Improving the quality and effectiveness of computer-mediated instruction through usability evaluations

Michael S. Crowther, Chris C. Keller and Gregory L. Waddoups

Michael S. Crowther is a QA/Usability Specialist, Chris C. Keller is the Testing and Quality Assurance Manager and Gregory L. Waddoups is the Director of Evaluation at the Center for Instructional Design (CID), Brigham Young University (BYU). Address for correspondence: Dr Gregory Waddoups, 3814 HBLL, Brigham Young University, Provo, Utah 84602, USA. Tel: 1 (801) 422-4734; fax: 1 (801) 422-8910; email: greg_waddoups@byu.edu

Abstract
Brigham Young University’s Center for Instructional Design (CID) creates online courses and multimedia instructional applications for university faculty. This paper asserts that including usability testing as a part of evaluation improves the quality and effectiveness of computer-mediated instruction. The paper describes the fundamental purpose and functions of usability testing, and also distinguishes between different forms of evaluation: accessibility, quality assurance (QA), usability and implementation. Through a detailed case study, we describe how usability testing improved the quality of a computer-based chemistry course and facilitated a clearer analysis of the learning effectiveness of this course.

We alone know what serves the true welfare and benefit of the State and People.
King Frederik VI of Denmark (Nielsen, 1994, 268)

Introduction
With the rapid increase of technology and the ever-rising number of students at Brigham Young University (BYU), computer-mediated learning is growing in importance. As more students turn to computer-mediated instruction, the demand and necessity for product quality increases. Indeed, poorly designed instructional applications are unlikely to be instructionally effective; therefore, those designing computer-mediated instruction have a moral, ethical and pedagogical obligation to create usable applications. Systematic formative and summative evaluation is the best way to determine an application’s quality. In addition to traditional evaluation at the end of the development cycle, however, developers need to incorporate more focused usability evaluation from the beginning of the development process. This paper focuses on a case
study outlining the relationship between usability testing and educational evaluation, explaining how they are integral to improving the quality and effectiveness of computer-mediated instruction. Additionally, this paper explains the challenges associated with integrating usability evaluation into the process of designing and developing computer-mediated courseware.

The Center for Instructional Design (CID) is a production department charged with creating online courses and faculty multimedia projects (see http://cid.byu.edu for examples of CID’s projects). Therefore, it produces and supports the design, development and implementation of technology-enhanced instruction for on-campus and off-campus students. The CID has 26 full-time employees and 150 student employees. Among these employees are professional instructional designers, programmers, art and media production specialists, production managers and educational evaluators. Representatives from these functional groups work together as members of design teams to complete CID projects.

Until recently, product evaluation at the CID has occurred during the implementation phase, after the production team had finished development and faculty were implementing the product. Evaluation often pointed out usability problems, but at this stage the production team could fix only major product flaws, leaving many minor problems untouched. The expense of these late-stage fixes and the lower quality of the final application prompted the evaluation team to include testing earlier in the development process. This change included hiring a full-time testing and quality assurance (QA) supervisor to work on the instructional design and development team. This person assumed the responsibility for usability testing during production. Several projects later, the results of his usability work has confirmed the need for and benefit of early usability testing—and the need to convince other team members of the importance of usability evaluation.

Usability
The International Organization for Standardization defines usability as ‘the effectiveness, efficiency, and satisfaction with which a specified user can achieve specified goals in particular environments’ (Dillon, 2001). The following ‘six general attributes: utility, learnability, efficiency, retainability, errors, [and] satisfaction’ extend this definition (Sun Microsystems, 2001, 1). Table 1 explains these attributes.

Usability testing of computer-mediated instructional products must include careful attention to these six attributes. As will be clear from our case study, the quality and effectiveness of an instructional product depends on accounting for these attributes of usability.

