AN EXPLORATION OF TOOLS, TECHNIQUES AND PROCEDURES FOR EVALUATING INFORMATIONAL/EDUCATIONAL MULTIMEDIA SOFTWARE

Serge Walberg

Thesis submitted for the degree of Bachelor of Communications (Honours)

Faculty of Communications, Health and Science
School of Communications and Multimedia

20th December 2000

Use of Thesis

Abstract

This thesis explored the evaluation of informational/educational systems. It investigated the features of the design of an evaluation program needed to conduct the summative evaluation of informational and educational software.

In designing this evaluation methodology, several questions were addressed:

- What is needed to be achieved from an evaluation?
- Which characteristics/features of a system required evaluation?
- How was the evaluation going to be conducted?
- How can future developers use the information obtained from the evaluation?
- What were the scope and the limitations of the proposed evaluation?
- How can evaluation tools, techniques and procedures developed by this research, be useful in the design of evaluation programs for similar informational/educational software

To answer these questions an in-depth review of the available literature and electronic resources that address the area of multimedia software evaluation was conducted. In order to identify the best tools to conduct an efficient and effective evaluation it was also imperative to examine and assess tools and techniques already developed. It was then possible to proceed with the design and construction of an evaluation program, based on a thorough familiarity with and understanding of previous work conducted in this area.

The EduKit2000 CD produced and distributed by Edith Cowan University to all commencing and external students was selected as a case study with which to implement and test the evaluation program. Because of the availability and the proximity of both the developers and the end-users, it was considered an appropriate choice.

The evaluation program developed involved the use of four evaluation tools. These consisted of an expert review, a questionnaire, user and novice testing. The questionnaire was sent out with the CD. The feedback received from the questionnaire was analysed in conjunction with the data returned from the other evaluation tools and the summative evaluation of the CD was conducted, in order to determine whether the product reflected the requirements of its developers. Analysis of the returned research data aimed to discover what future changes may be required to be implemented to the original product in order for it to fulfil those requirements, while still remaining a cost-effective, freely distributed product. More importantly, this evaluation of EduKit2000 was intended as a “road test” of the methodology developed by this

research. It made it possible to identify and improve some features of the evaluation plan and strategy that did not perform as well as anticipated.

The research conducted for the purpose of this thesis explored the evaluation of informational systems and achieved four distinct objectives:

1. A methodology was researched, designed and developed to comprehensively evaluate informational/educational systems
2. This methodology was tested by using a case study (EduKit2000) to which it was applied.
3. The methodology was evaluated and reviewed, and recommendations were formulated for improvements.
4. A comprehensive evaluation of EduKit2000 was produced, along with recommendations for improving the product.

By adequately identifying those areas that can be *improved* in order for the product to fulfil its objectives, the evaluation program developed by this research was demonstrated to be an effective, and cost-effective, methodology for evaluations of informational/educational software.

Declaration

I certify that this thesis does not, to the best of my knowledge and belief:

- incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;
- contain any material previously published or written by another person except where due reference is made in the text; or
- contain any defamatory material.

Signed:…….

Serge Walberg, December 2000

Acknowledgments

I would like to thank my supervisor, Dr Ron Oliver, for his tireless support and assistance during the course of this research project. His advice, recommendations and provision of material have been invaluable to the completion of this Honours thesis.

I also wish to acknowledge the help received from the four experts who voluntarily participated in the expert review of the case study. These were Dr. Arshad Omari, and Messrs. Joe Luca, Mark McMahon and Steven Doyle, to whom I am deeply indebted and eternally grateful for their collaboration.

Last, but certainly not least, I am grateful to the students who participated in the evaluation. This includes the 10 novice users, the 10 end users, and the 56 students who returned questionnaires.

An Exploration Of Tools, Techniques And Procedures For Evaluating Informational/Educational Multimedia Software

CONTENTS

Chapter 1. Introduction

The following stages were used in the conduct of the research:

1. *A methodology* was developed that will enable evaluators of informational/educational systems and similar products to have access to a program for evaluating this type of software, including detailed guidelines of tools, techniques and procedures.
2. *The evaluation program* designed was *implemented and tested* on EduKit2000.
3. *An improved methodology* was obtained as a result of a final review of the evaluation program developed.
4. *Recommendations were proposed to the developers* of EduKit2000 for improving the product: this will enable the university to produce and distribute a more effective (and cost-effective) product next year.

Universities and other education institutions are increasingly developing informational packages, including course information, collections of software, administration procedures and promotional material. In Western Australia alone, three universities have already produced and distributed such packages: Edith Cowan University (EduKit), the University of Western Australia (ELVIS: Electronically Linked Visual Information System) and Curtin University. These are seen as a cost effective means of communicating important information and/or training to students and potential students, and shall be referred to in this research as *informational/educational systems*. There is a growing use by Australian universities of interactive resources (e.g. Bennett, Priest, and Macpherson, 1999); and researchers maintain that "technology is increasingly being used to supplement traditional face-to-face communication in business and education." (McLaughlan and Kirkpatrick, 1999, 243-257). Clearly, interactive and online informational software is useful for educational purposes, primarily because of the flexibility of the learning environment it provides.

Informational/educational systems, both on-line and CD based, allow students to engage in "individual and collaborative learning at times and places that suit them" (Collings, Pearce and Walker, 1998, 9-16). They allow students to be "active participants in their own learning" (Collings et al., 1998, 9-16), an important element in the constructivist approach to learning (Jonassen, 1999; Ewing et al., 1998; Foxwell, 1998) using computer-mediated training.

Because of the growing use of informational/educational systems, there is a correspondingly growing need to develop methodologies for evaluating these systems in order to determine first,

whether they achieve the objectives they have been set, and secondly to enable the design of improvements to successive generations of the products.

What are the characteristics of the tools and procedures required for the evaluation of these informational/educational multimedia systems and software? This research endeavoured to answer this question by developing an evaluation methodology for informational/educational systems. The methodology was tested by applying it to the evaluation of a case study, the EduKit2000 CD produced and distributed by Edith Cowan University. A set of recommendations for improving the system resulted from applying this evaluation to the test case.

The following chapters describe the process of developing the evaluation methodology, which is intended to be generically applicable to the evaluation of any informational/educational system.

1.1. Significance of research

Because it is essential to evaluate informational/educational multimedia systems in order for them to be improved, it is important to develop a methodology for doing this that will be applicable to all that range of software. This is the basic rationale of the study, and although much of the research conducted constituted a usability study for a specific piece of informational/educational multimedia software, it has more generalised applications. By determining which evaluation tools, techniques, and procedures can be applied to the evaluation of this type of software package, a methodology was developed which can be adapted and applied to the evaluation of other similar software, or to future incarnations of the same product.

Most multimedia software (applications, utilities, games and educational programs) is nowadays upgraded on a regular ongoing basis, to adapt to fast improving technology. Informational/educational software needs regular upgrading for the added reason that it must reflect changes in the instructional *content* of the program, as well as its presentation in order to achieve optimum educational outcomes (Tweddle et al., 1998). Because of the growing use of informational/educational systems, there is a correspondingly growing need for evaluation methodologies. If software is not effectively evaluated and improved, it may not achieve its primary objectives, resulting in instructional failure and financial losses. More importantly, without evaluation no product can be improved. This research sought to provide a comprehensive methodology, including a complete set of tools, templates, and procedural guidelines to enable future evaluations of this or similar multimedia products.

While each individual product will require its specific parameters to be applied to the evaluation methodology presented here, the techniques and procedures are anticipated to be broadly applicable to all informational/educational systems.

Chapter 2. Developing a Framework using a Review of Literature

This chapter explores the available literature and research conducted by theorists and previous researchers who have addressed the evaluation of multimedia software. By comparing the available methodologies, a framework was developed identifying the principal areas of inquiry that need to be assessed in order to conduct an effective evaluation. An examination of tools and processes available for assessing each area of inquiry is also presented in order to proceed with the development of an evaluation strategy.

There are basically two types of software evaluations:

Formative Evaluation:

"Formative" evaluation refers to structured evaluation that is provided while the course is ongoing so as to permit improvements (Scriven, 1967). Formative evaluation can best be described as an ongoing assessment during the phases of design and production of a piece of software in order to improve it. All multimedia software undergoes some formative evaluation during the development phase. This is ensured using quality assurance procedures, and enables the resolution of problems and bugs during production.

Summative Evaluation:

"*Summative* evaluation is evaluation done after software design and production is complete in order to establish its performance and properties." (Draper et al 1997, 103). This form of evaluation presents the evaluators' conclusions relating to the quality, validity or worth of the multimedia product. The process begins after the product has been designed and produced, and is useful in determining ways in which it can be improved before final delivery to the end-user.

Summative Evaluation takes place after all modifications to a program have been made, after the program has been in place long enough to stabilise, and after the impact of the program has had a chance to be realised.

Increasingly, theorists support the view that the distinction between the two should be minimised and that evaluation activities should be seen as an integral part of good design practice and not something that is external to the design process (Schön, 1983).

During the timeline of a project's development, formative evaluation is generally seen as more important than summative evaluation. As time progresses, this relationship is inverted, as can be seen from Figure 1 below (Rasmussen, 1993).

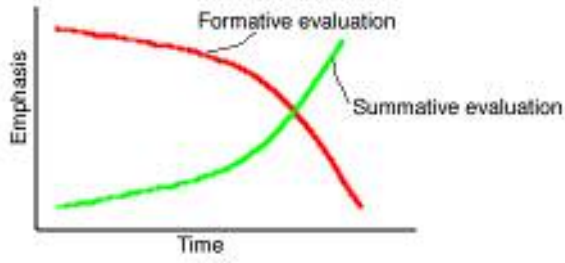

Figure 1. The relative importance of formative and summative evaluation over time

Evaluation of informational/educational multimedia systems must be an iterative combination of both summative and formative evaluations. Because it is often annually upgradeable (to reflect the needs of each generation of students in terms of courseware, software requirements, etc.) these systems must be summatively evaluated at the end of each production cycle. However since each year sees the initiation of a new production cycle, using evaluation results to improve the new incarnation, this constitutes a form of the summative/formative evaluation described by Schön (1983), above.

2.1. Evaluation methodologies

"The majority of software evaluations are not only flawed and largely inappropriate, but are far from being truly objective… " (Tucker, 1989, 8-16).

With education fast becoming a *user-pay,* profit-motivated, training scheme for corporate business' human resource requirements, a real danger exists that the quality of education delivered will deteriorate in the interest of economic rationalism. It must therefore be a primary concern for educators and learners alike to maintain as high a quality as possible on any educational or informational programme.

Evaluation of multimedia software has been interpreted in a wide range of different ways by theorists and previous researchers, some agreeing with Tucker (1989), that new methodologies for evaluation need to be developed. However most agree with Collings, (1998) that the general purpose of evaluation is "to provide input to an iterative and participative design process" (Collings et al., 1998, 287). Evidence of consensus over this view is abundant in the available literature (Nielsen, 1992; Monk et al., 1993; Lindgaard, 1994; Spool et al., 1999).

Notwithstanding, while there is general agreement as to the purpose of software evaluation, the particular areas of investigation and assessment (i.e. interface design, navigation, interactivity, content, scope, functionality, etc.) differ widely between researchers, often depending on their own areas of expertise. This work has aimed at identifying as precisely as possible exactly those areas of inquiry to be investigated so as to produce a methodology that will provide optimum results when evaluating any informational/educational system. In order to design the most effective evaluation program possible it was important to identify which areas of inquiry would be addressed.

2.2. Areas of Inquiry

Previous research in evaluation methodology has focused primarily on evaluating and improving the educational *content* and *instructional design* of informational/educational software, rather than the multimedia delivery platform itself (Alessi and Trollip, 1991). Some, such as Reeves (1993), have addressed the functionality of interface design.

Reeves, (1993) addresses the issue of interface evaluation using a tool he calls "Dimensions". While it is limited in its capacity to deliver precise values which can be used for comparative research, the tool does address most aspects of interface design using 10 "*Dimensions*": ease of use, navigation cognitive load, mapping, screen design, knowledge space compatibility, information presentation, media integration, aesthetics and overall functionality. Sims, (1999) is specifically concerned with educational outcomes, and adds 3 more "Dimensions"; control, adaptation and communication

Other prominent evaluators have also focused on the relative importance of interface design in informational/educational software; Laurel stresses: "Everything about the interface should engage the user to accomplish the task" (Laurel, 1990).

Barker and King's approach resembles Reeves 'Dimensions' by listing a set of categories as areas of evaluation:

1. Engagement (Interface Design)
2. Interactivity
3. Tailorability (Scope)
4. Appropriateness of multimedia mix
5. Mode and style of interaction
6. Quality of interaction
7. Quality of end-user interface

8. Learning styles
9. Monitoring and assessment techniques
10. Built-in intelligence
11. Adequacy of ancillary learning support tools
12. Suitability for single user group distributed use

However, not all research has focused on interface design. Many theorists also consider navigational structure and interactivity vitally important. "..a product is assessed on whether it offers both passive and active interaction with the user, and whether it provides the means by which a high degree of involvement is achieved." (Barker and King, 1993, 307-319).

