

HEURISTIC EVALUATION OF EDUCATIONAL MULTIMEDIA: FROM THEORY TO PRACTICE

Albion, P.R.

Department of Education
University of Southern Queensland

email: albion@usq.edu.au
<http://www.usq.edu.au/users/albion>

Abstract

Cost-effective methods for formative evaluation of educational multimedia are needed. Heuristic methods have been shown to be cost-effective in the area of user interface evaluation. This paper describes the use of heuristic methods to evaluate the user interface, educational design and content of an educational multimedia project.

Key words

Heuristic evaluation, formative evaluation, multimedia, educational design

1. Introduction

The development of educational multimedia inevitably requires the commitment of substantial amounts of time and money. Both are typically in short supply in educational institutions. In the interests of ensuring that resources are used to best effect it is important to ensure that both the processes and products of multimedia development are evaluated.

Evaluation serves multiple purposes in relation to the various stages and functions of the multimedia development process (Reeves and Hedberg forthcoming) and different methodologies will be employed at different stages. They may include reviewing published research and existing products, observing the responses of users as they interact with the product or assessing changes in performance resulting from implementation.

Evaluation is sometimes seen as less important than the processes of development. As a consequence there is pressure to minimise evaluation costs (Quinn 1996) by selecting and applying methods that make efficient use of resources.

Formative evaluation is especially prone to resource pressures. The results are needed quickly so that development is not delayed and it may be necessary to undertake multiple evaluations for successive cycles of development.

2. Background to heuristic evaluation

The need for effective and efficient methods of formative evaluation is not unique to multimedia development. Usability inspection has been developed as a formative evaluation technique in the area of Human Computer Interaction where the focus is on improving the ease of use of software.

2.1 Usability inspection methods

Usability inspection is the generic name for evaluation methods that rely upon considered judgement of inspectors. They have been contrasted with other approaches to evaluation.

Automatic evaluation by computer program does not work well as yet and formal mathematical analyses of usability do not scale well to complex interfaces. Empirical testing is expensive because of the need to involve multiple users for significant periods of time (Mack and Nielsen 1994). Usability inspection methods have been shown to be a cost effective alternative to these evaluation methods (Desurvire 1994). Various inspection techniques have been described, such as heuristic evaluation, pluralistic walk-through, cognitive walk-through (Nielsen and Mack 1994), and graphical jog-through (Demetriadis, Karoulis et al. 1999).

Heuristic evaluation is among the easiest methods to learn and results in problem reports that appear to be better predictors of end-user problems (Mack and Nielsen 1994). The method uses multiple evaluators who conduct independent inspections in which they compare interface elements with a list of recognised usability principles, the heuristics. An heuristic is a general guide for some activity, what might be described as a 'rule of thumb'. The heuristics compiled by Nielsen (1994, see Table 1 below) included such widely accepted principles of user interface design as "supports recognition rather than recall" and "prevents errors". The reports of the multiple evaluators are considered together in order to maximise the chances of properly identifying any usability problems. Studies have found that the use of 3 to 5 evaluators is the reasonable minimum that will ensure identification of about 75% of usability problems in a project. The use of more evaluators will result in only marginal improvements in the rate of detection (Nielsen 1994).

2.2 Heuristic evaluation of educational software

It is axiomatic that software of any type should meet basic standards for usability. In pursuit of this goal, usability inspection methods for user interface evaluation can be applied to educational software. However, Quinn (1996) proposed that usability inspection approaches might be adapted for the purpose of evaluating the educational design of software.

In Quinn's model the evaluators would include representatives from the target learner group, educational design experts and content experts for the relevant domain. The heuristics would comprise a compilation of elements of good educational design based upon tenets of relevant educational theories.

Quinn developed a draft list of eight heuristics based upon theories including cognitive apprenticeship, anchored instruction, problem-based learning and technology-mediated instruction. These were selected because, despite their differences in emphasis and sequencing, they are broadly constructivist and share characteristics such as engaging the learner in sequenced activities and guided reflection on learning.

