DOCUMENT RESUME

ED 466 153 IR 021 223

AUTHOR	Dimitrova, Maia; Sharp, Helen; Wilson, Stephanie
TITLE	Are Experts Able To Predict Learner Problems during Usability Evaluations?
PUB DATE	2001-06-00
NOTE	7p.; In: ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (13th, Tampere, Finland, June 25-30, 2001); see IR 021 194.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Software Development; Computer System Design; *Courseware; Designers; *Evaluation Methods; Foreign Countries; *Instructional Material Evaluation; Instructional Materials; Material Development; *Multimedia Materials
IDENTIFIERS	England

ABSTRACT

The usability of Instructional Multimedia (IMM) applications is vital for their success and for the satisfaction of their users, as the confusion resulting from using poorly designed programs can be particularly detrimental to learning performance. A number of approaches for expert-based evaluation of IMM have been proposed during the past few years. However, there is little evidence in the literature regarding how effective they are, especially in identifying real learner problems. This paper reports an empirical study that assesses whether experts can predict the problems experienced by students. The evidence suggests that expert evaluators, although successful in predicting usability problems, still have difficulties identifying certain types of learner problems, such as comprehension and learning support. The paper concludes that expert evaluations do not eliminate the need for tests with actual learners. Ways of improving their effectiveness are suggested. (Author/AEF)

Are Experts Able to Predict Learner Problems During Usability Evaluations?

Maia Dimitrova
Centre for Human-Computer Interaction Design
City University London
London EC1V 0HB, UNITED KINGDOM
maia@soi.city.ac.uk

Helen Sharp
As above
hsharp@soi.city.ac.uk

Stephanie Wilson
As above
steph@soi.city.ac.uk

Abstract: A number of approaches for expert-based evaluation of Instructional Multimedia have been proposed during the past few years. However, there is little evidence in the literature regarding how effective they are, especially in identifying real learner problems. In this paper we report an empirical study which assesses whether experts can predict the problems experienced by students. The evidence suggests that expert evaluators, although successful in predicting usability problems, still have difficulties identifying certain types of learner problems, such as comprehension and learning support. We conclude that expert evaluations do not eliminate the need for tests with actual learners, and suggest ways of improving their effectiveness.

Introduction

The usability of Instructional Multimedia (IMM) applications is vital for their success and for the satisfaction of their users, as the confusion resulting from using poorly designed programs can be particularly detrimental to learning performance. To avoid this, the evaluation of such software should assess how successful learners are at achieving learning tasks, and not just how effective and efficient they are while interacting with the application (Squires and McDougall, 1996). To measure the former, 'before' and 'after' knowledge tests are typically performed with learners (Draper et al, 1996). However, learner tests have been found to be expensive in terms of the time and effort required, and recruiting users can also be problematic (Dimitrova and Sutcliffe, 1999). Due to these problems, involving learners may not be feasible in many projects, and alternative evaluation methods need to be explored.

A number of expert-based methods for the evaluation of IMM have been proposed in the past few years, such as Interactive Multimedia Checklist (Barker and King, 1993) and Multimedia Taxonomy (Heller and Martin, 1999). However, there is little evidence in the literature regarding their effectiveness, especially in terms of identifying real learner problems. In a review of expert- and learner-based evaluations, Reiser and Kegelmann (1994) criticise the expert-based approaches for having poor reliability as the majority of them required evaluators to make subjective judgements. The authors also acknowledge that teachers and students rate software differently, however they do not explain the nature of these differences. Tergan (1998) also criticises checklist-based approaches for their inability to assess the instructional efficacy of the software. Although these reviews are useful, they do not provide empirical data to support the conclusions reached. The reviews also do not give details about the differences between expert and learner evaluations.

In this paper we report an empirical study which assesses the effectiveness of expert predictions using three different evaluation methods by asking the question whether experts can predict real learner problems. To address this, we compare the results produced by two types of expert - subject matter specialists and multimedia designers - to those from learner tests and discuss their similarities and differences in terms of the number and the type of problems predicted.

Study Design

The IMM Application

One section of a multimedia environment for learning Mathematics at university level was evaluated. The selected topic covers the principles of exponential functions and the three types of transformation of these functions – Scaling, Reflection and Translation. A series of 23 screens presents the Maths content in textual, graphical and animation formats. Interactive quiz-like tests are also provided, which enable the users to plot exponential graphs and test their knowledge of transforming them.