Interactivity is the degree of communication and feedback between the user and the system. It exists in a programme

- where the user can ask questions and receive answers,
- where the user can search for topics
- where the user can access a variety of information packages and media by navigating through the programme
- where the programme will respond in different ways (reactively) depending on the users input.
- where a user support system is in place for guidance and help

The navigational structure of a multimedia product is clearly of critical importance. Some, like Barker and King (1993) consider it the most important hallmark of quality. By the very nature of an informational/educational system (encyclopaedic and informational), the user is clearly not expected to read or assimilate *all* the material presented. The content is too vast, and users are assumed to be searching for specific information. For this reason, and particularly because of the huge spread and variety of content, efficient navigation is of prime importance (Utting and Yankelovitch, 1989). Users are required to be able to locate the information they are looking for with no (or few) prerequisite computer skills.

It is important that users do not become confused or lost in the CD's structure, as this would result in them abandoning their search. Mapping, or orientation is important with this kind of system and is generally considered one of the attributes of good navigation design.

"An important aspect of navigation is orientation" writes Yankelovitch "this is a critical variable because users frequently complain of being lost in an interactive program." Utting and Yankelovitch, (1989, 58-84).

However, navigation should also be designed to enable experienced users, such as administrative staff, secretarial personnel, and advanced students to by-pass time-consuming procedures and rapidly access those services they need. This may require parallel navigational structures, as “lowest-common-denominator” navigational architecture might not be the most efficient in every case (Luca, 1996).

Most researchers agree on the importance of interactivity/navigation. Morrison considers it *the* major criterion to be used in evaluation, and describes it as: ".. the learner in conversation with himself over the material to be learned" (Morrison, 1987, 134-138).

Finally the overall scope of the system must be evaluated. Is it too large? Too small? Too narrow? Particularly in the case of educational software, scope must be very accurately predetermined. "Organised distance education,…requires scope planning to be useful." (Holmberg 1989).

While these areas of inquiry and evaluation are clearly essential, other areas can have an equally important role to play in the comprehensive evaluation of informational/educational systems. One increasingly important area is functionality, both electronic and instructional. Collings et al., (1998) include it in the list of heuristics for their research, while Reeves (1993), Barker and King (1993) all include functionality as one of their primary evaluation "Dimensions". Because functionality relates not just to the instructional design of information but also to the electronic integrity of the system (working hyperlinks, missing graphics, slow loading pages) and of its overall performance, functionality cannot be omitted from the areas of inquiry.

It can be seen that most previous research has identified *content*, *scope*, *interactivity/navigational structure* and *interface design* as being of critical importance in a multimedia product of this kind. Even when a larger number of variables has been proposed, they still relate to these five areas of inquiry.

For example, the following table lists Reeve's (1993) Dimensions and Barker and Kings’ (1993) Categories. These can both be seen to ultimately address the same five areas of inquiry.

Table 2.1 A comparison of areas of inquiry

Reeves	Barker and King
Ease of use, (Navigation)	Engagement (Interface Design)
Navigation (Navigation)	Interactivity (Interactivity/Navigation)
Cognitive load,	Tailorability (Scope)
Mapping, (Navigation)	Appropriateness of multimedia mix
Screen design, (Interface Design)	Mode and style of interaction (Interactivity/Navigation)
Knowledge Space Compatibility,	Quality of interaction (Interactivity/Navigation)
Media integration, (Interface Design)	Quality of end-user interface
Aesthetics, (Interface Design)	Learning styles (Content)
Overall functionality (Functionality)	Monitoring and assessment techniques
Information presentation, (Scope/Content)	Adequacy of ancillary learning support tools (Instructional Functionality)
	Built-in intelligence (Electronic Functionality)
	Suitability for single user group distributed use

As a result of assessing the resources provided by previous researchers and theorists, using both published and WWW material, the areas necessary for the evaluation of informational/educational systems have been clearly identified. Previous research has consistently earmarked these areas, and they are consistent with the objectives of this research:

a) Interactivity/Navigational Structure
b) Interface Design
c) Quality of Content
d) Scope
e) Functionality

a) Interactivity/Navigational Structure

This area involves the evaluation of the mechanisms of navigation, the ease of use to the novice, the existence of alternative navigational mechanisms, the accuracy and integrity of all links and hyperlinks, the existence of mapping features to enable users to know where they are in the system, the existence of mechanisms for the users to input data and receive results, the existence of *help* and *search* facilities, the quality of the interaction and the mode and style of interactivity.

b) Interface Design

This involves evaluating the aesthetic appeal of the interface, the appropriateness of the multimedia mix (sound, graphics, animations, video, etc.), the legibility of the text, the appropriate use of colours, fonts, sizes, shading, layout, formatting of text and images, the clarity and intuitiveness of icons, signs and symbols, the engagement level of the interface and the order disposition of screen elements.

c) Quality of content

This includes the exploration of all content material (text, graphics, media) for accuracy and integrity. Also, text should be correct (grammar, spelling, syntax, etc.) and provide aesthetic appeal. Instructions provided for performing tasks are tested to verify that they do in fact enable users to perform the tasks. Learning styles and the presentation of information are assessed for efficiency and the effectiveness of presented material in achieving its objectives is established.

d) Scope

Here, the amount of material presented is assessed for its appropriateness. Missing areas of information are identified, as well as excessive or repetitive presentation of unnecessary material. The appropriateness of the material in relation to the target audience is also determined, as well as the existence of duplicate material or media. Scope should be based on an accurate assessment of the aims and objectives of the software investigated.

e) Overall functionality

Overall functionality can be subdivided into *instructional* functionality and *electronic* functionality. For instructional functionality the existence of features such as *help* and *search*, disabled users' functionality such as *zoom* options, etc. is investigated as well as such features as user-input processing, *Frequently Asked Questions* sections, and the provision of mechanisms for displaying media, downloading plug-ins and replaying videos.

Electronic functionality exists where a system has no broken links, missing graphics, slow loading pages, duplicate files, or missing pages. Devices such as volume controls, video playback controls, zooms, printers and converters should all function correctly.

2.2.1. Processes and Tools for Evaluation

There is an endless (and growing) number of evaluation processes available and in constant use, from postal surveys to door-knocking interviewers, and it is important to use exactly the appropriate ones. For the purpose of this research it was imperative to select precisely those

tools that would best evaluate the areas of inquiry described above. In order to do this it was necessary to investigate and assess all the commonly used instruments, and to determine their degree of appropriateness to our purpose.

Heidler, (1993) of the Multimedia in Manufacturing Education Lab at the Georgia Institute of Technology presents five on-line sets of multimedia development tools, hyperlinked to an in-depth description of each one. It includes, in addition to a set of Evaluation tools a description of customisable tools for analysis, design, management and production. Thirty-nine tools are presented, ranging from *analysis report templates* to *objective review checklists*. The section dealing with evaluation tools lists 10 items, which are described below: evaluation matrix, anecdotal record form, expert review checklist, focus group protocol, formative review log, implementation log, interview protocol, questionnaire, user interface rating form and evaluation report sample. While he presents a concise listing and description of the *modus operandi* of each tool, he fails the user by falling far short of providing any useful guidelines or instructions relating to the judicious *use* of the tools. While tools are essential for conducting software evaluation, *how* these tools are used is of even more critical importance.

As no method of evaluation will detect all errors, additional tools and techniques need to be constantly developed. A very effective combination for evaluation is to conduct user testing in conjunction with an expert review. There are two major reasons for selecting these tools. First, a heuristic (expert) evaluation can eliminate a number of usability problems without the need to "waste users," who sometimes can be difficult to find. Second, these two categories of usability assessment methods have been shown to identify distinct sets of usability problems; therefore, they supplement each other rather than lead to repetitive findings. Researchers such as Desurvire et al. (1992), Jeffries et al. (1991) and Karat et al. (1992) all came to this conclusion.

Heuristic evaluations require a team of system experts. The experts' report supplements the user tests findings, in a technique developed by Nielsen (1990).

2.2.2. Processes

With the panacea of available evaluation processes, it is important to identify and define the most important (and appropriate) ones. Heidler's (1993) web based system (described above) lists several of these. In order to select the most appropriate for the purposes of this research, several more were investigated:

1. Anecdotal Record Form

This is a means of collecting qualitative data in the form of an anecdote related by the user in *hir* own words. A template form is used recording date, time, and person(s) involved. Each anecdote relates to a single incident, and brings a "human" element to the evaluation. (Heidler, 1993). These can be expensive and time-consuming but have the advantage of returning very specific data.

2. Expert Review

A particular form of *heuristic* expert evaluation was developed by Nielsen and Molich (1990), and is a method for structuring the critique of a system. It involves several evaluators *independently* evaluating a system to identify potential user problems. Nielsen and Molich's (1990) experience indicates that five evaluators usually results in 75% of overall usability problems being identified. The evaluators then confer and their findings are aggregated. The recommended procedure is for a small number (between four and six) of evaluators to apply a set of '*heuristics*', of which below is a *sample* list:

- Visibility of system status;
- Match between the system and the real world (Accuracy of content);
- User control and freedom (Interactivity/Navigational Structure);
- Consistency and standards (Quality of content);
- Error prevention (Electronic Functionality);
- Flexibility and efficiency of use;
- Aesthetic and minimalist design (Interface Design); and
- Help and documentation (Instructional Functionality).

Experts regard heuristic evaluation as one of the most cost-effective and widely used usability investigation tools currently available (Nielsen, 1992). Heuristic expert evaluation is used to detect minor and major errors in the user interface of a product and does so extremely well. The rates of detection are about 42 percent for major usability problems and 32 percent for minor problems, when a single evaluator is used (Nielsen, 1992). Though this figure is impressive it is improved to approximately 75% for both minor and major errors when a team of five evaluators is used. (Nielsen and Molich, 1990).

Apart form the remarkable rate of error detection that a heuristic expert evaluation has to offer there are a host of other benefits. The expert evaluation also provides a high level of flexibility as the number of evaluators can be regulated, as well as the stage in the usability evaluation lifecycle at which it is implemented.

To carry out an appropriate evaluation of a multimedia product, only objective experts in the field have the necessary skills and expertise, especially in relation to the navigation and interface design aspects. Heuristic expert evaluation is not only the most appropriate but also the most cost-effective methodology to employ as it can extract the most out of the experts conducting the evaluation. An added advantage is its ability to evaluate the product in a short period of time, due to the fact that the experts have the opportunity to discuss their findings upon completion of their individual investigation. To aid in the rapid completion of the evaluation, investigations can occur simultaneously.

If none of the experts have been involved in the design and construction of the product, it can be assumed that the results and findings will not be biased in any way.

3. Formative Review Log

This process is used during development of a product, and is valuable as it provides feedback to developers while the product is still in production (Heidler, 1993). Tools used to conduct a Formative Review Log usually consist of a simple instrument with three columns, the first for recording the screen or format sheet number that the person is reviewing, the second for writing down observations (e.g., errors, confusing points, or ideas), and the third for recording what actions have been taken in reaction to the feedback provided by the end-user.

4. Questionnaire

A questionnaire is an excellent instrument for summative evaluation in that it can address any number of issues and can be implemented at a scale (superficial or in-depth) proportional to the *time/cost* constraints of the evaluation. The questionnaire once constructed is relatively cheap to implement, as volunteers are usually used. The only additional time factor involved is analysing the results and summarising the information for reports. The questionnaire should also detect any problems that the experts have missed.

As this form of evaluation is usually carried out by volunteers who have not been involved in the production of the product (and are of varying ages and abilities), it can be guaranteed that the results gathered and the reports generated will be truly objective and without prejudice, adding to the reliability of the results.

A *Likert* scale questionnaire is sometimes more effective than a standard one because more data for each question is generated (several possible responses per question). With more data available problem areas that would be missed with a standard questionnaire can more

effectively be deduced. The questionnaire also gives the ability to gather user opinions and areas they consider problematic.

Other processes abound. Collings describes processes that involve "Expert Walkthrough or inspections: others involve users who provide feedback by undertaking typical tasks." (Collings et al., 1998). Very extensive work has been conducted in this area by researchers like Faraday and Sutcliffe, (1995a).

5. Audit Trail recordings

Users can be observed, videotaped or recorded while they test the system being evaluated by performing typical tasks. Software is also available which can automate these observation sessions. For example, a piece of software called Lotus ScreenCam allows evaluators to record all mouse movements and navigational jumps (as well as recording verbal comments) made by the users while they are navigating through the site. This is a useful tool to use in conjunction with a questionnaire; together they illustrate how the software was used as well as what users thought of it. Audit trail observations/recordings are particularly useful in identifying navigational problems in software as they allow an analysis of the navigational paths, errors and obstacles of users. Scope and quality of content material are also assessed by observations, as well as the organisation of the presented information.

Other evaluation processes exist, such as interviews, walk-throughs, etc. All of these are useful in different specific cases. It is important to identify which ones are the most appropriate for the purpose of this research. Before a selection was made however, an investigation of the available tools and templates was necessary. These tools also need to be appropriate to the objectives of the evaluation, and may determine which processes are employed.

2.2.3. Tools and Templates

Most of the processes described above require the use of tools and templates to be applied. The questionnaire process for example might require the use of a rating tool (described below). Similarly the audit trail observations will require procedural guidelines for conducting observation sessions, and expert reviews often utilise a set of heuristics to be applied to the software investigated. Below is an exploration of some of these tools and templates.