Usability has two primary approaches: expert analysis and end-user testing. A usability professional conducts the expert analysis, identifying problems with the application based on usability guidelines or heuristics and past experience with similar products. End-user testing involves end-user representatives using the application at any stage of
Improving the quality and effectiveness of computer-mediated instruction

Table 1: Usability attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Usability attempts to measure:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td>How useful the application is</td>
</tr>
<tr>
<td>Learnability</td>
<td>How easy it is to learn</td>
</tr>
<tr>
<td>Efficiency</td>
<td>How much it streamlines the work process</td>
</tr>
<tr>
<td>Retainability</td>
<td>How easy it is to remember application operations</td>
</tr>
<tr>
<td>Errors</td>
<td>How many errors occur during work</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>How much people enjoy using the application</td>
</tr>
</tbody>
</table>

development and giving feedback to the design team. Users test the design team's assumptions on issues of order, navigation, layout, instructional aids, pace and flow of the material, and aesthetic appeal. The usability professional guides the process of collecting this user input, analyses the data, reports to the design team and facilitates the process of requesting changes and validating fixes with further testing.

Evaluation distinctions
Evaluation can occur during both the development and implementation of computer-mediated instructional material. It also occurs on a continuum from formative to summative evaluation. The continuum applies to both the development and implementation phases.

In creating instructional applications, evaluation is more formative the closer it occurs to the beginning of the development process. The farther along in the development process, more concrete and less amenable to change the application becomes; thus, the evaluation becomes more summative. At the end of the development process, the final evaluation gives a summative judgement of the application's value. Implementation evaluation information overlaps with usability and follows a similar process. Table 2 summarises these two processes. The shaded box suggests that usability evaluation should occur as a formative evaluation early in the development process.

Usability testing is not the only form of evaluation that CID evaluation teams use to improve the quality of computer-mediated courseware. The evaluation and QA team performs several other tests on CID projects. For example, a member of the QA team conducts accessibility testing as the product nears completion. This test ensures that colour-blind, blind, deaf and physically disabled individuals can use the application. To simplify the process, the team checks the application against the W3C consortium guidelines for digital media production. Among other techniques, the tester uses screen reader software to navigate the application. By designing the application to meet standard specifications for use by the blind, the developer increases the probability that the application will be accessible to disabled users. The tester also ensures that audio material has captions and the application uses appropriate colours to mitigate problems for the colour-blind.

### Table 2: Evaluation phases

<table>
<thead>
<tr>
<th>Formative</th>
<th>Development</th>
<th>Summative</th>
<th>Implementation</th>
</tr>
</thead>
</table>

Drawing conclusions based on early user testing

Applying user suggestions to development

Applying initial results of live testing to implementation/administration of the application to better suit users (usually little change to the application)

Testing and reporting application readiness (mainly technical issues)

Gathering conclusive results of application effectiveness after a period of usage

Applying these results to "version 2.X"

(This phase affects only the future product)

The tester also performs a functional test. This test requires that product stakeholders designate system requirements for supporting the application. Accordingly, the tester checks the application using different software components, including various browsers, plug-ins and add-ins. Finally, the evaluation team may also test the application with different hardware and operating system configurations.

The evaluation team may also perform a variety of tasks to measure the product quality, according to stakeholders' specifications during the implementation stage. They evaluate how well the application meets stakeholders' needs and design objectives. The evaluation team's work is more important for guiding administrative and implementation decisions, rather than for revising the current version of the application. These evaluations are formative and summative, and often include long-term, extensive and multi-method evaluations. The more extensive evaluations are in contrast to the quick, less formal tests that the usability team conducts. However, through the more extensive formative and summative evaluations the evaluation team can discover design problems and make recommendations for future changes. However, the problems they discover and changes they recommend are less likely to appear until the CID develops subsequent versions of the application. Regardless of the product quality, the evaluation team can improve the application's implementation through communicating openly about its challenges and strengths, training users and planning for product support.

The entire evaluation process ties together experts who monitor the quality and effectiveness of the CID's products. The QA team consults with the production team on the use of technology, measures application quality and ensures the application's availability to different users. Usability measures user input against design assumptions. Evaluators manage stakeholder expectations and report development trends and

Table 3: Elements of evaluation

<table>
<thead>
<tr>
<th>Represents</th>
<th>Quality assurance</th>
<th>Usability</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>End-users</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Stakeholders</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

implementation successes and failures to key decision makers. Table 3 sums up the distinction between these elements of the evaluation process.

Why usability evaluation is important to computer-mediated instruction

The CID is responsible for creating instructional applications that provide educational benefit comparable to a classroom experience. Obviously, usability problems can impair student performance (Tselios et al., 2001). Even considering the economic ramifications, a poor interface that prevents people from buying a consumer product has fewer serious moral and ethical implications than an interface that impairs a student's learning.