Learner Tests

Four students undertaking a course in Mathematics at City University London were involved in the learner tests. Before the experiment, pre-exposure knowledge tests were administered to establish students' prior knowledge of the material. Each student was then given four tasks to perform, which consisted of learning about the principles of exponential graphs and exploring the three different types of transformation. During the usability tests, the students were asked to think aloud while performing each task. After the students had completed the tasks, they were interviewed by the experimenter to determine their attitude towards different aspects of the application. The student sessions and the interviews were recorded on video. At the end, comprehension tests were administered to reveal the knowledge students gained while working with the software. The material covered by the students was divided into 20 knowledge propositions, of which the students were expected to have a reasonable level of comprehension after working with the application. Each proposition was tested in the post-exposure comprehension tests.

Expert Evaluations

Ten experts took part in the expert evaluations, including six multimedia designers (MMDs) with varying degree of design experience and four subject matter experts (SMEs), all of whom had significant knowledge in this area of Mathematics and experience in teaching it to students.

Each expert was asked to use one of three usability evaluation methods. The first method was Multimedia Taxonomy (MMT) (Heller & Martin, 1999), which represents a three-dimensional categorisation framework of multimedia issues, such as media types, their expression and contextual aspects like the target audience and the content. The taxonomy contains 120 cells, in each of which evaluators can ask questions regarding specific issues of media design. The second approach was Multimedia Cognitive Walkthrough (MMCW) (Faraday & Sutcliffe, 1997), which concentrates on cognitive aspects of multimedia presentations. It involves three steps of evaluation of the media design, the media combination and the media selection. Each step contains a set of guidelines against which the relevant presentation segments can be evaluated. Finally, the Interactive Multimedia Checklist (IMMC) (Barker and King, 1993) comprises twelve categories, such as engagement and interactivity, which embody essential principles of good design. The authors suggest 90 questions distributed amongst all categories, and experts are expected to answer the ones relevant to the application being evaluated. The MMT and the IMMC were used by two multimedia designers and two subject matter experts, whereas the MMCW was used by two multimedia designers, as recommended by the authors of the techniques. No subject matter experts used the MMCW because it concentrates on low-level multimedia design issues, and it would not be appropriate for such experts to use.

Results

Learner Tests Results

The video footage containing the student interactions, their verbal protocols and the post-exposure interviews was analysed to identify usability problems. Problems were identified using a set of nine criteria, such as 'the learner articulated a goal but cannot succeed in achieving it without external help from the experimenter' and 'the student expresses confusion while trying to achieve a task'. A total of 51 unique usability problems were found to match the criteria. The comprehension test results showed that students understood the concepts of Reflection and most of those of Translation. However, they had particular problems understanding the principles of Scaling, as well as some principles of Translation. In particular we found that the students had difficulties comprehending 13 of the 20 knowledge propositions. We defined comprehension difficulties as cases where at least two students did not grasp the essence of the knowledge proposition. Thus, as a result of all learner tests we found that the students encountered 64 problems in total, i.e. 51 usability and 13 comprehension problems.

Expert Evaluations Results

A total of 191 unique problems were identified by the experts. The total number of problems identified by each expert group is shown in Table 1. 27 problems were identified by both types of experts using the IMMC, and this number has been included in both totals given in columns 4 and 5.

Evaluation Method / Expert Type	Multimedia Taxonomy	MM Cognitive Walkthrough	Interactive MM Checklist	Total Number	Mean
Multimedia designers	43	34	69	**146**	**24.3**
Subject matter experts	32	-	40	**72**	**18**

Table 1: Number of problems predicted by each expert group

Analysis of the Evaluation Results

The main question to be answered was whether the experts were able to predict the problems experienced by the learners. Therefore, we compared the experts' predictions with the results from the learner tests. To be able to match the two problem sets, six matching rules were established. For instance, problems were matched if both problem statements described the same learner behaviour or if both described the same fault with the same design feature, although it may have been observed in a different page of the application. The results of the problem matching are depicted in Figure 1. As can be seen from the figure, only 28 of the 64 learner problems were predicted. It was found that in total 60 statements identified by the expert evaluators mapped onto 28 of the learner problems. In the following sections we discuss these results.