1. Evaluation Matrix

This matrix enables evaluators to consider a wide range of data collection methods for each question requiring to be addressed. (Heidler, 1993).

Questions are listed on the vertical side of the matrix, and a list of the feasible data collection methods is tabulated on the horizontal side of the matrix. Each question is considered carefully, and the most appropriate data collection method is selected.

The matrix is customisable by replacing both sets of variables. An evaluation matrix has been used in *this* research for determining which tools would be required for our case study, and is presented below.

2. User Interface Rating Form

Reeves' Interface Rating Tool, already described above is an example of a most commonly applied rating form.

Other tools exist relating to the evaluation of on-line material including the use of bulletin boards to collect feedback data (Millen, 1999), and the use of computer-to-computer conferencing supported by an on-line whiteboard (Hammontree et al., 1994, and Hartson et al., 1996).

2.3. Adopted Strategy

As demonstrated above, very little research has previously been conducted in the specific area of regularly upgraded informational/educational systems software. However useful work has been conducted on the design and use of evaluation tools. Resources on evaluating interface and style, made available by Barker and King (1993) as well as Reeves (1993) have been used. Laurel (1990) provided basic guidelines to human/computer interaction, critical to the evaluation of information systems. Nielsen's techniques of heuristic evaluations were the basis of our own heuristics, and several university Internet resources (referenced below) provided excellent (and up-to-date) research on the evaluation of instructional content and learning outcomes.

An Evaluation Matrix was constructed to identify the tools considered most appropriate for this research. Since the areas of inquiry had already been established, it was then necessary to map these against a grid of available tools. Using information obtained from research of the literature, and described above, the best tools identified by previous theorists for each area of inquiry were mapped on the grid. The Evaluation Matrix used is shown in Table 2.2.

Table 2.2 Evaluation Matrix for tool selection

Area of Inquiry / **Tool**	**Questionnaire**	**Expert review**	**Observation**	**User Testing**
Interface Design	X	X		X
Interactivity/Navigational Structure		X	X	X
Quality of Content	X	X		
Scope	X	X		X
Overall Functionality		X	X	

Bearing in mind that certain combinations of tools have been identified as returning optimum results (see above), the *expert review* combined with the *user testing* were adopted as the best tools for *interface design* evaluation. A combination of *user testing* and *observation* was adopted as the best for evaluating *interactivity/navigational structure*. Because of the large population of potential respondents, the *questionnaire* was considered the best tool for evaluating both the *scope* and the *quality of content*, as well as providing useful data as to *interface design.*

The processes selected to perform the evaluation were chosen because previous research conducted indicated they were the most appropriate for our purposes. They are listed below with a brief listing of the advantages each one represents, and an indication of the theorists and researchers (referred to above) who have recommended them:

1. Questionnaire (e.g. Heidler, 1993; Hannafin, 1988;)
    i. Best when the whole population can be used
    ii. Best identifies interface design problems
    iii. Cost effective
    iv. Simple to implement
2. Expert Review (e.g. Nielsen and Molich, 1993;)
    v. Best identifies content quality and scope problems
    vi. five experts can resolve 75% of usability problems
    vii. Experts can recommend improvements
    viii. Usually expensive, in this case free
    ix. Experts are available at Edith Cowan University (SCAM)
3. Audit trail observation (e.g. Draper, et al., 1994;)
    x. Most informative
    xi. Can be automated (Lotus ScreenCam)

xii. Best identifies instructional functionality problems
xiii. Selection of sample population
xiv. Usually expensive, in this case free

4. User testing (e.g. Faraday and Sutcliffe, 1995; Duchastel, 1987; Collings et al., 1998;)
xv. Most reliable data
xvi. Best identifies navigation problems
xvii. Usually expensive, in this case free

A combination of the four processes described above was adopted as the evaluation strategy for this project, and constitutes the methodology developed for evaluating informational/educational multimedia systems. The choice of tools was based on an assessment of their effectiveness at evaluating each of the areas of inquiry identified above as appropriate for evaluating this type of software.

The evaluation plan developed and described above, consisted of a *questionnaire* targeting the entire user population, a heuristic *expert review* conducted by in-house experts (who are also users of the software), an *audit trail observation* of novice users selected from the user population, and a *user testing* session conducted by a representative sample of the user population. These four processes will identify problems, strengths and weaknesses of any informational/educational multimedia system that they are customised and adapted to.

The next phase of the project consisted in testing the evaluation program using a case study, in order to assess the efficacy of the system, and is described in the next chapter. As a result of testing the methodology on the case study (EduKit2000), an evaluation of that software was also produced.

Chapter 3. Testing/Implementing the Evaluation Program

In order to determine the degree of efficacy of the evaluation program described above, it was tested by applying it to a case study. This was done to identify any problematic areas or inconsistencies, hence allowing the system to be improved. This chapter provides a history and description of the product selected for the case study used to test the methodology, and the reasons for its selection. As a result of the test, an evaluation of the case study was obtained, which constituted the third objective of this research.

3.1. Selection of case study

The EduKit2000 CD produced and distributed by Edith Cowan University was selected as an appropriate case study with which to implement and test the evaluation program. Because of the availability and the proximity of both the developers and the end-users, it was considered an appropriate choice as this allowed the product to be tested in the environment in which it was being used, and on the whole population of end-users. EduKit2000 is an informational/educational system directed particularly at external and commencing students of Edith Cowan University.

3.2. History of product selected for evaluation

In order to accurately evaluate any multimedia software it is important to clearly identify the original objectives of the product, and to determine the extent to which it has attained these objectives. The primary purposes and objectives of EduKit2000 can be categorised into three distinct areas:

As an aid to students

In order primarily to reduce the disproportionately large attrition rate for commencing students, it was decided to provide all commencing and external students with an information package that would facilitate their integration into the university environment. This package would include information not just about their academic courses, but also administrative procedures, library access information, a collection of useful (and recreational) software and plug-ins, a description of the available support services (such as counselling, medical, career, chaplaincy and related services), as well as an overview of extra-curricular activities (clubs and societies, sports facilities, social activities, etc.) that are available to Edith Cowan University students. A "Printer Section" would also allow students to print out any necessary administrative forms and applications such as assignment cover sheets, parking permit applications, etc.

Clearly, if attrition rates could be reduced by even a small percentage, this would render EduKit2000 extremely cost-effective.

As a public relations exercise

In order to smoothen the major shift in the university's focus towards a cost-effective corporate style of operation, it was considered important to pre-empt any student reaction to corporate streamlining by highlighting a focus on students, and on services to students.

Furthermore, as a broader public relations exercise directed at the general public, it expressed and emphasised Edith Cowan University's confidence in its own teaching standards by indicating that the quality of multimedia produced by its own students is at least as high as that available commercially.

As cost savings for the Student Service Centre

Finally, but perhaps most importantly it was seen as an important part of the efforts to streamline human resource payroll budgets. By providing commencing students with most of the information they required to fulfil essential tasks such as connecting to the Edith Cowan University modem pool, EduKit2000 would decongest switchboards and IT Help Desk staff, allowing expensive human resources to be re-allocated to other areas.

Also, by allowing students to print out their own application forms and other administrative documents, further savings could be achieved.

Other areas and departments would also, incidentally, achieve savings through a reduced load on their services. These would include the library, the counselling services, faculties' and schools' administrative staff, etc.

The CD is intended to benefit both the University and the students. EduKit2000 encompasses many different departments, each dealing with different areas. Because of this, the CD is huge and contains over 400 pages of information. When this much information is provided it is crucial to provide a clear and simple interface supported by an intuitive, consistent navigational structure that allows users to locate and access the particular piece of information that they require, efficiently and effectively.

3.3. Characteristics of the selected product

EduKit2000 was entirely designed, developed, produced and distributed by Edith Cowan University students. It is a CD based informational/educational multimedia system providing information about courseware, university administrative procedures, extra-curricular activities and library catalogs. It also includes a large collection of useful software for commencing students, and instructions for connecting to the ECU modem pool.

The system was developed using Macromedia Director7 on G4 Macintosh machines, and was intended to run on all platforms. 14,000 CDs were produced and posted to all external and commencing students.

Chapter 4. Conducting the Evaluation

This chapter describes both the approach and the process followed for testing the methodology described above by applying it to the case study, EduKit2000. A detailed description of the processes, tools and procedures is provided, including guidelines for the use of each of the processes and instruments selected.

4.1. Evaluation Plan

Summative evaluation is evaluation done after software design and production is complete in order to "establish its performance and properties." (Draper, 1997, 16-87). The tools and methodology used for the evaluation must therefore investigate what the desired outcomes of the software are, and seek to provide guidelines for improving performance.

4.1.1. Approach

It is crucial for an institution to learn whether or not its informational/educational system is fulfilling the needs of the end-user, and of the institution itself. The approach to the evaluation and the methodology followed depended on a number of factors, which include:

- the length of time earmarked for the project
- the type of product being evaluated
- the particular issues being investigated – (interface design, navigational structure, scope and content)
- the goals and purpose of the product (educational, promotional, public relations, financial)
- the target audience
- the available budget

4.1.2. Methods

The methods of inquiry that were undertaken were determined by the particular issues being investigated and evaluated, and an analysis based on an in-depth, critical review of the available literature pertaining to previous research in this area.

The choice of data collection methods depended not only on the questions that needed to be answered but also on logistical factors. These include the time frame available to carry out the evaluation, the availability of expertise, access to equipment and end users, and cost effectiveness.

4.1.3. Scope of Evaluation

As outlined above, the extent to which the interface design, interactivity and navigational structure, scope, content and overall functionality of EduKit2000 support the goals and purpose of the CD as intended by the designers and the their sponsors (these goals are outlined in section 3.1) was evaluated.

It must be noted that due to the enormous size of this product it was not feasible to examine and evaluate every individual page. For this reason, a few of the key areas have been identified, which are considered to be representative of the total CD and the evaluation has been limited to these sections only. The results of an analysis of the selected areas was considered representative of the standing of the CD in general. Since the purpose of the evaluation was primarily to test the methodology, rather than the product, this will not affect the final outcome of this research.

The areas that have been chosen for evaluation, as described above are:

- Navigational structure
- Interface design
- Quality of content matter
- Scope
- Overall functionality

4.2. Evaluation Strategy

The following is a description of the strategy that was adopted to implement the use of each one of the evaluation tools selected for our case study.

4.2.1. Questionnaire

The purpose of using a questionnaire was to evaluate interface design, quality of content and scope, as described above. For this reason, it was designed to return data primarily related to these areas of inquiry, using a total of 17 questions. Personal details of the users, computer skills and hardware assets were also considered important data to collect, to establish a framework for analysing the returned material. Legibility and style were the main features of interface design to be assessed and questions were designed to obtain that data. Quality, particularly accuracy, of the material presented was assessed by referring respondents to specific sections for feedback.

Pilot study

Before the final questionnaire could be printed and included with the packaging, it was necessary to test the final draft by conducting a pilot study. This was done to avoid the expense of having to reprint and re-mail a large number of copies should errors be detected, or to attempt to change questions later. The pilot study was conducted by sending the questionnaire draft to 10 members of the stakeholders' committee, which had collected, collated and provided the content material that the university wished to include in the CD. These were all Edith Cowan University staff members, and very proficient with computer use. The object of the test was exclusively to evaluate content quality and format of the proposed questionnaire. The original draft was designed with the specific objective of obtaining the most useful possible data back to the developers in order to improve the product, and specifically addressed areas relating to interface design, content quality and scope.

As a result of the Pilot study, a number of changes were made to the original draft of the questionnaire before it was finally printed, packaged and distributed. These included some changes of wording, the omission of a “fax-back” option, more personal data from users, and changes in the layout and presentation of the questionnaire questions.

The final changes were first approved by the university, and 14,000 copies of the questionnaire were printed.

Implementation

The *tear-away* questionnaire form was deliberately made as an awkward appendage to the CD packaging, to encourage people to *tear it away*, and hopefully to post it back. Return postage was prepaid by the university in order to maximise the number of returned questionnaires. The final version of the questionnaire is reproduced in Appendix 1 (*Tools and Templates, Questionnaire.)*

The questionnaire was part of the packaging of the CD, and was sent to all commencing and external students by post at the beginning of Semester 2/2000. Extra copies were made available to all other students of Edith Cowan University at the *megalab* on each campus.

Processing of questionnaire data

56 questionnaire responses were returned in time for this study, representing a 0.4% response rate. The data from the returned questionnaires was qualitatively and quantitatively recorded. This involved the systematic analysis of each questionnaire, recording all pertinent details.

The qualitative data, i.e. the comments provided, were all summarised into a master list for each question. From this record it was possible to produce a list of both positive and negative critiques of the product.

The *quantitative* data, which included the personal details of the respondents, i.e. their age, gender, computer proficiency and mode of study, plus the *Likert Scale* responses to the questions, were statistically computed to provide a view of the spread of results that were achieved throughout all of the evaluations.