Such an evaluation of the educational design of software would not replace usability inspection. However, since there is likely to be some overlap in the problems identified, Quinn suggested that the numbers of evaluators for each process could be kept low for a total of 6 to 8 evaluators.

Quinn's original paper did not report on the results of any trials of the method. Nor do there appear to be any published reports of subsequent trials.

Others have recognised the potential of heuristic evaluation methods in relation to educational software. Squires (1997) distinguished between predictive evaluation of software as undertaken by teachers prior to purchase and interpretive evaluation of the software in use with students. Arguing that established predictive evaluation techniques, such as writing software reviews or using checklists and frameworks, ignore context and are time consuming, Squires advocated an heuristic approach to the predictive evaluation of educational multimedia.

More recently, Squires and Preece (1999) proposed an approach to predictive evaluation of educational software based on a set of heuristics that integrate usability and learning issues. They identified cognitive and contextual authenticity as important dimensions in the evaluation of software for use in socio-constructivist learning environments. Under each of these dimensions they located key aspects related to credibility, complexity and ownership in the case of cognitive authenticity and collaboration and curriculum in the case of contextual authenticity. These aspects were considered in the light of the 10 usability heuristics identified by Nielsen (1994) and possible inter-related issues were identified for 19 of the possible 50 areas of interaction. A set of 8 'learning with software' heuristics were derived but empirical testing of the heuristics is yet to be conducted.

Heuristic evaluation methods appear to offer potential benefits in the evaluation of educational multimedia and it was this potential that led to the adoption of the heuristic approach which is described in this paper.

3. Application of heuristic evaluation

The evaluation described here was conducted on the beta version of an educational multimedia product at the University of Southern Queensland. The instructional design and development of the project have been described elsewhere (Albion 1999; Albion and Gibson 1998a; Albion and Gibson 1998b; Albion and Gibson 1998c; Gibson and Albion 1997).

3.1 Early formative evaluation

Several evaluation strategies were employed during the design and initial development of the project. Members of the team engaged in iterative walk-throughs of the scenarios as they were defined and refined. Simple representations of screen designs were constructed to facilitate visualisation and content was carefully reviewed.

One of the four scenarios envisaged for the final product was laid out in detail and the content prepared before serious work was begun on the remaining three. This permitted the creation of a working prototype which, while not identical to the final user interface, was sufficiently complete to permit testing of the design concept.

This preliminary evaluation was conducted with a group of 30 students representative of the intended target group. Students were observed while working with the materials and were prompted or assisted as necessary to compensate for the incomplete interface. Data

obtained from observations, interviews and a brief questionnaire indicated that students found the materials both motivational and informative. Their comments on specific components were noted and used to inform subsequent development.

3.2 Selection and use of heuristic evaluation

The design of the package evolved as a consequence of the evaluation that occurred throughout development. Thus there was no opportunity for evaluation of the complete package until a beta version was available.

The purpose of evaluation at this stage was to identify problems that should be remediated before release of the package for implementation. Heuristic evaluation was selected as a suitable approach for this purpose.

The instructional design implemented in the package was relatively novel (Albion and Gibson 1998a) and there was a desire to validate the educational value of the design. Moreover, its constructivist orientation matched the theories on which Quinn (1996) had based his proposal for heuristic evaluation of educational design. Hence Quinn's method was selected for use.

Although Quinn referred to the inclusion of 'content experts' among the potential evaluators, his heuristics did not specifically address content issues. The nature of this package and its use of content to create context in the scenarios made it important to evaluate the authenticity of the included content in addition to the interface and educational design. Hence a third set of heuristics directed towards content was developed.

The heuristics proposed by Nielsen (1994) and Quinn (1996) were adapted with minor changes to the wording of some descriptors to facilitate understanding by evaluators some of whom were from non-technical backgrounds. The three sets of heuristics are shown in Tables 1, 2 and 3.