One goal of computer-mediated instruction is to encourage higher-order thinking. For example, ChemLab is a CID-developed chemical education simulation that lets students conduct multiple virtual chemistry experiments. Using the programme, students can focus on problem solving, experimenting and interpreting results (higher-level thinking) instead of concentrating on simple recipe following ('cookbooking') or the time-consuming physical setup details that are necessary procedures in a physical lab room. However, to measure ChemLab's success in producing higher-level thinking skills, an evaluator must be able to analyse how usability issues affect student learning. Indeed, if usability problems impair student performance, it is difficult to assess the true educational value of the application (Kim et al., 2001). Poor student performance demonstrates that the product is not achieving the design team's objectives, but it does not identify whether the fault lies with the educational concept, the presentation, the interface or all three. Iterative user testing can distinguish between conceptual, pedagogical and interface problems, thus preventing poor design from skewing evaluation measurements of educational effectiveness (2001).

As Johnson (2000, 13) says, 'Software should neither be designed for the user nor by the user, but rather with the user'. That means with the user literally, not just 'with the user in mind'. Most recently, CID instructional design teams have learned about the benefits of involving users early through the development of a computer-based introductory chemistry course (Chemistry 105). The development of this course has become a significant case study that validates the importance of conducting usability testing during early stages of the development process. The following case study describes the ways in which usability evaluation improves the quality and effectiveness of an online introductory chemistry course and the difficulties often associated with implementing a usability process.
Case study: Chemistry 105, introductory chemistry course

Product background

'Chemistry 105' (Chem 105) is an introduction to basic chemistry that the CID is designing and developing as an online course. This course is part of a larger programme in which high-enrolling general education courses are redesigned to use technology to improve the efficiency and effectiveness of teaching and learning in these large lecture courses. The development of these courses requires the collaboration of an instructional design team that consists of instructional designers, programmers, artists, project managers and evaluators.

The user interface (UI) in Chem 105 is a futuristic lecture hall (see Figure 1). The professor appears on the stage at key moments to introduce material. At other times, he continues teaching in a small video window in the top left of the screen (see Figure 2), while textual material, animations or interactive learning objects appear in the centre of the lecture-hall stage. Navigation buttons retract in and out of a framed box in the bottom left of the screen. A chalkboard, whiteboard and video/animation screen in the background give the lecture hall a familiar, classroom feel. As the lesson proceeds, those boards zoom out from the background to fill the stage (see Figure 3).
Some objects on the boards are linear animations, synchronised with the video and audio so that the student cannot directly control them. The speed of these objects is dependent on the timing of the lesson presentation. Some examples of the animations are chemistry formulas derived from beginning to end (see Figure 6), or a visual reminder that the concept of planets orbiting a sun resembles the relationship between cell particles orbiting a nucleus (see Figure 7). Other objects are interactive and interrupt the linear progression of the lesson, requiring students to interact with the objects to move forward (see Figure 4). Objectives precede each interactive object; questions and answers follow (see Figure 5). There is no limit to the time students may take to work with the interactive objects.

The overall goal of this course is to create an immersive learning environment for students to explore principles of chemistry. The team’s usability work on the project included an expert analysis of the low-fidelity prototype and, months later, a high-fidelity functional prototype of one lesson within this course.

**Testing technique**
From the beginning of the project, the instructional design team’s overriding objective was to create an application that would improve students’ learning in introductory
chemistry, specifically focusing on concepts that were particularly difficult for beginning chemistry students. The usability team's evaluation efforts intended to measure the design team's success in meeting their objective. In this case, the usability work was a subset of an overall effort, concentrating on the student's use of the application and the extent to which the multimedia lessons increased student's learning of complex chemistry principles.