4.2.2. Expert Review

The expert review was primarily aimed at evaluating interactivity/navigational structure, content quality and scope of the information provided on the system. Four highly qualified, professional multimedia developers who are also faculty staff of the School of Communications and Multimedia were recruited as the experts for conducting the heuristic expert review. Because they themselves were potential end users of the product and well acquainted with the content material, their assessment more accurately pinpointed errors and omissions.

Guidelines

To ensure consistent data was produced from each of the experts that where conducting the *heuristic* expert evaluation, guidelines where created. The guidelines where designed so that all of the experts evaluated all of the targeted aspects of the CD and that no aspect of the product was overlooked by any evaluator. It is standard practice in conducting heuristic evaluations to utilise a set of guidelines aimed at standardising procedures and areas of investigation.

To assist in focusing the scope of the heuristic evaluation, the guidelines created addressed two main areas, the design of the *navigational structure* and the quality of the *content,* since these were the areas identified by this research as being best evaluated by an expert review. The navigational issues addressed were:

- The *cognitive load* (mental stress imposed) the site created for the user.
- Whether the site had effective mapping and the user was always aware of where they where in the site.
- Whether the *help* system and its structure was useful and effective.
- Whether the speed at which a user could navigate was appropriate.
- Whether the organisation of the content was logical and appropriate.
- Whether the site had effective and useful cross-referencing.
- Whether manoeuvrability throughout the site was consistent.

The accuracy and integrity of the content (as well as an assessment of scope) was evaluated by the experts using a geographical site-map of the CD, listing and locating all the content material that required testing. Both the guidelines developed for the use of the experts and the site map of the CD's content matter are shown below and are included in the *Tools and Templates* appendix.

Guidelines were also created for the *interface* evaluation. These guidelines included:

- the general 'look and feel' of the interface.
- The cognitive load that the interface would create and if this would be acceptable for an inexperienced user.
- Whether the metaphors were appropriate to a wide range of users.
- Whether the interface was consistent throughout the web site.
- Whether the text was legible and
- If the use of media was appropriate.

These guidelines are reproduced in Appendix 1 (*Tools and Templates, Guidelines For Expert Review Of Edukit2000.)*

Experts were also provided with a copy of the geographical location of the areas requiring testing for accuracy of content material. This was identical to that provided to the end users for testing, and is also presented in Appendix *1 (Tools and Templates).*

Processing of Expert Review data

The heuristic expert review consisted of our four in-house experts, and a statistical presentation of the data obtained was not considered useful. Data returned was processed and compared qualitatively, then inserted in a grid to highlight areas of concurrence between the four reports received.

Summary of Expert Review process

This process rendered a large amount of valuable data, and was conducted relatively fast. Because experts were participating on a volunteer basis, the timeline was left flexible resulting in some delays in returning all the data. However, most experts were able to submit reports within a week of receiving guidelines. No other problems were encountered in implementing the heuristic expert review. The amount of time each expert actually devoted to the review process was left discretionary for each expert, and was estimated as ranging between two and six hours.

4.2.3. User Testing

User testing aimed primarily at returning data about interface design and navigational structure. Graphic design in the packaging of the CD was also included in the areas for evaluation by users. Because the product targeted mainly 1st year students, it was decided to conduct user testing on a sample of that population. However it was important to use students who had some skills and experience in evaluating multimedia, so a whole class of 1st year Interactive Multimedia students were requested to evaluate the CD as part of their unit training in Software Evaluation. By establishing a classroom environment for the testing, results were optimised by ensuring functioning hardware and software, technical assistance and tutor guidance. 10 students completed the evaluation, using a tool broadly modelled on Reeves' (1993) "Dimensions". This is a user-rating tool using a Likert rating mechanism. It was accompanied by a set of guidelines similar to the ones provided for the experts (Appendix 1, *Tools and Templates.)*

Tool and Guidelines

The guidelines chosen here were broad, general indications used to insure that all important aspects of the CD were critiqued. Reviewers were told to go beyond what a typical checklist would require. Because this was a *qualitative* evaluation, they were asked to write as much as necessary in response to each question. A checklist was provided to map all the *geographical* areas of the CD that needed to be visited and evaluated.

The guidelines and rating instrument used for user testing are presented in Appendix 1 (*Tools and Templates, Guidelines and Rating Instrument for User Testing.)*

A time limit was set on the evaluation (the scheduled time length of the class), and users were permitted to collaborate and to seek assistance from the tutor, both in understanding the concepts and performing the evaluation.

Processing of User Testing data

The data returned from the User Testing rating tool was treated and processed in exactly the same way as that returned from the Expert Review. Graphical representations of the Likert scale responses returned from the rating tool used were created, and written input from users was recorded and analysed.

User Testing Process

Because of the voluntary basis of user participation, only a relatively small number of users could participate in testing. The session was scheduled with their course coordinator to coincide

with one seminar session (2 hours), however since these are not compulsory, some students may have hurriedly completed the rating tool with little observation of the software. Also, the number of participants was limited to those who turned up to class on that day. No other problems were encountered with user testing.

4.2.4. Observation of Novices

Audit trail observation of novices was conducted using Lotus ScreenCam software, which records all user activity (mouse clicks, cursor moves, keyboard use and even verbal comments). The addition of Lotus ScreenCam to the evaluation has been a very useful part of user testing. As the navigation methods that respondents use are an important part of this usability evaluation, this needed to be observed and investigated. Human observation of users would consume vast amounts of time, money and would require the training of observers, as well as influencing responses by the physical presence of the evaluator. Lotus ScreenCam was therefore employed to do this (with guaranteed electronic accuracy), as well as recording any verbal comments users made into the microphone. To ensure that all data gathered by *Lotus ScreenCam* was consistent, a document was created to advise evaluators on how to conduct the session. It was established that the candidate would be advised that their movements on screen would be recorded for further evaluation and that an audio recording of their verbal comments would also be conducted. It was decided to advise the respondents of what was happening so that they were not apprehensive or nervous.

Upon completion of the Observations, all of the comments that users had made during the *Lotus ScreenCam* recording where entered into the *Session Coordination Observation Forms*.

As Lotus ScreenCam also creates a video recording of a user's movements, this information was analysed using a form that was created for this task[1]. This Lotus ScreenCam analysis tool gathers data about how effective the navigation of the web site is. It allows evaluators to compare information such as:

- The number of mouse clicks to complete a task.
- The methods that where used to complete a task.
- The way in which the user navigated
- Where any difficulties occurred and
- the time it took to complete all the tasks.

The implementation of the audit trail observation was conducted in several stages, each of which is described below:

Creation of list of tasks for novices to perform.

The list of tasks was carefully selected to enable testing of a wide range of features, as well as covering a geographically large area of the CD. Users were expected to penetrate several layers of navigational structure as well as to discover the optimum method of navigation and search. There was no time limit set for completion, to impose minimum stress on the participants.

The task list form presented to participants is shown in Appendix 1 (*Tools and Templates, Novice User Tasks.)*

Creation of release forms

All participants were asked to sign a release form allowing the copyright-free use of all material recovered from the session. The reasons for this were explained to each participant. This form constitutes one of the documents in the Tools and Templates appendix.

Preparation of guidelines

Guidelines were set for the evaluators (in this case only myself), in order to standardise the returned data. These are reproduced in the *Tools and Templates* appendix, and are summarised below. It was important to specify how much help, if any, participants would receive from the evaluators. It was established that help would only be offered in the case of participants becoming unable to proceed further with the tasks.

The general conditions of the testing environment (room, atmosphere, desktop condition, accessories provided, etc.) were established, as well as the procedures and human interaction between the novice and the session coordinator. Instructions were given to evaluators to verbally explain the process before requiring novices to sign release forms. The guidelines also contained indications on how much assistance could be given to users, and at which point, as well as how to make a record of the session. Finally, evaluators were informed how to debrief participants after the test and how to end the session.

Recruiting novices

10 computer novices were all recruited, all Edith Cowan University students with little or no computer skills. No age, gender, faculty or other criteria were applied to the selection of participants.

Implementation of Audit Observations

1. Installation of software on work stations

An educational version of Lotus ScreenCam was installed on the system intended for the observation. All other unnecessary icons were removed from the desktop, the computer system's resources were optimized and the Lotus ScreenCam icon and controls clearly placed and identified.

2. Preparation of environment

The observation environment was prepared in accordance with the guidelines shown in the Tools and Templates appendix. A clean, cool, ventilated room was provided, as well as writing materials, soft background music and a hot drink.

3. Signing release forms

Participants were verbally explained the implications and significance of the observation before being asked to sign the release forms.

4. Performing the observation

A form was prepared to assist evaluators during the Audit trail observation session. This tool is included in Appendix 1 (*Tools and Templates, Questionnaire.)*

5. Debriefing the user

Finally, participants were debriefed after they finished (or abandoned) the session. This was a brief (five minute) friendly chat, aimed at identifying any problems with the conduct of the observation session, rather than about the CD evaluated. Notes were taken by the evaluator, and appended to the *Session Coordinator Observation Form*.

6. Analysing returned data

The analysis of the Lotus ScreenCam results involved viewing all the Lotus ScreenCam movie files and assessing what the users actually did when they were navigating the site. Statistics were drawn from these observations, which include whether or not the user used the buttons on the top toolbar, how many mouse “clicks” it took the user to find the information they were after, and how long it took the user to complete the tasks overall.

Verbal comments made by the user were automatically recorded by Lotus ScreenCam. These were later transcribed into the observation forms. Finally, all this data was summarised to provide further recommendations for improving the product. The form used to record this data is included in the Tools and Templates appendix, and consists of an analysis of

1. the number of clicks to find the results
2. the methods were used to locate information (home, toolbar, help facility, search facility, FAQ, automatic paging, back button, scrolling)
3. the time it took to complete the tasks
4. the sections the users got noticeably confused in, and why
5. did the user go home between tasks or did they use the toolbar to move between sections?
6. did the user need to use the zoom option?
7. the users general impressions/comments about the task and their interaction with the CD
8. Spontaneous comments

Because observation of novices was aiming particularly at evaluating navigational structure, interactivity and scope, the information listed above aimed at identifying problems in these areas.

Summary of Novice Observation Process

Novice observation was conducted over a large period of time because of the difficulties encountered in adjusting to individuals' schedules, as well as the problems related to installing Lotus ScreenCam software on more than one machine. The highest degree of flexibility was finally achieved by installing the software on a laptop computer, which could then be transported to the location most convenient to each novice.

No time restrictions were applied, and novice users took between 15 and 35 minutes to complete the ten tasks.

Conducting the tests was laborious and time consuming, yet yielded a large volume of valuable data. Work on novice observation spanned a period of six weeks.

Chapter 5. Results and Discussion

This chapter describes in detail the results of applying the methodology to the case study. The processed data is presented in graphic and textual forms, with a description of the implications of each group of results, as well as a discussion of the processes involved in analysing the data. Sample user comments are provided for illustration, but the totality of returned user comments is not pertinent to this research and is not presented here.

Evaluating a multimedia product is usually done in one of two ways: qualitative or quantitative. Both methods are effective but if used separately may not gather enough information to adequately evaluate the product. When combined, a more meaningful interpretation of the returned data can be made.

Quantification of results

Those results returned by the questionnaire that are provided as simple yes/no results were compiled and appropriate statistics obtained and provided as a clear explanation of the findings. Statistical data was also produced describing, gender, age and bio-data distribution. Likert responses were amalgamated, and values such as mean, mode, and standard distribution were derived.

Qualification

The *heuristic* expert evaluation results and those from the user testing were transcribed and compiled then compared to form a list of recommendations for the product's owners (the University) to review and act upon as they see fit.

Evaluation

The results from the heuristic expert evaluation were compiled and compared to determine if there were common themes appearing from the four experts' evaluation of the site. The results collected from the questionnaire were compiled mathematically and analysed in comparison with researched results where appropriate. Data returned from the user evaluation sessions and the novice user audit trail observations where also processed to produce further information, and an overall evaluation was synthesised using the data returned from the four processes. When compiled, the results provided a clear indication as to what recommendations are appropriate for the EduKit2000 CD.

5.1. Presentation of findings/results

This section discusses the information returned on the EduKit2000 system. The findings and results of the evaluation are presented in a graphical format consisting of a series of graphs (pie, bar, line, etc) accompanied by a set of recommendations for changes to the CD, which are substantiated by the findings of the evaluation.

5.1.1. Questionnaire Responses

56 responses were received to the questionnaire. The results from the questionnaire have been broken down and graphical representations of the results can be seen following. Figure 2. shows the distribution of the study modes of questionnaire respondents.

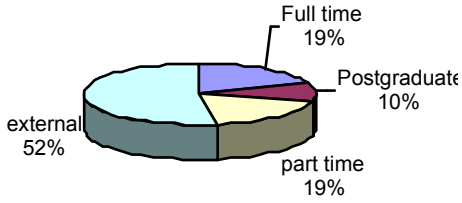

Figure 2. Study Mode of Questionnaire Respondents

Because the CD was sent exclusively to external and commencing students (the representative sample population that it was intended to test), a large proportion of the respondents were external students. Only 10% of respondents were postgraduates, which was to be expected since very few of these would need the resources presented in EduKit2000.

Figure 3. shows the age distribution of questionnaire respondents.