Table 1: Interface design heuristics [after Nielsen (1994)]

| | |
|--|---|
| <i>Ensures visibility of system status</i> | The software keeps the user informed about what is going on through appropriate and timely feedback. |
| <i>Maximises match between the system and the real world</i> | The software speaks the users' language rather than jargon. Information appears in a natural and logical order. |
| <i>Maximises user control and freedom</i> | Users are able to exit locations and undo mistakes. |
| <i>Maximises consistency and matches standards</i> | Users do not have to wonder whether different words, situations or actions mean the same thing. Common operating system standards are followed. |
| <i>Prevents errors</i> | The design provides guidance which reduces the risk of user errors. |
| <i>Supports recognition rather than recall</i> | Objects, actions and options are visible. The user does not have to rely on memory. Information is visible or easily accessed whenever appropriate. |
| <i>Supports flexibility and efficiency of use</i> | The software allows experienced users to use shortcuts and adjust settings to suit. |
| <i>Uses aesthetic and minimalist design</i> | The software provides an appealing overall design and does not display irrelevant or infrequently used information. |
| <i>Helps users recognise, diagnose and recover from errors</i> | Error messages are expressed in plain language, clearly indicate the problem and recommend a solution. |
| <i>Provides help and documentation</i> | The software provides appropriate online help and documentation which is easily accessed and related to the users' needs. |

Table 2: Educational design heuristics [after Quinn (1996)]

| | |
|--|---|
| <i>Clear goals and objectives</i> | The software makes it clear to the learner what is to be accomplished and what will be gained from its use. |
| <i>Context meaningful to domain and learner</i> | The activities in the software are situated in practice and will interest and engage a learner. |
| <i>Content clearly and multiply represented and multiply navigable</i> | The message in the software is unambiguous. The software supports learner preferences for different access pathways. The learner is able to find relevant information while engaged in an activity. |
| <i>Activities scaffolded</i> | The software provides support for learner activities to allow working within existing competence while encountering meaningful chunks of knowledge. |
| <i>Elicit learner understandings</i> | The software requires learners to articulate their conceptual understandings as the basis for feedback. |
| <i>Formative evaluation</i> | The software provides learners with constructive feedback on their endeavours. |
| <i>Performance should be 'criteria-referenced'</i> | The software will produce clear and measurable outcomes that would support competency-based evaluation. |
| <i>Support for transference and acquiring 'self-learning' skills</i> | The software supports transference of skills beyond the learning environment and will facilitate the learner becoming able to self-improve. |
| <i>Support for collaborative learning</i> | The software provides opportunities and support for learning through interaction with others through discussion or other collaborative activities. |

Table 3: Content heuristics

| | |
|---|---|
| <i>Establishment of context</i> | The photographs, documents and other materials related to the simulated schools create a sense of immersion in a simulated reality. |
| <i>Relevance to professional practice</i> | The problem scenarios and included tasks are realistic and relevant to the professional practice of teachers. |
| <i>Representation of professional responses to issues</i> | The sample solutions represent a realistic range of teacher responses to the issues and challenge users to consider alternative approaches. |
| <i>Relevance of reference materials</i> | The reference materials included in the package are relevant to the problem scenarios and are at a level appropriate to the users. |
| <i>Presentation of video resources</i> | The video clips of teacher interviews and class activities are relevant and readily accessible to the user. |
| <i>Assistance is supportive rather than prescriptive</i> | The contextual help supports the user in locating relevant resources and dealing with the scenarios without restricting the scope of individual responses. |
| <i>Materials are engaging</i> | The presentation style and content of the software encourages a user to continue working through the scenarios. |
| <i>Presentation of resources</i> | The software presents useful resources for teacher professional development in an interesting and accessible manner. |
| <i>Overall effectiveness of materials</i> | The materials are likely to be effective in increasing teachers' confidence and capacity for integrating information technology into teaching and learning. |

The heuristics were presented to the evaluators on a form where each heuristic was accompanied by a rating scale (1= poor to 5= excellent with an additional rating of NA for "Not Applicable") and space for comments. The heuristic evaluation method as described by Nielsen (1994) does not use such a rating scale although evaluators may be asked to rate the severity of problems they identify. In the present evaluation it was considered that the addition of a rating scale might lend itself to obtaining an overall assessment of the perceived quality of the materials. Evaluators were asked to rate the package on each characteristic and to add any relevant comments in the spaces provided. To ensure ample space for comments, the forms were printed on one side only of the paper and evaluators were encouraged to add additional pages as necessary.

