CRITICAL REFLECTION ON A PROGRAM OF DEVELOPMENT EVALUATION

Kennedy, G.E.

Biomedical Multimedia Unit
The University of Melbourne

Email: g.kennedy@medicine.unimelb.edu.au

Abstract

This paper reflects on a program of development evaluation, a year after its implementation in the Faculty of Medicine, Dentistry and Heath Sciences at the University of Melbourne. This paper reports on what has been learnt about the evaluation program from the involvement of over fifteen computer facilitated learning development teams. It provides an overview of the program's implementation and discusses six themes: evaluation criteria, data collection techniques, evaluation perspectives, timing of the program, coaching of staff and general implementation. Areas where improvements can be made are noted.

Keywords:

development evaluation, multimedia development, computer facilitated learning, instructional design, project management.

Introduction

In 1998 the Biomedical Multimedia Unit (BMU) established a program of development evaluation in the Faculty of Medicine, Dentistry and Heath Sciences. This program aimed to assist staff with the design and development of educational multimedia and to maintain the quality of educational multimedia being developed in the Faculty. This paper reflects on the program of development evaluation, a year after its implementation in the Faculty.

A number of contextual factors needed to be considered before the program of evaluation was established. These contextual factors determined many features of the program. The first issue was the number and range of projects the program needed to cover. Seventeen educational multimedia projects were undertaken in 1998 and this number was increased by an additional 16 projects in 1999. There was a great deal of variety in the projects across the Faculty. This was evidenced by the number of disciplines covered by projects, the variety of pedagogical underpinnings and the different levels of multimedia sophistication.

The individuals who were involved in the program were universally academic staff who specialised in the content area of the multimedia package they were developing (eg. physiology, microbiology, physiotherapy or dental science). The academic staff (henceforth called "developers") were often unfamiliar with many aspects of multimedia production and while many had taught for a number of years, most were not well versed in educational theory. Additionally, while many had participated in scientific research for some time, most were not trained in the process of evaluation. It was imperative that the program of evaluation be accessible to developers. If the program was to be successful developers needed to see its utility and needed to understand the reasons for carrying out an evaluation in a particular way. If the evaluation process was at all cumbersome, developers, who were also busy academic staff, would soon tire of the procedure even if they recognised its value.

Finally, it was acknowledged that only one part-time position was available to develop and manage this evaluation program. There was a need, therefore, to establish a program which, to a certain extent, ran itself. While the program needed to support developers in their evaluation endeavours, much of the onus to complete evaluations needed to be placed on the developers themselves.
As mentioned above these contextual factors needed to be considered when establishing the goals of the evaluation program and planning its structure. Broadly, this program had three goals:

- to refine and improve computer facilitated learning (CFL) packages being developed in the Faculty
- to establish a program of evaluation which would lead to a "culture of multimedia evaluation" in the Faculty
- to coach multimedia developers in the process of evaluation so that they would become more independent and be able to manage their own multimedia evaluation

These goals are implicit in the key structural features of the program which were:

- an emphasis on the iterative nature of the development evaluation program which is tied to the multimedia development cycle
- the need to consider three different evaluation domains and specific criteria within these domains
- the need to consider different evaluation perspectives including both internal and external review
- the need to employ a variety of data collection techniques

The themes discussed in the paper fall into six areas: evaluation criteria, data collection techniques, evaluation perspectives, timing of the program, coaching of staff and general implementation. Over fifteen development teams participated in the program to varying degrees. This paper reports on what the BMU has learnt about the program of development evaluation from its involvement with these development teams. It provides an overview of the program's implementation, notes the successes and failures and provides examples of specific experiences with development teams. The paper notes what has been learnt so far from the implementation of the program and what alterations will be made to the program in the future. It is hoped that others involved in the development evaluation of educational multimedia can learn from this review.

Evaluation criteria

Development evaluation criteria were predominantly based on characteristics of CFL packages. The development and nature of these criteria are reported elsewhere (Kennedy, Petrovic & Keppell, 1998) and will not be reviewed in detail here. Evaluation criteria were classed into three domains (instructional and conceptual design, interface and graphic design, and user attitudes and affect). Each of these domains contained a series of specific sub-criteria. The three evaluation domains and

the sub-criteria were useful guiding the evaluation process and provided a coherency within and across evaluations. While general comments and perceptions of packages were actively encouraged, it was important to have some consistency in the focus across the responses of reviewers.

The evaluation domains and criteria provided a detailed and clearly specified framework which was used to introduce concepts of evaluation to developers. The framework alerted developers to the issues which should be considered in development evaluation and the complexities inherent in the evaluation process. Many developers embarked on their project with misconceptions or assumptions about the process of evaluation. Many, for example, considered evaluation of learning material to be simply about determining whether students had met the learning objectives. The evaluation domains and criteria were a useful tool in dispelling these misconceptions. Many developers commented that they had not considered a number of the computer related aspects of development which were articulated in the domains of instructional and conceptual design and interface and graphic design.