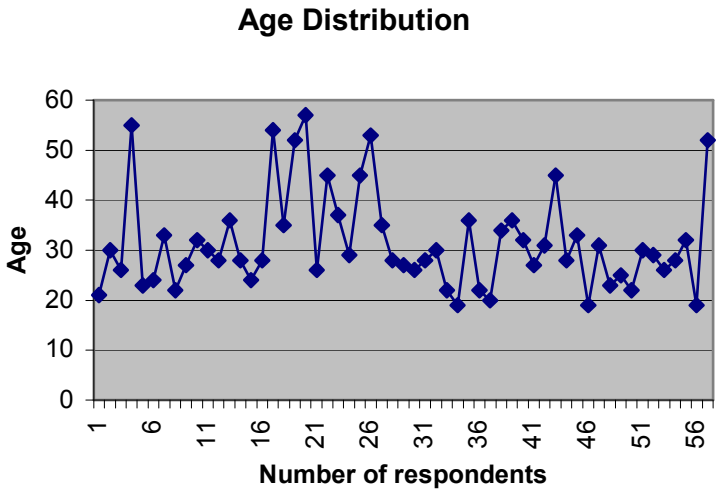

Figure 3. Age Distribution

As anticipated, the age distribution reflected the age of commencing students through the entire university population, i.e. mostly between 20 and 30 years.

Figure 4. shows the gender distribution of questionnaire respondents.

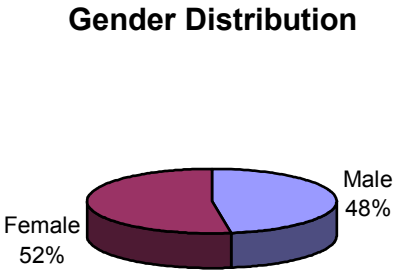

Figure 4. Gender Distribution

The gender distribution closely resembles that of the entire student population of Edith Cowan University, i.e. a slightly higher number of females.

Figure 5. shows the computer skills of questionnaire respondents.

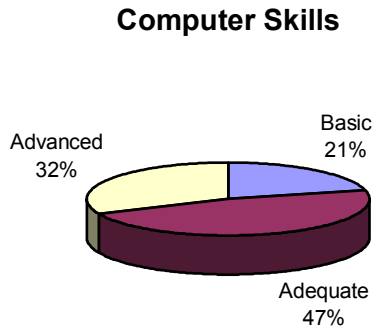

Figure 5. Computer Skills

Most respondents claimed adequate computer skills. This is consistent with the profile of the general student population.

Figure 6. shows the hardware and software assets of questionnaire respondents.

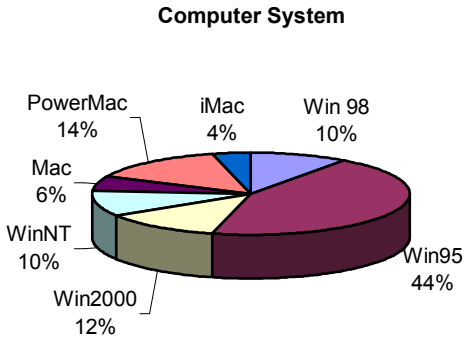

Figure 6. Computer system

As anticipated, the hardware and software assets of respondents spanned a wide spectrum. This is important to the evaluation as data returned from testing on a wide range of systems provides the best information to evaluators.

Figures 7 to 23 show the distribution of responses to the EduKit2000 related questions in the questionnaire.

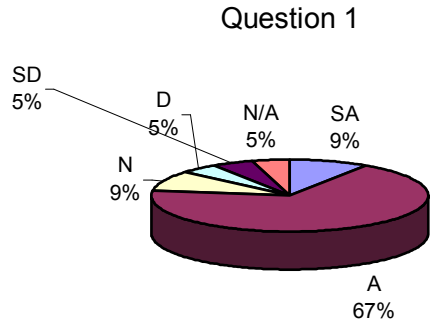

Figure 7. EduKit2000 is a useful service provided by ECU

A summary of user comments indicates that although much of the content is available online it is convenient to have it in one compact package. An analysis of results suggests that although an overwhelming majority of respondents agreed with the question (67% + 9%), with only 10% in total disagreeing, 5% did not answer this question and another 5% thought it inapplicable. Most users considered the product useful.

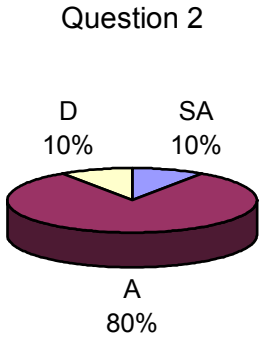

Figure 8. It is well presented and attractive

The questionnaire was especially used to test user attitudes to interface design. Clearly the data returned from this question unequivocally demonstrates that the end users liked the interface. However some problems were noted: 25% of respondents felt that the icons on the interface did not represent the services they stood for. The summary of user comments reveals that some icons were more representative than others. This indicates that some icons do not really work very well if their function is to inform users.

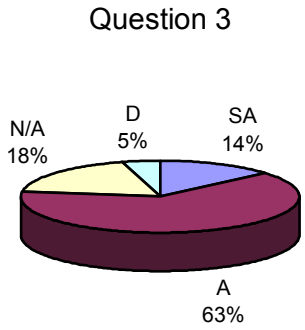

Figure 9. The text was easy to read

77% of respondents found the text clear and legible, and only one respondent thought the text was "*perhaps a little small*". All others though it clear and easy to read. This would seem to suggest, that the current text size is adequate. A number of respondents were older people with eyesight problems and they still found the screens legible (with their glasses on).

Legibility is a prime concern for developers of informational/educational systems. With only 5% disagreeing with this question, the product can be seen as successful from this point of view, perhaps because of the "Magnify" feature built into it's interface.

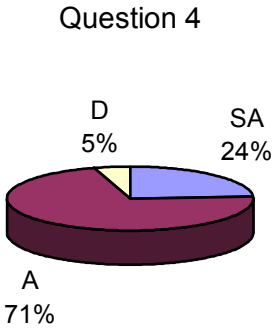

Figure 10. You were able to navigate easily through EduKit2000

With a total of 95% of respondents satisfied with the navigation mechanisms (usually one of the biggest problems with systems of this complexity), these can be seen as effective and easy to use. To summarise user comments: Most users found it 'easy' to see where they were and where they wanted to go. Many did not have a clear idea of their location but they still managed to navigate. Three users suggested more obvious use of headings or icons to let you know what sections you were in. The percentage of users who felt orientated is very high. Users should be

able to recognise where they are, by looking at the current page. They should not need to recall their path from the homepage. Clearly navigation was not a problem to users.

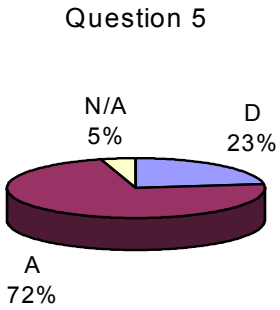

Figure 11. The "FIND" function was useful

Because the "Find" function could only "*find*" a finite number of items, users were either very happy when they found what they wanted, or very unhappy when thy didn't. 72% of users agree that the find facility was useful, 5% had no opinion or did not use it, and 23% found the search facility to be of little use.

A summary of user comments suggests that some users did not know that there was a search facility available. Many of those who used it found it did not recognise the word they had typed – nor did it provide further help or suggestions. Some felt that not enough information was provided when a search was successful – it only provided the link to a page.

It is difficult to know if the search facility is adequate at this point. The evaluation indicated that it was reliable but limited– when it didn't work it was useless, and when it did work it was very good. A greater effort needs to be made to maintain this important facility. The fact that it was not used much indicates that users are able to locate their information using the menu system.

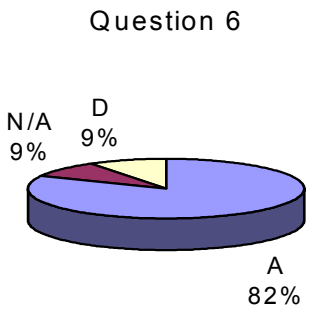

Figure 12. The collection of software was useful

Users commented that some games did not install properly, and there was disappointment at the quality of some games. Most comments indicated that the software was considered useful. An analysis of the responses indicates that a very large majority of respondents thought the software provided was useful. Those who responded N/A probably represent those who had not yet needed to download any of the software. This coincides with the results of the user testing, which rated content higher than any other feature.

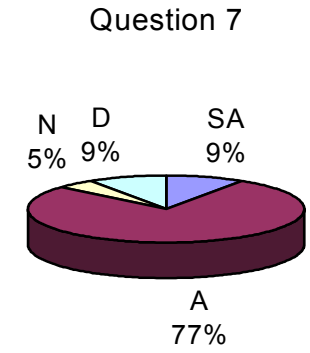

Figure 13. The course information was useful

Users commented that more information on specific units was needed, including previous class averages and exams, assessment details, etc. Here again, only a very small minority found the course information not to be useful.

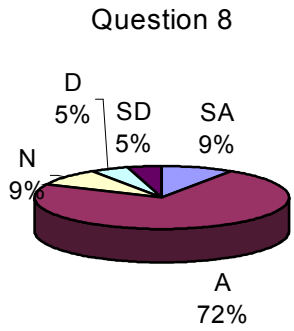

Figure 14. The services information was useful

Most respondents considered the services section useful, with only a total of 10% disagreeing.

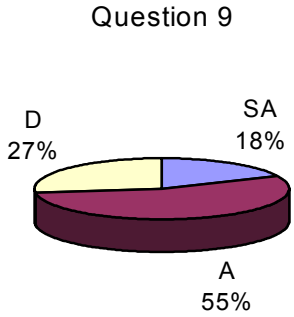

Figure 15. The lifestyle section was useful

User comments suggested that more descriptions of clubs and societies, services such as the cafeteria, bar, etc. are required. A substantial proportion did not find this section useful. This data coincides with other data returned about the lifestyle section. Because of the small volume of content available on the CD that relates to student lifestyle, many respondents did not find enough information to provide them with a complete description of lifestyle issues.

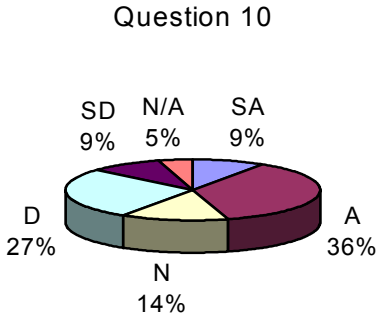

Figure 16. The modem pool/remote access instructions were useful

User comments suggested that animated demos would be useful. Generally, responses to this question provided very useful information, especially since connecting to the modem pool was seen by the developers as one of the prime objectives of the product, as described above. At least 19% of respondents did not address this question, either because they were already connected or had not yet tried to connect. Of those who did, almost half encountered problems with the connection instructions provided by the CD.

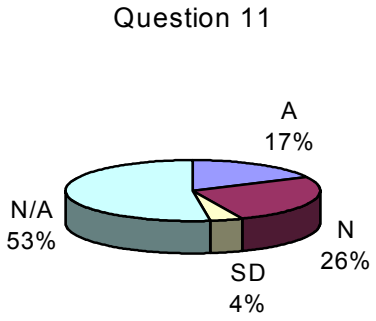

Figure 17. The library information was useful

There were no user comments in response to this question. An analysis of returned data indicates that because of the very large number of N/A responses most users did not attempt to use the library information. This could be because the questionnaire was completed too early in the academic semester, or because the very large proportion of external students have no interest in library information.

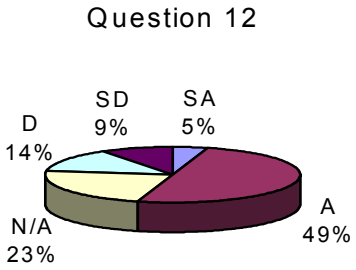

Figure 18. The "HELP" function was useful

Users commented that the page is too cluttered; cursor error on rollover and content is limited.

As with the "search" function, this feature of the product was only programmed to provide help with a finite number of potential problems. Not all conceivable help is available, which explains

why a large 23% disagree with the question. A total of 23% of respondents did not however need to use it.

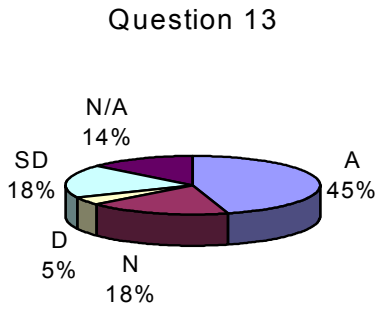

Figure 19. Software from the collection installed correctly

A huge total of 32% had not yet tried installing software from the collection. However, of those who did, more than two thirds found it installed correctly.

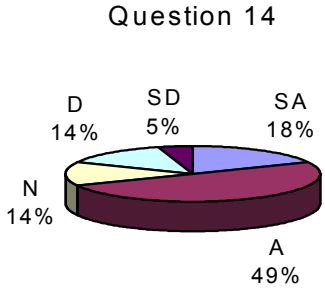

Figure 20. The videos played correctly

Some users commented that the video didn't work, was boring, jerky, slow-loading. The wide range of responses received here was anticipated. They reflect the varying levels of sophistication of the hardware assets of the student population. Those possessing highly advanced multimedia technology clearly had fewer problems viewing videos. Those students with more basic computer systems will be represented by the 19% who experienced difficulties. Better video compression is required.