3.3 Evaluation procedures

The group of evaluators was chosen to include persons with expertise in user interface design, instructional or educational design and teaching. In addition, two undergraduate students were selected to provide reactions representative of the intended user group.

This project was designed for CD-ROM delivery for access using a web browser. A beta version of the CD-ROM was supplied to evaluators for use on their own equipment. This approach provided for evaluation under conditions approximating those of intended use

on a variety of computer systems with different browsers. It was also convenient for the evaluators who would otherwise have been required to commit a substantial period of time to work through the material in a test facility. However, this flexibility introduced some problems with providing support to the less technically adept evaluators with installing ancillary software or dealing with minor problems which arose. It also increased the likelihood of delays in obtaining responses.

4. Results

The time taken by individual evaluators to return their forms varied from a few days to several weeks and in some cases reminder notices were necessary to obtain the data. The reasons for the delays varied but were mostly related to the other commitments of the evaluators who were all volunteers undertaking the evaluation on their own time.

Descriptive comments obtained from the evaluators revealed that their experiences varied substantially. Some reported no difficulty in accessing the materials while others, despite following the instructions provided for installation and use, apparently experienced problems with the operation of their browser and associated components. A small number of evaluators commented on the lack of specific instructions for certain tasks and of options for accessing resources they had seen previously. When questioned about their use of the 'Help' button or the device for accessing resources they revealed they had not actually used those facilities.

Numerical ratings were summarised by averaging. For the seven completed responses available at the time of writing averages on the 28 heuristics ranged from 3.6 to 5.0 on the 5 point scale with just two averages below 4.0. Inspection of the results revealed that some evaluators had rated certain items as low as 1 or 2 confirming that they had not simply scored at the same point on the scale for each item.

Both positive and negative comments were received for each of the three sets of heuristics - interface, educational design and content - and in some instances the evaluators offered constructive suggestions for improvement. Examination of the responses revealed that many of the problems had been noted by only one or two of the evaluators. The identified problems were used to develop a list for further investigation and remediation in the final version of the materials.

5. Discussion

With the benefit of hindsight, the delays in obtaining responses were probably predictable. The evaluation took place at a busy time of the academic year when several of the evaluators had commitments to assessment work, residential schools and conferences. Moreover, the volume of material on the CD-ROM was substantial, including over an hour of QuickTime™ video clips and some hundreds of pages of text. Although the interface and other aspects of the design could probably be adequately 'inspected' for the purposes of this evaluation in an hour or two, there was a tendency among evaluators to want to be as thorough as possible. Thus they were reluctant to begin working on the material until they had a substantial amount of time available. Clearer

briefing notes about the nature of the inspection process and the amount of time that might be expected to be committed would probably have helped to speed the evaluation process.

The apparent inconsistencies among the experiences reported by evaluators reflected differences in their expectations of the package, their computer systems and their technical expertise with those systems. In addition to the variability in response that underlies Nielsen's finding that several evaluators are needed in order to be confident of identifying most of the serious problems (Nielsen 1994) there appeared to be significant differences in the experiences of evaluators even when they reported using similar computing platforms. In some cases the issues they identified might have been resolved by use of the online help or other facilities in the package but they had either not located those or had chosen not to use them. The latter group of issues were addressed by insertion of additional explanatory material in the introductory sections of the package and by changes to the interface to make key support features more easily accessible.

Nielsen's (1994) description of the process of heuristic evaluation included the possibility of a debriefing session. Circumstances in this evaluation made a session with the group of evaluators impossible but individual evaluators were contacted to discuss issues they had raised. These discussions helped to locate the causes of problems and sometimes elicited suggestions for improvement.