The chemistry course development project began before the CID adopted its current usability practice, so the usability team members did not participate in the initial design of the UI, nor did they contribute to designing the information flow of the lesson. Instead, the usability efforts began with an expert analysis of the UI at the low-fidelity prototype stage. This expert analysis consisted of the tester viewing screen captures of the UI in the different stages of the lesson (much like the screen shots included in Figures 1–7) compiled in a Microsoft PowerPoint slide show. Though this gave the usability team an opportunity to critique some design and layout standards, they could not evaluate the user interaction, information flow or user control of the pace of the lesson. The usability manager made a series of suggestions to improve the product, but the design team did not invite him to participate in the discussion and resolution of
these problems. Instead, the team decided to continue with development without following the advice of the usability expert. Unfortunately, this is not an isolated incident. Our usability efforts at the CID have demonstrated the difficulty of convincing instructional design team members of the need to include usability evaluation early in the process. They often reject suggestions and refuse to implement changes that could improve the usability of a product if those suggestions require alterations in the original design. In this instance, making the suggested improvements would have significantly changed the original visual design of the chemistry lesson.

Later, as the first lesson prototype neared completion, the usability specialist had another chance to work on the project. By that time, the usability process had proved its value in improving other projects at the CID, and these successes were creating a culture of formative evaluation within the CID design process. As a result, the Chem 105 production team invited the usability team to meetings and asked the members to review objects and give feedback concerning various elements of usability. The usability team also helped develop plans for conducting high-fidelity usability testing at the completion of prototype development. The high-fidelity prototype was a fully functional lesson within the course.
Although the increased involvement of the usability team was a move in the right direction, it happened late in the production timeline, which minimised the positive effect the usability testing could have on the prototype. The usability team for this project included one usability expert and one evaluator who together performed 10 usability tests with students concurrently enrolled in a section of Chem 105. These students had recently completed the same course material covered in the prototype computer-mediated lesson. The team followed a similar test plan with each participant. They told each participant to use the application as though the testers were not there. If the participant became confused, the usability expert asked the participant to explain what happened, what would help resolve his/her confusion and what his/her next action would be. The team told the participant to stop and review items or move forward, skipping items as needed. They asked the participant to speak aloud as he/she made decisions. The usability expert also told the participant that navigational clues would be available when the participant expressed frustration and could not navigate through the course.

The participants spent 45 minutes moving from the beginning of the chapter through several animations and interactive objects. Along the way, the team recorded participants' verbal and non-verbal feedback and questions. At times, the evaluator asked
Factual questions to measure participants' comprehension of the material or asked behavioural questions about participants' decisions or reactions to options presented on-screen. During the last 15 minutes, participants answered open-ended questions about the experience, the lesson material and the online approach to chemistry.

The team's findings revealed much about the product, as well as the ways in which usability influences the measurement of learning in computer-mediated learning environments. Generally, the participants liked the idea of making more material available online, saying it was both entertaining and stimulating. In particular, they liked the flexibility it gave them to study at their convenience. They liked the interactive objects most and wanted to spend time outside of class using them to reinforce difficult principles.

They generally viewed this product as a supplement to the classroom work rather than a replacement. They also had difficulty with the navigation. For example, they expected a table of contents but did not know where to find it; the two small tabs on the top and to the right of the box in the lower-left corner of the screen were not enough indication of a table of contents. Once they found the navigation interface, the icons confused them. The pace of the material also frustrated them. At times they wanted to stop and
repeat small chunks of material, and at other times they wanted to skip forward to specific objects or past material they already knew. The navigation let them move forward and backward only by subsections, some of which lasted up to 10 minutes (see Table 4).

Student learning
The team members also discovered that their high-fidelity testing gave only preliminary ideas about the application’s ability to improve participants’ overall ability to learn. The functional and navigational challenges of the application impeded the team’s ability to even begin to draw serious conclusions about actual student learning. Participants spent the majority of their time talking about challenges with finding objects, moving forward and backward and controlling the pace of the material. Only occasionally, and especially on the self-paced learning objects, did they say much about learning. In short, the usability difficulties inherent in the UI impeded students’ ability to engage in significant learning and prevented the evaluation team from measuring whether the lesson prototype promoted higher-order thinking.

This high-fidelity user testing revealed the importance of usability testing for assuring the quality and effectiveness of computer-mediated courseware. Indeed, in this case the
Table 4: Results of high-fidelity user testing

Positive user responses
- Participants generally liked the look and feel of the virtual lecture hall.
- They liked the flexibility to study on their own time and wanted to be able to control the pace of the material.
- They liked the audio/video that accompanied the material.
- They were captivated and entertained by the synchronisation of different media in the material.
- They liked the interactive objects that gave them extra time and exposure to learn difficult concepts.
- They liked the animations that clarified the instructor's lecture.