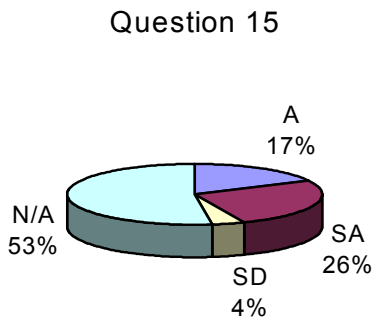

Figure 21. You successfully connected to the ECU Modem Pool

43% of respondents successfully connected to the modem pool using these instructions. Again, the large proportion of N/A responses reflects those users who did not attempt to use EduKit2000 to connect to the modem pool. This could be either because they were already connected or had not yet attempted to connect at the time of completing the questionnaire. However, of those who did attempt 81% were successful.

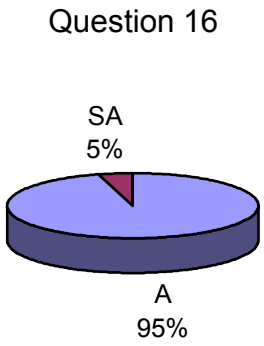

Figure 22. EduKit should continue to be designed and developed by students

Users commented that only students could understand the needs and interests of other students. This was the only question in the questionnaire that received 100% favourable responses. Clearly there is overwhelming student support for the concept of having the product developed by students.

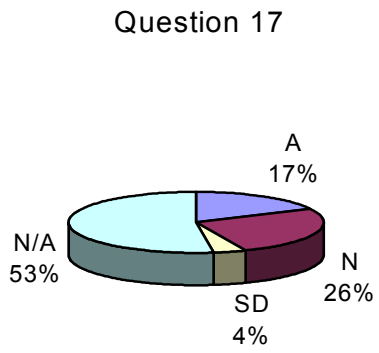

Figure 23. You printed off many documents

An analysis shows that the large number of N/A responses indicates that more than three quarters of the respondents had not yet tried to print any documents from the CD. This is because of the timeframe at which the questionnaire was completed. Assignment cover sheets and other forms would be in demand at a later time in the academic semester.

Summary of Questionnaire findings

As can be seen from the graphic data presented above, most of the feedback received from the questionnaire was positive. Although it is understandable that students will react positively to the free gift of an informational/educational system, and that responses are likely to be positively skewed, the overwhelming positive experience described by students through the questionnaire indicates that the product has been reasonably successful in achieving its objective of appealing to the student population.

Areas of concern expressed by users have been identified, but generally reflect a very good product with few systemic bugs or instructional failures. Problems need to be addressed with regards to icons, video compression, content integrity and scope. It has been found that many of these coincided with the results returned from the other processes used for the evaluation.

The questionnaire was intended to provide data on interface design, content quality and scope. It performed adequately in that function, providing valuable (but little) information about the software collection and the use of some icons and buttons.

Interface design

The questionnaire data provided information on several design problems, although it generally reflected satisfaction and approval of layout and graphic design. In particular, some icons were considered unintuitive by respondents (print, volume, control, etc.).

An important concern in interface design is the legibility of presented material, and the questionnaire returns suggest there are few legibility problems with the product. Questionnaire data has therefore proved useful in providing information about specific interface design issues

Content quality

Most of the questions were designed to evaluate the quality of the content. Returned data was able to pinpoint many areas where the content quality was lacking, such as information on clubs and societies, library access and modem pool instructions. Conversely, were content quality was adequate (e.g. the software collection), the data provided that information to evaluators.

Scope

The scope of the CD and of many of its component features (search and help facilities, software collection, etc.) were adequately evaluated in the returned data, many respondents concurring in their written comments. The questions relating to scope (e.g. Which sections were useful? Was the product useful? Etc.) successfully highlighted scope issues in the product.

The biggest problem encountered with the questionnaire process was the very low level of response (0.4%). However the real value of questionnaire data is when it is applied to a large proportion of the user population, and this was not the case.

A full transcription of all comments has been done and recorded in a database file, for the use of future developers.

5.1.2. Expert Review

The expert review was intended to evaluate interface design, overall functionality and quality of content. It has however also provided valuable data about scope, navigation and interactivity. Reports from the four experts selected for the expert evaluation were analysed and interpreted. Common trends were noted as well as multiple references to particular issues, bugs, problems or advantages. Because the guidelines provided to the experts identified the areas of inquiry to be addressed, this made it easy to compare the experts' reports on the basis of these areas. A tool was also provided to the experts to conduct the testing. This Testing Sheet or Review Log is reproduced in the *Tools and Templates* appendix. It allows testers to identify and describe problems, then recommend changes.

Each expert's report was processed to provide data on the areas requiring evaluation, and these are presented below:

Expert 1

Interactivity/Navigational Structure

1. Mapping and mental model of the product are problematical.
2. Cursor rollovers are required
3. Good maneuverability and navigational ease
4. High cognitive load
5. Cross-referencing may be confusing

Interface Design

1. Good engagement quality
2. Low cognitive load
3. Many Windows metaphors used which could confuse Macintosh users
4. Good legibility, especially with zooming
5. Background colours too strong
6. Interface should resize to monitor preferences
7. Volume control icon not intuitive

Quality of Content

1. "Fun" software may be unnecessary
2. Only three Clubs and Societies presented

Overall Functionality

1. Error in playing video on some platforms

Expert 2

Interactivity/Navigational Structure

1. Remove inactive "back" and "forward" arrows

Interface Design

1. Window defaults to middle of screen
2. Good main Menu page
3. Too much scrolling needed
4. Paging metaphor not clear
5. More chunking of text
6. Not all icons intuitive or consistent (print, volume, etc.)
7. Rollovers and cursor changes on all active links

Quality of Content

1. Software list should be presented immediately as a list

Scope

1. Software area should be more intuitively accessible

Overall Functionality

1. Slow startup, may require a "wait message or progress bar
2. Slow navigation
3. Slow scrolling

Expert 3

Interactivity/Navigational Structure

1. Avoid menu animation for each change of section
2. Pop-ups do not always disappear

Interface Design

1. Engaging interface, good design
2. Close button on right of screen could confuse Macintosh users
3. Volume control icon not intuitive, also zoom icon
4. Backgrounds a little to strong
5. Cursor changes not all accurate

Quality of Content

1. Good use of multicultural images
2. Omissions in Credits page
3. Games are of low quality

Scope

1. Indicate length of time video runs
2. Paging not intuitive

Overall Functionality

1. Video: logo pixelates on zoom-out
2. Find option needs to prompt user to use "enter"
3. Slow loading
4. Slow scrolling

Expert 4

Interactivity/Navigational Structure

1. Cross-referencing may be confusing
2. Enhancing the nav path and maybe making it clickable could probably be used whilst page indicators could be removed
3. Poor mapping
4. Quit screen is confusing. There should be a message saying “Are you sure you want to quit?”
5. The text menus used are often indistinguishable from the main body text

6. Back function (next to find) means back to a previous text content section but does not include the navigation from the home page … back should eventually get you to the main menu.

Interface Design

1. The general look and feel of the CD is of a high standard
2. Poor multimedia mix
3. Choice of some text colours and backgrounds negatively affect readability
4. Cursor changes on rollover
5. Icons seem oddly selected and are not really distinctive to their function. For example: links looks like a phone, printing and audio icons are oversimplified
6. Text should be anti-aliased
7. Packaging graphic design incompatible with CD graphic design
8. Skip video button needs to be a button and needs to be anti-aliased
9. Minimise and quit buttons need to be labelled

Quality of Content

1. More synthesising, reducing and reformatting of text for readability, the shear volume of text contained within the system
2. Poor image quality (low level JPEGs)
3. Violation of both the Quicktime and the Macromedia licensing agreements here, a screen is needed which is displayed for at least four seconds on quit showing respective logos.

Scope

1. More multimedia content should be used for instructional purposes, e.g. demos, animations, interactions
2. Probably too much content

Overall Functionality

1. Slow scrolling
2. Video poorly compressed and slow loading
3. As a CD product the interactions provided could have been enhanced and made more visually and instructionally stimulating
4. Menu options work on mouse down rather than mouse up
5. FAQ: a lot of questions presented in a form to lose the novice… maybe questions at the top, hyperlinked to the answer below, this would avoid painful scrolling
6. Campus maps don't work… clicking on buildings does not tell you what the building is.

Summary of Expert Review

A summary of the experts' reports has been represented in grid format and is presented below:

Table 5.1 Summary of expert results

Area	Expert 1	Expert 2	Expert 3	Expert 4
Interactivity/Navigational Structure	Mapping Cursor rollovers Good maneuverability and navigational ease High cognitive load Cross-referencing confusing	Remove inactive "back" and "forward" arrows	Avoid menu animation for each change of section Pop-ups do not always disappear	Cross-referencing confusing Page indicators could be removed Poor mapping Quit screen is confusing. Text menus indistinguishable from text Back function does not go home
Interface Design	Good engagement quality Low cognitive load Many Windows metaphors Good legibility, with zooming Background colours too strong Interface should resize Volume icon not intuitive	Window defaults to middle of screen Good main Menu Too much scrolling Paging metaphor not clear More chunking of text Icons not intuitive or consistent (print, volume, etc.) Rollovers and cursor changes on all active links	Engaging interface Good design Close button on right of screen could confuse Macintosh users Volume control, zoom icon icons not intuitive Backgrounds too strong Cursor changes not all accurate	Good look and feel Poor multimedia mix Poor choice of some text colours and backgrounds Cursor changes on rollover Text should be anti-aliased Packaging graphic design incompatible with CD Skip video button Minimise and quit buttons labelled
Quality of Content	"Fun" software unnecessary Only 3 Clubs and Societies	Software contents should be a list	Good use of multicultural images Omissions in Credits Low quality games	More synthesising, reducing and reformatting of text Poor image quality (low level JPEGs) Violation of QuickTime and Macromedia licensing agreements
Scope			Indicate length of time video runs Paging not intuitive	More multimedia for instructional purposes, e.g. demos, animations, interactions Too much content
Overall Functionality	Error in playing video	Slow startup Requires a "wait message or progress bar Slow navigation Slow scrolling	Video: logo pixelates on zoom-out Slow loading Slow scrolling	Slow scrolling Video poorly compressed and slow loading Menu options should work on mouse up FAQ: reformat Campus maps don't work

The grid above clearly demonstrates that the experts concur on most of the issues they have addressed, with some problem areas being identified by all four experts (such as the Volume Control and Print icons, the scrolling speed, etc.)

To summarise the expert review, the general look and feel of the CD is considered of a high standard. The model used in the development is based upon a text-oriented delivery of information rather than any use of higher-level interactive content, very much in the model of a simple web site delivered in CD.

The interface design and information model selected, while able to deliver a large amount of information to the end user does not always assist comprehension or clarity of the information presented. The end result is much like a book that has been put onto a CD, with few images or supporting media to may make the content understandable or digestible to the wide range of often techno-illiterate users.

From an instructional design perspective, more effort needs to be put into synthesising, reducing and reformatting for readability, because of the shear volume of text contained within the system. Most screens of content are essentially text files with limited formatting and few hyperlinks. Icons are not really distinctive to their function. For example: links, printing and audio icons are oversimplified. These issues increase the cognitive load especially for the novice user (the target audience).

With regard to legibility, selections in the area of text size and colour (especially in relation to background images) make the content difficult to negotiate at times. Examples of these issues are:

- dark blue text on complicated backgrounds,
- small typeface size in menus,
- non-anti-aliased text
- and awkward alignment.

Video has been poorly compressed and plays back in an awkward, jerky fashion. Some images are blocky, demonstrating the same edge effects and blurring that take place when JPEG images are saved over multiple generations. Given the delivery medium is CD not NET, higher graphics and multimedia quality should be developed.

Multimedia could have been used in a more complete way to enhance the content and to explain some of the more complicated elements in a way which is more visual. For example, connecting to the remote access system could have been done with something other than text. The use of cursor changing on rollover and 'hot' text elements would have enhanced the interaction provided.

Navigation is clear and well-organised when navigating between major sections and finding content.
A reassessment is necessary however of the following navigation tools:

- Navigation Path indicator (top left)
- Back button
- Find Option
- Forward and back options for multiple linear pages
- Page x of y indicator
- Hyperlinks

The overall effect is to produce an information system in which the user can quickly get lost. The combination of "page x of y" and the nav path indicator is a little awkward, especially with sections containing 200+ pages. Enhancing the nav path and maybe making it clickable could probably be used whilst page indicators could be removed.

The relatively consistent colour scheme and look and feel all lend themselves to an impression of being highly generic, no real visual cues exist within the content to indicate which section you are in. This is a feature of the book metaphor used throughout the product. Mapping needs to be clearer so that users are not 'lost'.

The content is mostly relevant, appropriate, comprehensive and useful, though its presentation and navigation could be improved.
Programming of the product is generally well executed with all major sections working efficiently according to specifications. The main usability issue is the slow scrolling speed of the text elements, this can make navigating the CD content, which is almost exclusively text based, painful.