Despite the delays occasioned by the manner in which the evaluation was conducted the method proved effective in identifying issues requiring attention in all three of the design aspects addressed by the heuristics. Because the evaluators worked independently of each other it was possible to begin work on modification of some elements before all responses had been returned. Although the heuristics for educational design and content can undoubtedly be improved on the basis of experience in this and other evaluations they do appear to represent a useful foundation for cost effective evaluation of educational multimedia.

6. References

- Albion, P.R. (1999) 'PBL + IMM = PBL²: Problem-based learning and multimedia development', in Price, J.D., Willis, J., Willis, D.A., Jost, M. and Boger-Mehall, S. (eds) *Technology and Teacher Education Annual 1999*, Charlottesville, VA: Association for the Advancement of Computing in Education, pp. 1022–1028.
- Albion, P.R. and Gibson, I.W. (1998a) *Designing Multimedia Materials Using a Problem-Based Learning Design*. 15th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, Woollongong, Australasian Society for Computers in Learning in Tertiary Education.
- Albion, P.R. and Gibson, I.W. (1998b) 'Designing Problem-Based Learning Multimedia for Teacher Education', in Price, J.D., Willis, J., Willis, D.A., Jost, M. and Boger-Mehall, S. (eds) *Technology and Teacher Education Annual 1999*, Charlottesville, VA: Association for the Advancement of Computing in Education, 1240–1244.
- Albion, P.R. and Gibson, I.W. (1998c) 'Interactive multimedia and problem based learning: Challenges for instructional design', Ottman, T. and Tomek, I. (eds) *Educational Multimedia and Hypermedia 1998*, Charlottesville, VA: Association for the Advancement of Computing in Education, pp. 117–123.
- Demetriadis, S., Karoulis, A., et al. (1999) "'Graphical" Jogthrough: expert based methodology for user interface evaluation, applied in the case of an educational simulation interface'. *Computers & Education*, 32, 285–299.
- Desurvire, H.W. (1994). 'Faster, cheaper!! Are Usability Inspection Methods as Effective as Empirical Testing?' Nielsen, J. and Mack, R.L. (eds) *Usability Inspection Methods*, New York: John Wiley & Sons, pp. 173–202.
- Gibson, I.W. and Albion, P.R. (1997) 'CD-ROM based hypermedia and problem based learning for the pre-service and professional development of teachers', in Conway, J., Fisher, R., Sheridan-Burns, L. and Ryan, G. (eds) *Research and Development in Problem Based Learning*, Newcastle: Australian Problem Based Learning Network, 4, 157–165.
- Mack, R. and Nielsen, J. (1994) 'Executive summary', Nielsen, J. and Mack, R.L. (eds) *Usability Inspection Methods*, New York: John

Wiley & Sons, pp. 1–23.

Nielsen, J. (1994) 'Heuristic Evaluation', in Nielsen, J. and Mack, R.L. (eds) *Usability Inspection Methods*, New York: John Wiley & Sons, pp. 25–62.

Nielsen, J. and Mack, R. (1994) *Usability Inspection Methods*, New York: John Wiley and Sons.

Quinn, C.N. (1996) 'Pragmatic evaluation: lessons from usability'. *13th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, Australasian Society for Computers in Learning in Tertiary Education.

Reeves, T. and Hedberg, J. (forthcoming). *Evaluating Interactive Learning*.

Squires, D. (1997) 'An heuristic approach to the evaluation of educational multimedia software'. *Computer Assisted Learning Conference*, University of Exeter.

Squires, D. and Preece, J. (1999) 'Predicting quality in educational software: Evaluating for learning, usability and the synergy between them'. *Interacting with Computers*, 11, 467–483.

© Albion, P.R.

The author(s) assign to ASCILITE and educational non-profit institutions a non-exclusive license to use this document for personal use and in course of instruction provided that the article is used in full and this copyright statement is reproduced.

The author(s) also grant a non-exclusive license to ASCILITE to publish this document in full on the World Wide Web (prime sites and mirrors) and in printed form within the ASCILITE99 Conference Proceedings. Any other usage is prohibited without the express permission of the author(s).