Evaluation of a Learning System – and Learning to Evaluate

Helen Keegan
University of Salford
H.Keegan@salford.ac.uk

Abstract

This paper describes the process of the formative evaluation of an acoustics computer-aided learning (CAL) package. The CAL package, entitled “Waves: Analysis and Synthesis”, is a multimedia-rich learning support package for students on sound-related undergraduate courses. The evaluation took place over a two-month period in the School of Acoustics and Electronic Engineering at the University of Salford. In addition to being an evaluation of the CAL itself, this was also an exercise in learning to evaluate. The methodological mix is discussed along with the evaluation process as a whole.

1. The evaluation

This was effectively an exercise in learning to evaluate - accordingly there was no fixed methodology or set of tools from the onset. Advice was sought from ‘expert’ evaluators, along with a thorough survey of the evaluation literature from books and journals. Consequently this was an opportunity for the evaluator to employ different instruments and approaches in order to see what worked best (in this case). The sequence of the evaluation and the instruments used were mostly dictated by time and resource constraints.

The approach taken in the evaluation of the CAL was largely illuminative [1], using a combination of different evaluation techniques and exploratory data analysis to compare, triangulate and interpret findings. Quantitative methods were employed to obtain easily comparable data of fixed measures; for example, a post-task questionnaire (PTQ) was designed to elicit data on aspects of design, navigation and the user interface. Qualitative methods were employed in order to detect and explore unanticipated issues. In what was essentially an ethnographic approach, the learners were observed using the CAL. This proved to be an extremely rich source of data in terms of usability. Interviews and focus groups enabled us to follow-up the findings in more depth. Pre and post knowledge testing was used in an attempt to measure any learning gains as a result of the intervention.

In reality practicalities and time constraints largely dictated the combination of tools in each case. Five groups of learners participated in the evaluation of the CAL, which took place over a two-month period.

GROUP 1 - USABILITY: The approach to evaluation with group one was informal. The three volunteers were final year students who were asked to take the software home and use it in their own time, noting how long they spent using the package and making notes of any problems encountered as they worked through the CAL. All three students made comprehensive notes, which mostly related to usability. They also completed the PTQ before being interviewed on a one-to-one basis, which enabled us to discuss the issues that they had raised in more depth. The findings contributed to usability, design, and the user interface. Problems had been identified which were ‘ironed-out’ before the next stage of the evaluation in order to avoid re-evaluating the same usability issues.

GROUP 2 - USABILITY/PEDAGOGY: The evaluation approach was far more structured and intensive with group 2, which comprised of four final year students who had been away for a year on industrial placements. This stage of the evaluation was split into two sessions. During the first session the subjects completed Student Profile and Task Experience questionnaires before using the CAL. The same test was administered post-task in order to elicit two comparable sets of data to assess the pedagogical effects of the CAL. The post-task questionnaire, which focused on Content, Usability, Design and Navigation, consisted of Likert-scales and open-ended questions. Supplemental observation was employed in order to study the human-computer interaction. Any difficulties or unexpected problems were then recorded in the session logs, which were used to identify key issues surrounding the usability of the CAL. The logs were also used as a prompt in the follow-up interviews. The group interview allowed the subjects to discuss their experience of using the CAL, which proved to be a rich source of data.

GROUP 3 - USABILITY/PEDAGOGY: The approach taken and the instruments employed mirrored group 2,
except that there was no group interview due to practical constraints. In this case the group comprised of four first year students, which enabled us to compare the pre/post knowledge tests of groups 2 and 3. By choosing 2 groups with a divergent range of abilities and domain knowledge we could observe the pedagogical effects of the package upon different groups of learners.

GROUP 4 – USABILITY: Group four comprised of eleven second-year students taking a Digital Audio module who used the CAL in place of a Sound Synthesis lecture. With this group there was no questionnaire or knowledge testing. We were concentrating on their open responses to the CAL, in light of the modifications that had been made to the software. The group were asked to write a ‘letter to next years students’ explaining the nature of the CDROM [2]. This proved to be an excellent way of evaluating their overall impressions of the CAL. However, this didn’t tell us anything that we didn’t know already, although it would have been valuable at earlier stages of the evaluation. The session ended with a focus group interview, where the group were asked specific questions related to issues that had been addressed through further development of the CAL in light of the findings from groups 1, 2 and 3.

GROUP 5 – USABILITY: Group five comprised of fourteen second-year students who were taking a Midi and Music Technology module as part of a Professional Sound and Video Technology HND. The CAL replaced a two-hour lecture on Sound Synthesis. This was the only stage in the evaluation where the evaluator was not known or present. This was intentional as the overall responses from the other groups had been consistently positive. It was decided that by implementing an ‘anonymous’ evaluation in which the lecturer (rather than the evaluator) explained and directed the study, we may be able to detect whether the findings were influenced by a Hawthorne effect. The learners were given post-task questionnaires. However, there were no significant differences or findings in the feedback (Open-ended questions) or the overall results (Likert scales). This does not necessarily mean that the prior evaluations were unbiased – perhaps the students also wanted to ‘please’ the lecturer.

2. Conclusion

This paper has described the process of a formative evaluation of an acoustics CAL package. The iterative nature of the evaluation resulted in a cycle of testing and amendments to both the CAL package and the evaluation tools. The evaluation progressed from being initially illuminative and open-ended to being more integrative in nature [3]. The evaluation took shape over the 5 sessions, as usability problems were detected and modifications were made before evaluating the CAL with the following group. The responses were easily anticipated by stage four of the evaluation, which indicated that the study had been fairly comprehensive. This process is shown schematically in Figure 1.

![Figure 1. Evaluation process.](image)

This framework facilitated a holistic perspective on the data, identifying which aspects of the CAL needed to be redesigned, and enabling the evaluator to ascertain the relative strengths and weaknesses of different evaluation techniques. The open-ended measures were by far the most valuable instruments employed in the evaluation. Learning gains proved to be extremely difficult to measure.

The initial stages in this phase of the evaluation were small-scale, highly domain dependent, and relatively low in authenticity [4]. Future work will involve the integrative evaluation of the CAL intervention as part of a longitudinal study.

3. References


Further details

A comprehensive discussion of the CAL itself and the evaluation as a whole is beyond the scope of this paper. For the full version, please contact the author.