As expected, developers were keen to evaluate specific or unique aspects of their packages. Most CFL packages used particular instructional or multimedia techniques as part of their teaching and learning strategy. For example, one package made particular use of video and audio and attempted to create a “virtual interview” situation for students. Another package aimed to individualise instruction by creating “learning loops” within the package which were activated when students prior knowledge was deemed to be lacking. The use of these unique strategies on the part of developers demanded a flexible approach to the evaluation criteria. A flexible approach enabled specific aspects of particular packages to be incorporated into the evaluation.

While the evaluation domains and criteria generally served their intended purpose, it became clear there was a need to focus more attention on the pedagogical underpinnings of the CFL packages being developed. The sub-criteria in the domain of instructional and conceptual design provided a useful guide for developers at a micro level. That is, developers were required to consider design issues such as the sequence and structure of their content, the level of interactivity, the learning outcomes of the package and the system of navigation. However, often an evaluation of the overarching philosophy of the package was neglected. Once a package reached a coherent form, often it was not clear whether development had been guided by particular learning and instructional theories or pedagogical philosophies.

There seems to be no easy solution to this problem—altering the epistemological and educational beliefs of academic staff is not easy. As with the research of Bain, McNaught, Mills and Lueckenhausen (1998), Reeves’ (1992) pedagogical dimensions may be useful in this area. Dimensions such as epistemology, pedagogical philosophy and underlying psychology could be used to challenge researchers about the macro structure of their package. One aspect of the solution would involve educating developers about pedagogical approaches which may be drawn upon when constructing a multimedia package (eg situated cognition, problem or case-based approaches). While some developers are aware of these approaches, many are not, and only by making such macro strategies explicit and accessible will change take place in this area. Part of the solution also seems to be the need to ensure a close association between instructional designers and the development team. The involvement of an

instructional designer early on in the development process would be instrumental in the development teams adoption of a more general educational philosophy in the design of their package.

Data collection techniques

Four data collection techniques were advocated in the program of development evaluation (questionnaire, focus group, observation and expert review). Data collection techniques were not prescribed for developers, rather they were able to make choices from a number of options at different stages of the evaluation program (see Kennedy, 1999). Questionnaires were used universally by developers and all developers included at least one other data collection technique in the evaluation of their package.

Initially there was an emphasis on more quantitative measurement with the questionnaires. This emphasis quickly shifted to a greater reliance on qualitative measurement as qualitative data was regarded as more useful in meeting the goal of refining and improving CFL packages. However, the use of quantitative data was not discarded entirely. Quantitative approaches were useful for highlighting areas in a package which needed attention—usually indicated by extreme scale responses. However, *specific modifications* which were required were usually determined through qualitative means (whether open-ended questionnaires, focus groups or observation).

For example, when reviewing a CFL package on DNA transcription a questionnaire was developed which incorporated both scaled and open-ended questions. It was clear from the scaled responses that students were often confused about the transitions and links between screens. When this (quantitative) result was followed up in a focus group discussion on the package it became clear to the developers that students were not aware that an incorrect response to a quiz question sent them off on a “learning loop” to explain why their answer was incorrect. From this evaluation, developers were able to modify the package so that there was greater continuity in the transition between quiz questions and learning loops.
The use of student observation was particularly useful in the refinement and improvement of packages. Reeves (1993) noted that there may often be discrepancies between the perceptions of developers and students with regard to components of a CFL package. Observation is a useful method for drawing out these possible discrepancies. For many projects it became clear how students were using the package and how this differed from how the developers thought students would use it. One package, “The Skin Atlas”, was specifically designed to allow students to compare and contrast images of normal and abnormal skin. Observation revealed that students almost universally did not use the compare and contrast function. Rather, students moved between sections of the package to compare images—a much more inefficient way to make comparisons. The reason why students did not use the compare and contrast function seemed to be that many simply did not know it was available. Most students ignored the optional instructions section of the package. The developers of The Skin Atlas are considering more visible instructions which explain the functionality of the package.

On a more positive note observation was instrumental in documenting how students engaged with CFL packages. In one package, “Communicating with the Tired Patient”, students played the role of a doctor in a “virtual doctor-patient interview”.

Students were required to discriminate between different types of interview questions. After listening to a range of doctors' questions students made a selection then observed and listened to the patient's response. When attempting to discriminate between different types of doctors' questions, students were often observed to close their eyes so that they could concentrate fully on the audio. It seemed that students felt the need to concentrate in this way so that they selected the most appropriate question in their "virtual interview" with the patient. Such engagement would be difficult to document using any other form of data collection technique.