As a CD product the interactions provided could have been enhanced and made more visually and instructionally stimulating.

Package design and implementation is beautiful and well constructed. The major problem with the package is that its look and feel is incompatible with that of the CD interface. The visual style, typefaces, logos and graphics on the CD are different from those on the package. The packaging should reflect the visual design of the actual product being packaged.

The expert review was intended to evaluate interface design, overall functionality and quality of content. It has proved to be the most effective process at achieving its objectives, providing critical information in these areas of the software. Major and minor flaws in the system were identified, indicating that the tool was an appropriate choice for evaluating the three areas described above. The expert review has also provided very useful information about the other two areas (interactivity/navigational structure and scope).

5.1.3. Observation of Novices

The Lotus ScreenCam statistics show how the user interacted with the CD; the interactions have been broken down into three areas. The first area shows how many clicks it took the user to complete each task that was set, the second shows how long it took to complete all 10 tasks and the third area shows whether or not the bottom toolbar was being used.

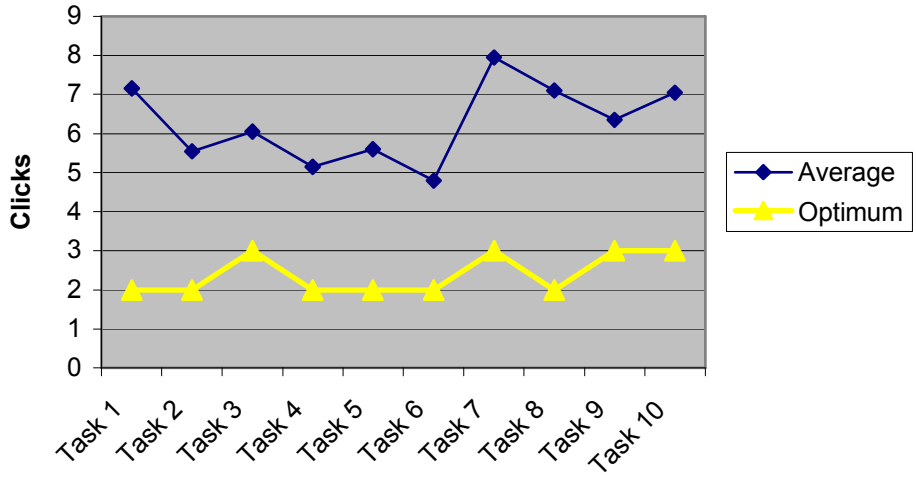

Figure 24. Average Number of Clicks

Most users used too many clicks (relative to the optimum number) to find the information. However this was to be anticipated. Because these were novice computer users, unaware of elementary navigation techniques, a larger number of mouse clicks was expected. Figure 26. shows the average number of clicks taken by the user in order to find the information they were after. The optimum number of clicks for each task is shown. These results show that the perhaps the information can not be found as obviously as it should be (i.e. the average user may have to

look in several areas before finding the area they are after.) To get the optimum number the special facilities (Search, Help, etc.) needed to be used. This also shows that these were not used or not used effectively.

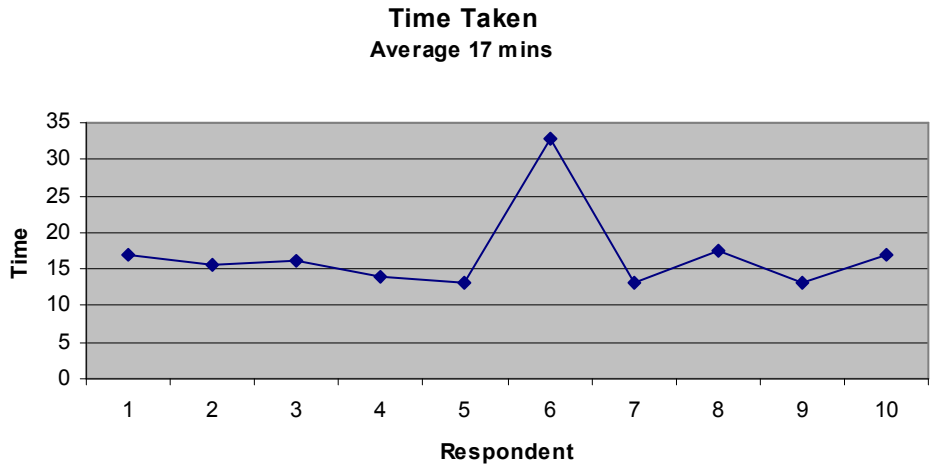

Figure 25. Time taken to complete tasks

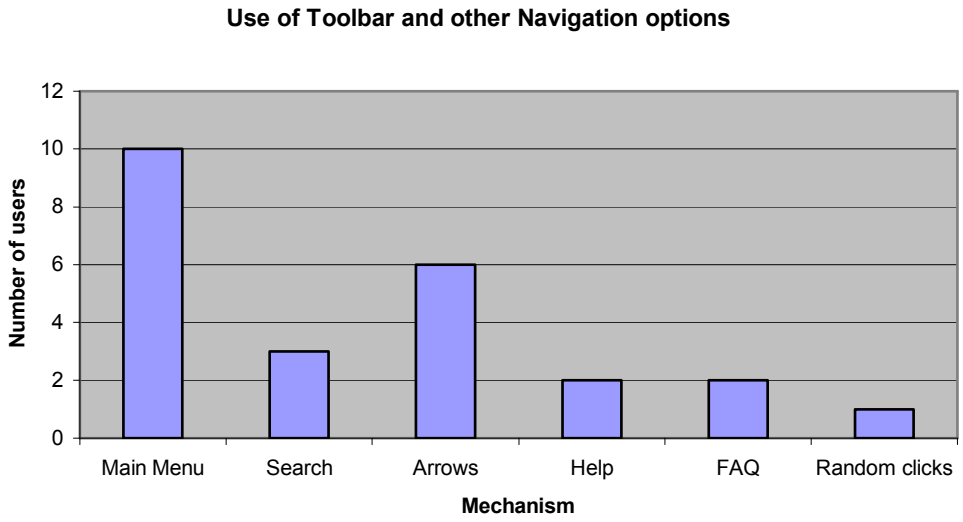

Figure 26. Use of toolbar and other navigational options

Using the search facility was the most efficient method of obtaining answers to the 10 tasks, yet only three participants used it. Most questions contained keywords that would yield results within 2 or 3 clicks using the search facility, which itself was only 1 click away from the Main Menu. However the average number of clicks users needed was more than twice that number, indicating that the search facility, while extremely efficient at yielding results, is not the obvious choice of navigation mechanism for novice users.

The results of the analysis of the data obtained during the Lotus ScreenCam Observation of 10 computer novices revealed the following trends:

a) The toolbar at the bottom of interface:

The feedback indicates that some of the respondents were unaware of this toolbar. Consequently these people did not notice that there was a main menu button and therefore were using the back button to back main menu. If the user was more than one screen away from main menu then it would take more than one click to get back.

b) The Search facility:

Of the few people that tried to use the *search en*gine, some failed to derive any benefit from this facility. Users of this facility tended to be more experienced respondents and therefore had fewer problems than complete novice users adopting typical styles of navigation. The feedback received from this observation reveals that the respondents found the terminology used in the questions was too *ambiguous*.

c) The Help facility:

The users found the help facility too cluttered and confusing. Users also identified an important programming bug in the system here: a cursor change on text-box rollover indicates a link, yet the information boxes are not hyperlinked.

d) Size of the CD:

There was so much information and options available in the CD that some users became easily confused and/or lost.

e) Verbal Comments:

The respondents were encouraged to talk aloud about their experiences during the observation. The following comments were typical examples from the users:

- Too much content
- The pages took too long to download
- Don't know what this button does
- Frustrating
- Confusing
- Many users were surprised at the ease with which they completed the tasks.

The full transcription of all comments has been done and recorded in a database file, for the use of future developers.

5.1.4. User Testing

The data returned from user testing was treated and processed in the same way as that returned from the expert review.

The open question responses returned by the users were transcribed and processed in conjunction with the extra comments, if any, which were volunteered at the end of the session. The data returned here is of particular importance as it represents tests conducted on a representative sample of the user population, 1st year Edith Cowan University students.

Most users rated the product very highly, in all areas of inquiry. The Likert scale rating tool used for the evaluation instructed users to rate the product on a scale of 1 to 5, ranging from Bad to Good, for each area of inquiry. Figure 29 shows the level of user satisfaction with the software in the five areas of inquiry under evaluation.

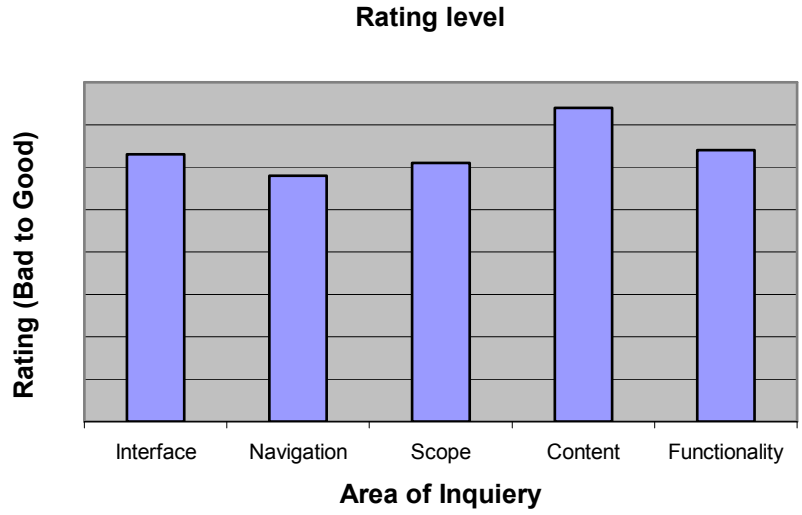

Figure 27. Rating level

Summary of comments:

Interface comments

- volume and print icons difficult to identify
- very good but video/graphics need upgrading
- background colour is not consistent

- aimed at target audience
- CD is pleasing but colour too bright,
- video takes too long to load
- sound can be annoying
- scrolling not good, too slow
- more pictures

Navigation comments

- recommend change cursor on rollovers
- help is not effective, button should be elsewhere
- better and bigger buttons.
- Arrows are a problem

Scope comments

- packaging can be improved

Content comments

- video interviews of students would be useful
- more graphics (photos) to support info. e.g. accommodation.
- instructions work and content accurate
- important points should be highlighted.

Functionality comments

- music slows down users' inputs and responses.

Other Comments

- a good software, but navigation could be less confusing
- too much text on each screen
- should have lower sound, less stressful.
- crashed after viewing the welcome video
- should improve icons.

Summary

An analysis of the comments made by the users shows a remarkable similarity to the conclusions arrived at by the experts. The full transcription of all comments has been done and recorded in a database file, for the use of future developers.

The information obtained from the integration of the processed and analysed data sourced in the four processes comprising the methodology was then constructed into a list of recommendations for improving the product, which are presented in the next chapter.

.

Chapter 6. Assessment of Evaluation Methodology

In order to determine whether the methodology that was developed by this research is an effective tool for the evaluation of informational/educational systems, it was important to compare the attributes of a good system to that of the system developed. This information was indispensable for determining what improvements could be brought to the methodology.

6.1. Attributes of a good methodology

The attributes of a "good" evaluation were derived from an exploration of the evaluation models developed by previous research and discussed in earlier chapters. These are:

- Ease of use

The evaluation system should be easy to conduct. It should be fast yet accurate in the data returned. Any logistical problems in conducting the evaluation would impact negatively on the results obtained, affecting the overall value of the evaluation. Procedures that are too complicated are more likely to incorrectly be applied, thereby affecting the accuracy of results.

- Accuracy

The evaluations resulting from the use of the methodology must produce accurate results. These results should be able to be confirmed by testing outcomes against objectives of the product. All deficient areas in the product, as well as bugs, weaknesses or omissions should be accurately pinpointed by an accurate evaluation methodology.

- Efficiency

It should be efficient and cost-effective. An efficient system has got the highest accuracy vs. cost ratio. No evaluation can be applied to a product if the cost of the evaluation is greater than the overall benefits derived from possible improvements to the product. However it is important that enough data, and of the appropriate kind, is returned from the use of the tool to provide a framework for the effective redevelopment or improvement of the product. The methodology designed should produce the most effective evaluation of the product at the lowest possible cost in time, budget and human resources.

- Consensus of data

The various components, instruments or process comprising the methodology should produce results of a consistent nature. Without delivering duplicate results (which are a waste of resources) they should identify trends or patterns by triangulating observation data. When several components of the methodology identify the same problem areas, this consensus of data allows greater assurance that the results returned are accurate.

- Scope of data gathered

The range of components of the methodology (processes, tools, etc.) should span the full spectrum of system attributes, in order to produce the most comprehensive evaluation of the product possible. Data returned should be both quantitative and qualitative. Quantitative data is important for obtaining statistical information, which can be presented graphically for better analysis. However qualitative data should also be collected, representing views, experiences, specific observations, etc. When it is more effective to collect written or verbal comments, this should be done, and an effective evaluation methodology should have the necessary capabilities to do so built into its design.