Negative user responses
- Participants quickly became frustrated trying to discover how to navigate.
- They had difficulty remembering the meaning of the icons.
- They were impatient with the pace of the lessons. They wanted to control the speed of the audio and video components.
- They did not like the lock-step approach to some of the objects and associated questions that they had to answer before moving on.
- They wanted a way to move forward or backward in small increments. Instead, the forward and back buttons took them to subsections, which could last up to 10 minutes.
- The nature of the application (CD-based) obligated them to have access to a computer. This dependency on a computer made 'quick studying' while in the hall, waiting in line or prior to class very difficult.
- The instructions for interactive objects confused them. They often could not remember them once they started and they did not have a way to return to them quickly.

difficulties associated with using the product limited the learning effectiveness of the courseware. During the low-fidelity test, the usability expert cautioned the design team about potential problems with navigation and usability of the product, but the design team insisted on maintaining many of the problematic design features. This situation reveals a tension that often exists between those responsible for the aesthetics of courseware and those responsible for usability. It is clear, however, that the early low-fidelity usability testing did reveal interaction problems that became apparent again in the high-fidelity test. Ideally, along with the expert analysis of low-fidelity prototypes, the team would have performed a more formal low-fidelity test early in the design phase, creating a paper sketch of each interactive object and its features. Nevertheless, the low- and high-fidelity testing performed by the team validated the CID's improved usability process and encouraged them to think of ways to involve users in the earliest stages of design.

Test conclusions
Though students were initially excited about the flexibility and options the CD-based chemistry course offered them, they quickly became frustrated with the confusing demands of navigating through the course. They were frustrated that the lesson ran like a video with limited rewinding ability. At times they expressed boredom with the slow pace of the lesson; at other times they were frustrated that they had missed

something important and could not rewind just a little to hear that phrase or word. The transitions between video clips and the associated learning objects often confused them. They expected more instructions throughout and expressed a desire to have control over the order and pace of the material.

It was especially interesting to see how student opinion differed from the design team's ideas. The design team hoped that course modules would replace both the textbook material and the face-to-face classroom experience. Though a few students thought the modules could replace a lecture here or there, not one of the test participants agreed that the computer-mediated approach could effectively replace the entire course experience. Instead, they saw the modules as a supplement to the textbook and in-class sessions, not as a replacement for them. Aligning the goals of the design team and the goals of the end-users is an important role of the usability and evaluation team.

The conclusions from the high-fidelity user test led the design team to believe that user testing is critical earlier in the design process. As the Chem 105 project continues, usability is now an integral part of the development process, which ensures that the usability team has authority to bring in users to validate future decisions. The design team is expected to take their suggestions seriously. In this case, the evaluation team concluded that the functional challenges impeded their ability to measure learning, and though using participants who already knew the material sped up the testing, a more thorough test, with students who are learning the material for the first time, will likely reveal much more about learning. With the resolution of the functional problems, the team hopes to measure learning in the next phase of testing, which will happen during the summer of 2003.

**Conclusion**

The CID has begun a more concerted effort to address usability issues at the beginning of project development, based on experiences with products such as Chem 105. This process has taken time, but demonstrating to instructional design teams the merit of early involvement of the usability team has facilitated its implementation, as has documenting the potential difficulties of ignoring or neglecting usability. This process of change often is accompanied by tension between those interested in the look and feel of the course and those concerned with usability. As the Chem 105 case demonstrates, the process of instituting early user testing into the development process has not been easy, and has required a significant change in the culture of quality within the instructional design team. Indeed, as the literature suggests, the CID discovered in its Chem 105 user testing that application problems can affect student performance and weaken evaluators' ability to draw conclusions about the level of student learning associated with the computer-mediated courseware. Unlike traditional classroom teaching, which is generally accepted as sufficiently effective, institutions developing computer-mediated instruction are under pressure to 'prove' that their products can be at least as effective as traditional classroom techniques. Thus, the necessity of providing accurate measures of actual learning makes usability and application quality especially important to those involved in the design and development of learning objects and
multimedia courseware. At the CID, the design teams are learning that it is best to find out how users understand applications, rather than to assume that they know what best serves the needs of learners.

References