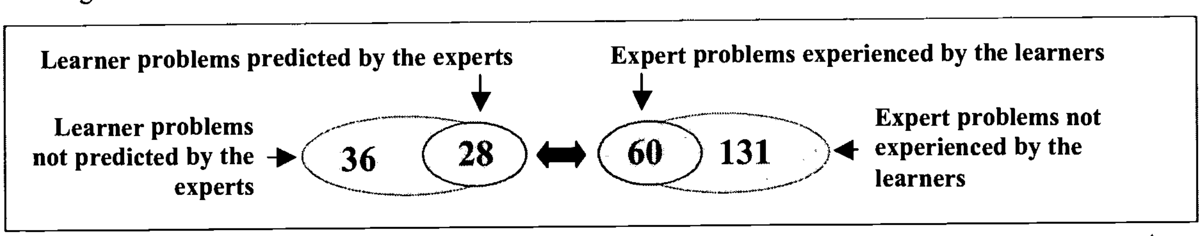


Figure 1: Similarities between learner and expert problem sets

Number of Correctly Predicted, Unidentified and Unobserved Problems

We first analyse the number of correctly predicted versus the number of unidentified learner problems. We then discuss the number of problems the experts predicted, which the students did not encounter in their interaction with the application.

Correctly Predicted Problems
From Figure 1 it can be seen that the experts predicted 28 of the 64 learner problems, or 44%. In particular, we found that 24 out of the 51 usability problems were identified by the experts, or 47%. However, the experts could predict problems with only 4 out of the 13 knowledge propositions which caused comprehension difficulties to the students, which is less than a third. The multimedia designers predicted more of the usability problems, whereas the subject matter experts identified more of the comprehension problems.

Unidentified Problems
The expert evaluations failed to predict certain problems that the students did encounter. We found that in total 36 of the 64 learner problems were not predicted by the experts, or 56%. In particular two thirds of the comprehension

difficulties and nearly half of the usability problems the students encountered were not predicted by the experts. The above results show that the experts had difficulty identifying potential comprehension problems, but they were more successful at predicting usability problems which the learners experienced.

Unobserved Problems
Apart from the 60 problems which were matched with the learner ones, the experts also found 131 other problems. We divided these into two categories – specialist problems and false alarms.

The *specialist problems* category includes 81 problems, which students cannot be expected to identify. These problems concern a variety of issues, such as the accuracy of the Maths equations and the notation used. We found that a significant proportion of the problems identified by the SMEs fell into this category (in total 60% of their predictions), whereas only 25% of all issues predicted by the MMDs were specialist ones.

False alarms are issues which experts identified as problematic but which did not cause problems to the learners either while interacting with the software or during the knowledge tests. We found 50 false alarms in total, which amounts to 26% of all expert predictions. Most of them were raised by the multimedia designers. One reason for this could be that the MMDs were more critical about the design of the application, pointing out minor issues which did not cause problems to the learners.

The analysis so far only provides information on the proportion of the learner problems predicted or not by the expert evaluators. The next part of the analysis aims to provide a more detailed review of the types of problems which the learners and the experts focused on during the evaluation of the IMM application.

Types of Problems Identified

From the analysis we found that although there are some similarities between the problems identified in the learner tests and the expert evaluations, each group paid attention to different aspects of the IMM application.

Types of learner problems the experts could predict
One area where the experts predicted all learner problems is *affordance*, which encompasses difficulties relating to students not being able to identify which part of the presentation affords certain actions or what action a particular button affords. An example of such problem is shown in Figure 2 (a), which illustrates that after reading the instruction circled the students had difficulty identifying where to click for the graph of 10^x. Both expert groups also detected some issues of *learner engagement*, i.e. how interesting and challenging (or not) the application was to the students.

