, more manageable chunks of texts-with-illustrations [17].
- presenting illustrations and relevant texts in proximity to minimize the visual search process [15, 16].
• minimizing extraneous information or sensory input [17].

• using auditory input to accompany diagrams, to minimize excessive resources that may be spent in visual search or construction of visual representation while at the same time reading accompanying text (a second and possibly competing visual process) [17].

Iding [14], pp. 410, argues “it would seem that providing a rather sparse, simplified environment would most effectively facilitate learning, at least among novices to particular content areas.” These observations lead to three guidelines:

**Guideline:** Multimedia environments should be designed to minimize cognitive load (i.e., unnecessary integrations or redundancies).

**Guideline:** Extraneous inputs should be minimized.

**Guideline:** Relevant visual, verbal, and auditory chunks should be small, and presented in proximity.

5. Conclusion

In this paper, we provided a broad overview of issues to consider in designing effective evaluations of web-based instructional materials. We first considered factors that would be important to educators adopting these tools, then we visited relevant HCI, educational psychology and web research that would provide descriptions of research-based factors to consider in evaluating instructional software. Finally, we introduced some relevant concepts with regards to the creation of usability evaluations. By incorporating research from diverse disciplines and areas we hope to provide a useful foundation for computer scientists, designers, and educational evaluators to dialogue in the design of effective instructional software usability evaluations.

6. Acknowledgments

Dr. Iding was partially supported by DARPA Space and Naval Warfare Systems via ONR grant number N66001981891. Dr. Crosby was supported in part by ONR grant number N00014970578.

7. References