Evaluation perspectives

Most staff appreciated the need to consider a number of perspectives in the evaluation of their package. While for some projects there were minor difficulties in obtaining an appropriate sample of content experts or students, generally this did not present too many problems. Developers usually found content experts among their colleagues who were able to complete reviews of their packages. The BMU has a number of graphic designers who made themselves available for reviewing the interface and graphic design of packages. Instructional designers and educational evaluators were also available at the BMU who were employed regularly for expert review.

One concern in this area was the potential over-evaluation of students, given the number of CFL packages being developed in the Faculty. As this was predicted to be a problem, the program of evaluation was designed so that student numbers were kept to a minimum. Packages were reviewed by small numbers of students for the majority of the evaluation program and it was not

until the package approached the beta version stage that it was put in front of large numbers of students. This strategy seems to have alleviated the problem of student over-evaluation. However, it is predicted that this will be a continuing difficulty with more packages approaching beta version in the next six months.

Timing of the program

There were two key features regarding the timing of the development evaluation program. First, the program was tied to a typical CFL package development cycle (Keppell & Bennett, 1997). Second, the program itself was divided into four discreet stages with each stage having a number of evaluation resources and strategies allocated to it (Kennedy, 1999). An important goal of the program was to ensure evaluation strategies were accessible to developers. Aligning the program of evaluation with the CFL development cycle ensured that development teams could easily participate in the program regardless of their stage of development. It became clear that this was of practical significance as some development teams thought of evaluation as a last minute "add-on" to the development process. Tying evaluation to the development cycle and having defined stages meant that these development teams were not disadvantaged from having not participated in the entire program of evaluation.

The only difficulty encountered with the timing of the program of evaluation was that, to a certain degree, evaluation was contingent on the development team following a "typical" CFL package development cycle. While this was often the case, difficulties

did emerge at certain stages in the evaluation program. For example, the early stages of development were more difficult to evaluate especially when developers were reluctant to commit their ideas to paper. Instructional and conceptual design was evaluated in a very informal manner until concept maps or story boards were produced by developers. On a number of occasions such documents were never produced, thereby undermining this stage of the evaluation process. Closer project management of the development team would be required to circumvent this difficulty.

Coaching of staff

A number of strategies were developed to coach Faculty staff in the process of evaluation. Evaluation kits were developed which contained details of the evaluation domains and criteria, the evaluation stages, the strategies and techniques involved and a number of template questionnaires (Kennedy, 1998). A series of letters were also written to developers, detailing the nature of the program of evaluation and inviting them to participate. The response to these overtures was less than overwhelming. While most developers expressed an interest in participating in the program, they were often distracted by project management, providing content or producing the multimedia rather than evaluating it. It became clear that evaluation was not seen to be as important as completing the next piece of content or finalising the design of an animation.
There was little the BMU could do to counter this attitude among developers, other than maintaining its position that evaluation was integral to the effective development of CFL packages and, in the long run, would save time and money. The BMU attempted to raise the profile of evaluation by using a number of other strategies:

- Projects which were closely associated with the BMU were guided through the evaluation program, a process which helped exemplify its practical benefits.
- Evaluation was emphasised in project management meetings as an important aspect of the multimedia development process.
- The BMU conducted lectures and a workshop on multimedia design and development which incorporated evaluation strategies.
- The BMU assisted Faculty staff with multimedia grant applications, where the importance of evaluation was also emphasised.

In addition to these internal proactive strategies, the committee which funds multimedia development for teaching and learning in the University assisted the BMU’s cause by stipulating that evaluation needed to be addressed both in grant applications and in reports on the progress of grants once funding had been received.

Over time these strategies combined with a clear evaluation framework and “on-call” advice have achieved the desired effect—to highlight the importance of evaluation and to foster a culture of evaluation in the Faculty. Once developers were actively involved in the program of evaluation they were very receptive to it. However, a proactive approach on the part of the BMU was still required to maintain developers attention on evaluation, even when they were part of the program. Thus, the goal of making developers independent in their evaluation efforts was not entirely achieved. By using the evaluation kit and template questionnaires, developers found it easy to plan their evaluation strategy, however, most developers needed to be reminded to actually implement this strategy. Evaluations are still being planned at the last minute

and it seems unlikely this will change unless more time is allocated to the management of the program.

General implementation

In general, the program of development evaluation was positively received in the Faculty. The clearest evidence of this is the fact that the program has over 15 multimedia projects associated with it. The areas of evaluation criteria, data collection techniques, participants and timing were seemingly well planned and their implementation was, on the whole, very successful. There were, however, a number of issues associated with the implementation of the program which presented unexpected difficulties.