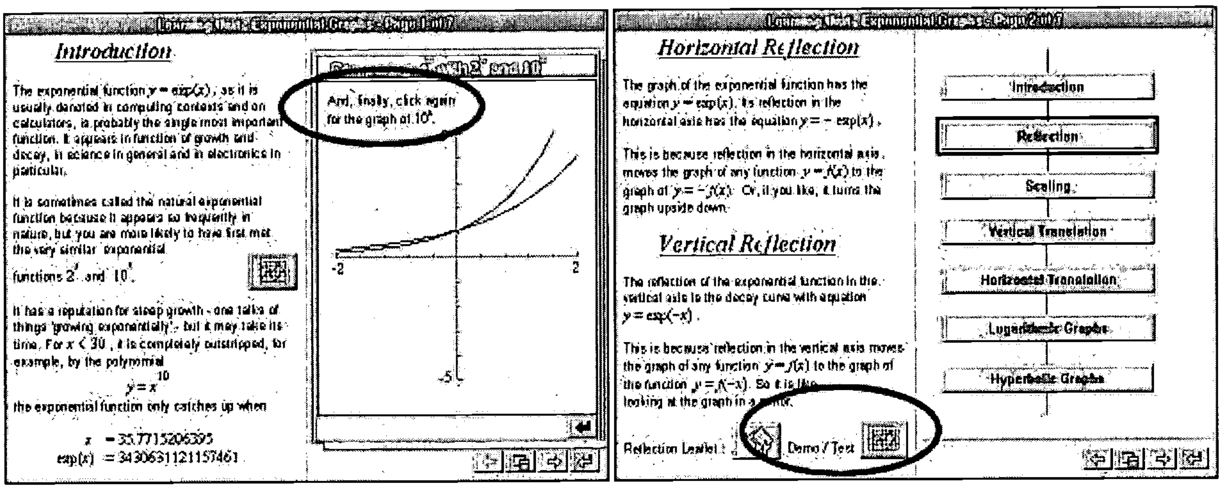
The multimedia designers also focused on problems with the *design and appearance of the media resources* used, such as the design of the graphics, graph lines, quality of the icons and the pop-up message boxes. This kind were also identified by the students. The experts further spotted some problems with *synchronising* time-varying media resources, such as animated text which changes too quickly for the students to read. The MMDs also identified some problems with the *navigation* within the application. Finally, mostly the SMEs, but also two of the MMDs, pointed out some areas in the presentation which they believed were not sufficiently clear, and the students actually had difficulties understanding these sections.

Types of learner problems the experts could not predict
On the other hand, a number of learner problems eluded the attention of the experts. These fall into three categories: *learning support, comprehension* and *missed interaction*.

Learning support problems deal with how much explanation of the material the students required. This greatly depends on the students' prior knowledge. Most students requested more help with Scaling and Translation, especially Scaling, since they had no previous knowledge of these concepts. Although before the evaluation sessions the experts were told to assume none or little prior knowledge of the subject matter, none of them could envisage where students may need further explanation of the material. Furthermore, none of the evaluation methods explicitly asked the evaluators to consider students' prior knowledge in order to identify such issues.

The *comprehension* problem category describes which parts of the material the students had problems understanding. Although the experts identified some areas of the material which could potentially cause such difficulties to students, they missed out a significant number of them. One factor found to influence the comprehension was the varying complexity of the Maths material. The higher the complexity of the material the greater the cognitive task requirements were on the students. Reflection was found to be the simplest concept, the principles of Translation were slightly more complex, and those of Scaling were the most complex of the three. The comprehension test results showed that all students grasped the concepts of Reflection, the majority of them got the Translation right as well, however most of them experienced difficulties with understanding Scaling. None of the evaluation methods suggests that the complexity of the material or the cognitive task requirements should be considered, and none of them correlates these aspects to how media resources could be used and designed to represent complex concepts in order to enable students to comprehend them easier.

Finally, *missed interactions* are situations where the students did not perform an interaction which is considered important for achieving their learning tasks. One such situation arose on the Horizontal Reflection screen, illustrated in Figure 2 (b), where a student skipped the test regarding Reflection, which would have helped them reflect on what they had learned about it. Such situations occurred predominantly because the learner's attention was not explicitly drawn to the important parts of the presentation. As can be seen from Figure 2 (b) the icon to start the test is placed at the bottom right-hand corner of the main presentation screen where the learner is not likely to look very often.



(a) An example of an *Affordance* problem (b) An example of a *Missed Interaction*

Figure 2: Sample screens from MathWise (NAG ©) illustrating learner problems

Types of problems the experts predicted but the students did not encounter

As mentioned earlier, the experts predicted a number of problems which the students did not experience, which we categorised as *specialist problems* and *false alarms*.

Specialist problems include pedagogical and instructional design issues, which fall into four categories. Firstly, many predicted problems concerned the accuracy and completeness of the Maths content and the notation used. Such problems were identified by the subject matter experts. For instance, two SMEs identified a mistake in one of the equations of Vertical Scaling. Secondly, issues regarding the adequacy of different monitoring and assessment techniques were identified. A third set of issues questioned whether different expert system facilities are required to support learners. Finally, the experts also made suggestions as to how the design of the application could be improved. Some of these specialist issues can potentially point to usability and learning problems. However, they were specified in a way that only revealed design faults, without identifying the likely effect of the faults on the learner's behaviour or performance.

In the *false alarms* category we include issues which experts identified as problematic but did not cause problems to the learners. Most false alarms were due to experts making wrong assumptions about students' sense of orientation within the application and the information presented, their control over the application and preferences regarding customisation of program settings. For example, one expert thought that students could lose a concept of where they are in the application, however when asked during the interviews none of the students reported experiencing such confusion. Such comments were made predominantly by the MMDs. The multimedia designers also commented on design faults which did not seem to bother the students. Perhaps because the students were so engaged in grasping the Maths material, they did not seem to notice presentation imperfections, such as some of the letters in the titles not being properly drawn. Such issues, however, are valid design considerations and can be useful for redesigning the application. Finally, the SMEs presupposed that learners' attention and concentration could not be maintained consistently, which was not the case with the students. However, the experimental nature of the evaluation could have caused the students to be more focused.

Conclusions

The results of the study presented in this paper show that the experts were successful at predicting a number of usability problems the students encountered. However, despite using formal usability evaluation methods, the evaluators did have difficulty predicting certain types of learner problems, particularly comprehension, learning support and attention to important information. One explanation of this is that the experts and the learners showed differences in focus. The subject matter experts emphasised matters of the content, the multimedia designers paid particular attention to the media and presentation design, media synchronisation and navigation, while the students were more concerned with how understandable the material was. Another critical issue that emerged from the study is that expert evaluators tend to uncover design and content faults, but rarely try to infer what consequences such faults may have on learners' behaviour and performance. Even when they did try to predict the effects on the learner, they often made wrong assumptions. The evaluation methods also did not support experts in making such predictions. The evidence presented above suggests that expert evaluations, although effective, do not eliminate the need for actual tests with learners.

The prediction rates of expert evaluations could be improved by training the experts in how to use learner data more effectively, so that they can make better assumptions about students' interaction with the IMM, and their behaviour and performance. Furthermore, more research is required into how the design of IMM should take into account relevant learner characteristics, such as their prior knowledge, metacognitive skills and personal motivations, and incorporate the findings into evaluation methods for use by experts. The existing usability evaluation methods also need to be enhanced to consider how the major factors contributing to effective IMM design - the learner, the subject matter content, the instructional approach adopted and the context of use – all relate to each other. This will provide a more integrated approach for evaluating the effectiveness of IMM.

References

Barker, P. and King, T. (1993) Evaluating Interactive Multimedia Courseware – A Methodology. *Computers in Education*, 21(4), 307-319.

Dimitrova M. and Sutcliffe A., (1999) Designing Instructional Multimedia Applications: Key Practices and Design Patters. *ED-MEDIA 1999*, AACE Publications, Seattle, WA, 358-363.

Draper, S.W., Brown, M.I., Henderson, F.P. and McAteer, E. (1996) Integrative Evaluation: An Emerging Role for Classroom Studies of CAL. *Computers and Education*, 26 (1-3), 17-32.

Faraday, P. and Sutcliffe, A. (1997) Evaluating Multimedia Presentations. *The New Review of Hypermedia and Multimedia,* 3, 7-37.

Heller, R.S. and Martin, C.D. (1999) Multimedia Taxonomy for Design and Evaluation. *Handbook of Multimedia Computing*, B. Furht (ed.), CRC Press, 3- 16.

Reiser, R.A. and Kegelmann, H.W. (1994) Evaluating Instructional Software: A Review and Critique of Current Methods. *Educational Technology Research and Development*, 42(3), 63-69.

Squires, D. and McDougall, A. (1996) Software evaluation: a situated approach. *Journal of Computer Assisted Learning*, 12, 146-161.

Tergan, S. (1998) Checklists for the Evaluation of Educational Software: Critical Review and Prospects. *Innovations in Education and Training International*, 35(1), February 1998, 9-20.

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").