One problem was that it was not clear how the results of an evaluation should be handled by the development team. If evaluations are comprehensive, suggestions about many of aspects of the CFL package will arrive from a variety of sources and these opinions may not always concur. It was not stipulated in the design of the evaluation program who should take responsibility for deciding what changes would actually be made as a result of an evaluation. Often this difficulty was avoided as a consensus was reached among the development team. However, there was a need to have a designated individual who would make decisions about alterations when consensus was more difficult or when changes were constantly being suggested (which often occurred when the package was close to completion). This person could potentially be the content expert, project manager, instructional designer, graphic designer or evaluator. Once the role of the BMU was established and accepted in the Faculty, generally a “gate-keeper” was appointed for each project through informal discussion with the development team.

One of the most useful evaluation techniques was the discussion which occurred in the regular meetings between developers, graphic designers, instructional designers, content experts and evaluators. These meetings were not specifically planned in the evaluation program but, because of their fruitfulness, were actively encouraged in the development process. Many critical and influential decisions were made about the development of the package in this environment. There was a need, however, to ensure that developers and/or evaluators documented these more casual evaluation procedures. Such formal documentation was advocated to record the decisions which had been made by the development team, and to assist with the process of writing reports and publications at a later date.

One of the contextual constraints on the evaluation program was that there was only one part-time position available to manage it. As a result the program was designed so it would, to a certain extent, run itself and developers were encouraged to be as independent as possible. It became clear that despite the BMU’s best efforts, the scope of the evaluation program necessitated a high degree of involvement by the BMU’s evaluator. This difficulty has been alleviated to a certain extent by stipulating core responsibilities of the BMU’s evaluator. Any additional evaluation needs require financial support from the development team. Development teams are generally content with this arrangement with many including an evaluation budget in their grant proposals.

Conclusion and areas of improvement

The process of critical reflection described in this paper has allowed the BMU to make conclusions about the overall efficacy of the program. The program has, on the whole met the goals outlined at the beginning of this article (except with regards to the time invested in the management of the program). Certain issues clearly still need to be addressed. Some changes, such as the emphasis on qualitative rather than quantitative measurement, have already been instituted. A number of areas where improvements could be made are listed below.

- More attention needs to be given to the pedagogical underpinning of projects, especially early on in the development process. This highlights the critical role of instructional designers.
- Careful documentation should take place when informal evaluations of packages are carried out. This is especially apparent early on in the development cycle when often critical design decisions are made.
- A constant proactive approach is required with developers regarding evaluation. Even when developers are aware of their evaluation options they need to be reminded to implement these options.
- Early evaluation, unless very informal, is dependent on the development team following an established design and development process and committing ideas to concept maps or story boards. Close project management is required in these phases of development if more formal evaluations are to be established.
- Designated decision makers ("gatekeepers") are required in the development team to institute changes which have resulted from evaluations. This is particularly apparent in the final stages of development.

References

Bain, J.D., McNaught, C. Mills, C. & Lueckenhausen, G. (1998). Understanding CFL products in higher education in terms of academics' educational beliefs: Enhancing Reeves' analysis. In R.M. Corderoy (Ed.) Proceedings of the fifteenth annual conference of Australian Society for Computers in Tertiary Education (ASCILITE) (pp. 49-58). University of Wollongong, Australia.

Kennedy, G.E. (1998). *Computer Aided Learning: Formative evaluation questionnaires.* Biomedical Multimedia Unit, University of Melbourne.

Kennedy, G.E. (1999). *Defining the dimensions of a formative evaluation program: A Multi-method, multi-perspective approach to the evaluation of multimedia*. In B. Collis & R. Oliver (Eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecomunications (Ed-Media 99) (pp. 604-609). Charlottesville, AACA.

Kennedy, G.E., Petrovic, T. & Keppell, M. (1998). *A program of evaluation for multimedia technologies being developed in the Faculty of Medicine, Dentistry and Health Sciences.* In R.M. Corderoy (Ed.) Proceedings of the fifteenth annual conference of Australian Society for Computers in Tertiary Education (ASCILITE) (pp. 407-415). University of Wollongong, Australia.

Keppell, M. & Bennett, S. (1997). *Digital Learning. A guide to multimedia and on-line development in the Interactive Multimedia Unit*. Central Queensland University.

Keppell, M., Elliott, K. & Harris, P. (1998). *Problem based learning and multimedia: Innovation for improved learning of medical concepts*. In R.M. Corderoy (Ed.) Proceedings of the fifteenth annual conference of Australian Society for Computers in Tertiary Education (ASCILITE) (pp. 417-424). University of Wollongong, NSW.

Reeves, T. C. (1992, September). Effective dimensions of interactive learning systems. Information Technology for Training and Education Conference (ITTE `92) (pp. 99-113). University of Queensland, Australia.

Reeves, T.C. (1993). Evaluating interactive multimedia. In D.M. Gayeski (Ed.) *Multimedia for Learning. Development, application, evaluation* (pp. 97-112), Educational Technology Publications.