- Tailorability

No two informational/educational systems are identical. It follows that no single evaluation program will apply unilaterally to the effective evaluation of all of them. The optimum methodology should posses a degree of flexibility allowing it to be customised to the specific characteristics of different products. The degree of a program's tailorability is a measure of the flexibility that program demonstrates in adapting to the particular requirements of different informational/educational systems.

- Generic applicability

Perhaps the most important feature of all in a good evaluation methodology is its ability to perform effective evaluations of all possible products in the category they are designed for. Because different informational/educational systems will have different educational, promotional and public relations objectives, the methodology developed should encompass mechanisms that can be successfully applied to as wide a range of products as possible.

6.2. Results and analysis of testing methodology

As stated above, the objective of evaluating EduKit2000 was two-fold: to test the *methodology* using the case study, and reciprocally, to test the *case study* using the methodology. The results of evaluating the case study (EduKit2000) were presented above *(Presentation of findings/results)*. It was then necessary to assess the quality of the *evaluation methodology*, and this was done by comparing its features to the attributes of a "good" methodology, described above.

Ease of use

Most of the processes and procedures developed or adapted while constructing this methodology are reasonably easy to apply. However, because it was desired to design the most accurate and comprehensive methodology possible, this resulted in a relatively complex strategy that may not be as easy to apply as is desirable. Because most informational/educational

systems are developed by universities and other educational institutions, it was assumed that an adequate supply of experts would always be readily available to perform the expert review. However, this is not necessarily true, as some institutions might not offer multimedia as a course, yet still wish to distribute an informational/educational system.

Questionnaire:

While easy to implement as a tool, it presented certain difficulties in its design, resulting in insufficiently accurate data. The need for a pilot study before implementation constituted an added degree of difficulty in applying the tool. Another problematic area was the difficulty of obtaining a sufficiently large proportion of responses. This necessitates the use of a mechanism for maximising returns, such as encouragement prizes.

Transcription and analysis of qualitative data returned from the questionnaires is also a lengthy and laborious process, and in this particular case study the value of the information obtained from the questionnaires may not have been important enough to justify the volume of resources allocated to applying this instrument. This was probably due to poor design of the questionnaire.

Expert review:

Implementation of an expert review is a relatively easy process, providing the availability of experts. Little work is required in the preparation of detailed guidelines since experts are assumed to know what is required to be tested. If the experts recruited are themselves end users of the product this makes implementation of the tool even easier, and the application of the methodology to the case study has resulted in few difficulties with the expert review, other than adjusting to the loaded timetables of the in-house experts.

Observation of novices:

Observation of novices presented difficulties in the recruiting and preparing of participants. A large number of tools and templates was required including volunteer release forms, coordinator forms and guidelines, task lists, etc. The implementation of the observation also involved a degree of difficulty particularly with debriefing sessions and the analysis of the returned data. Audit trail observations are not easy to conduct, requiring qualified personnel, expensive software and an adequate supply of novice users. However the value of the information gathered using this instrument is extremely high, and in the case of this test, has delivered a valuable evaluation of some features of the case study.

User testing:

Few difficulties were encountered with user testing. Because a sample of the user population was readily available, the only difficulty involved was the design of the guidelines and the rating tool used for testing.

User testing is relatively easy to conduct, especially in educational institutions, where the user population is physically available every day. However guidelines created for the users must be useful, accurate and specific to obtain best results, and this may represent a degree of difficulty in applying the instrument.

Clearly, an apparent weakness of the methodology developed is the degree of difficulty in its application. The test conducted for this research was not considered difficult, but the specific situations of the Edith Cowan University case study may have provided certain advantages that would not otherwise be available.

Accuracy

Until testing on outcomes is conducted on the case study (EduKit2000), it is impossible to determine how accurate the information obtained from the evaluation is. The product needs to be assessed in terms of the impact of its distribution on the user population, and the extent to which the objectives of the parent institution (the product's distributors) have been achieved. This may be the subject of further research, however with the information obtained from this research, it can be deduced that enough triangulation of observation views has been conducted to maximise the accuracy of results.

Questionnaire:

A high degree of accuracy cannot be guaranteed from the questionnaire, as it is impossible to determine to what extent respondents considered their responses, especially if a prize is offered to winning responses. Unless the questions are designed to elicit the maximum accuracy of responses, the value of the questionnaire may be wasted. Most respondents will often only answer the closed questions, requiring less consideration, yet these are the least useful in providing test information.

Observation of novices:

This process provides very accurate data, especially if video recording of the sessions is conducted. Data returned from audit trail observation of novices has been the most useful in identifying navigational and interactivity problems. If novices can adequately navigate the system then it can be assumed that the more experienced user population will experience even fewer difficulties. Observations can very accurately pinpoint specific navigational problems, and suspected problems can be included in the task list to improve accuracy of results.

User testing:

As with questionnaire returns, data derived from user testing did not produce the most accurate results. Users conducting the volunteer sessions will not always want to commit the required time and effort to produce accurate results. Stricter supervision, and mandatory activities, may be useful to ensure better results.

Efficiency

As was the case with *ease of use*, the specific situation on the ground for testing the methodology was especially convenient, providing extremely efficient results. This may not necessarily apply in all situations, and costs are likely to be considerably higher, affecting overall efficiency of the methodology. Because of the wide array of processes and tools involved, a considerable amount of resources are required.

Questionnaire:

Questionnaire responses will only be received in sufficiently high numbers if the postal return is prepaid, as was the case with EduKit2000. Depending on distribution numbers, this cost is likely to be quite high. A prize may also need to be offered to encourage maximum response, adding to costs and reducing efficiency. While the use of a questionnaire can provide very useful data, this is only true if a large percentage of responses is received. This was not the case in this project, and results obtained from the questionnaire were not as useful as was anticipated. Unless a mechanism is implemented to ensure high returns, the use of questionnaires may not be appropriate in every case. Also the cost in time will be considerable if significant numbers of results are received and processed.

Expert review:

Expert reviews (especially heuristic evaluations which require at least four experts) are extremely expensive to implement, and were only possible for this case study because the four experts volunteered their services *gratis*. This would almost never be the case in normal circumstances, and the expert review component of the evaluation program is likely to be the highest-ranking budget item in generalised applications of the methodology.

Observation of novices:

Expensive software is required (LotusScreen Cam) as well as the services of qualified supervisory personnel, all of which may affect efficiency by increasing costs. The value of the returned data however is high compared to that of other instruments, such as the questionnaire or user testing.

User testing:

While the costs of implementing user testing are very low (compared to a questionnaire or expert review), the overall efficiency is low because of the poor quality of data returned. This was probably due to the design of the guidelines, and the rating tool used.

Consensus of data

Consensus of data was generally quite high, and demonstrates a degree of reliability of the information produced. For example, the methodology identified certain icons as being problematic by a substantial concurrence of observations. Three of the four experts indicated the *volume* and *print* icons were not intuitive. This observation was confirmed by the questionnaire, the observation of novices and the user testing. In other words, every process, tool and instrument used has specifically identified these two icons. While this demonstrates a high degree of consensus of data, it is interesting to ask the question: Why did only three of the four experts identify the problem? The answer is probably that the 4th expert did not examine the icons concerned. In order to maximise consensus of data it is important that the different components of the methodology be accurately *scoped*, and perhaps initial scoping of the evaluation strategy could have been more meticulously conducted.

Questionnaire:

Results obtained from the questionnaire concurred with those obtained from user testing. However no significant concurrence was noted with the results obtained from the expert review or the user tests. While this demonstrated that the different elements of the methodology identified different problem areas, it raised the question of why some processes did not identify all the problems. Clearly the design of some tools was more successful than others.

Expert review:

Expert review results generally concurred with most data returned from the other processes, indicating that the expert review produces the largest amount of concurrent data, even about areas that it was not specifically intended to evaluate.

Observation of novices:

Data returned by this instrument concurred substantially with much of the other data, as was the case with the expert review. This may have been due to the design of the task list set for novices.

User testing:

User testing data concurred with questionnaire results, however this may have been due to the fact many of the questions in the these two processes were identical. If the design of these questions was faulty this could have affected the value of the information derived from both instruments.

Scope of data

This is one criterion where the methodology developed rates quite satisfactorily. Because of the use of automated audit trail recordings, a wide range of data was collected, including written and spoken comments, video recordings of navigation, and user impact assessments from the

debriefing sessions. Both quantitative and qualitative data was collected using the questionnaires, and in depth reports were received from the experts. However, while the scope of the data gathered may be quite large, the quality and usefulness of that data may not be very high. There may have been a tendency during the design phase of the methodology not to distinguish clearly enough between the *quality* and the *quantity* of the data collected.

Questionnaire:

The scope of the data returned from the questionnaire was restricted by the number of questions it is reasonable to expect respondents to answer. A longer questionnaire would have resulted in even fewer returned forms, reducing the reliability of the inferred information. However, the range of the questions was designed to span the full spectrum of usability problems that the questionnaire was intended to evaluate (interface design, content and scope).

Expert review:

Expert review data provided some of the most useful information returned by the evaluation program. The scope of that data depended on the individual experts, since little direction was given to the experts. Because very specific areas of inquiry were targeted by the expert review, the scope of the data returned was limited to that range. However, the scope of data can be customised (broadened or focused) by adjusting the guidelines provided to experts, and by customising the reporting tool they were provided with. Because expert reviews are expensive, it is preferable to restrict the scope of the data gathered by this instrument to those areas best evaluated by the experts. This was not the case with the EduKit2000 case study, resulting in incomplete evaluations, long delays, and some superficial results. The overall quality of the data however, in spite the wide scope required by the guidelines, was generally of a very high standard.

Observation of novices:

Data returned from audit trail observation of novices was limited to navigation and functionality. It is difficult to assess appreciation of aesthetics by recording navigational paths, and impossible to assess accuracy of content material using a computer novice. Information derived from the observations was quite deliberately restricted in scope, in order to provide the most useful possible information on those areas that are best evaluated using this instrument.

User testing:

User testing data was restricted in scope for similar reasons to the observation of novices. The data returned targeted specific areas of inquiry because longer or more complicated user testing sessions may have been counter-productive, causing participants to curtail, abandon or expedite the evaluation at the expense of thoroughness.

Tailorability

Questionnaire:

Of all the processes used in this program, the questionnaire is doubtless the most tailorable, flexible and customisable, making it an extremely important instrument. Questions can be open or closed, can have a rating or grading response scale, can be as few or as many as required, and can be presented in the most appropriate format for the user population. Because of its complete tailorability, the questionnaire can provide a wide scope of data

Expert review:

Because the expert review was not restricted to specific procedures it demonstrated a large degree of tailorability. Guidelines and heuristics can be adjusted to correspond more precisely to the evaluation of different software, giving this process a large degree of flexibility.

Observation of novices:

The tools and templates used with the novice observations provide a large degree of tailorability: tasks can be varied and timing adjusted.

User testing:

As with observation of novices, user testing was found to be easily tailorable by adjusting the parameters of the tools and templates used (e.g. the rating instrument).

Because of the broad scope of the processes used, the methodology can be considered to possess an adequate degree of tailorability. Specific tools used are completely customisable (the questionnaire, the novice tasks, the user rating tool and the experts heuristics) allowing the methodology to be adapted to the specific characteristics of most informational/educational systems. The guidelines used with many of the instruments allow a degree of tailorability by enabling the evaluation to be specifically directed at those areas or features which require evaluation. Because tailorability was seen as a major concern right from the start of the project, the final methodology developed demonstrates an adequate degree of tailorability. However not all systems tested will produce as successful an evaluation as the test case, and an improved methodology might provide the possibility of combining a wider range of tools and processes to apply to any software in the same generic category.

Generic applicability

All of the tools and processes developed for this evaluation program are generically applicable to other similar software, making the methodology effective for a wide range of systems. This is due to the flexibility built into the design of the individual components, as described above.

Questionnaire:

The questionnaire affords a high degree of generic applicability because of its flexibility and customisability. Most informational/educational multimedia systems have similar features and

objectives, and the use of the questionnaire in one form or another is appropriate to the effective evaluation of these systems.

Expert review:

The availability of in-house experts was the only limitation to the generic applicability of the expert review. Most systems in the product range of the case study used here can be evaluated using an expert review, if these are available.

Observation of novices and user testing:

Both these instruments also exhibit a high degree of generic applicability.

6.3. Recommendations for improvement

An analysis of the comparative assessment presented above clearly suggests that the methodology developed by this research is far from perfect and can be considerably improved. This could conceivably comprise the object of future research, however a sound foundation has already been established. Improvements to the methodology presented here would include, (but not be restricted to) the following considerations:

Questionnaire

- A more studied approach to the design of the questionnaire is needed so that the value of the returned data can justify the resources expended in applying the tool.
- A more effective strategy needs to be developed to maximise the number of returned questionnaires (only 56 were returned, out of several thousands distributed). Perhaps a prize can be drawn from returned questionnaire numbers to encourage responses.

Expert Review

- More consensus of data could be achieved by a more precise set of guidelines and procedures for conducting the Expert Review

Observation of Novices

- The tasks set to the novice users should be better designed with the aim of returning the most useful data, and to span all possible types of interactions with the system. Perhaps a smaller number of harder tasks should be set (requiring deeper levels of navigation, using mutiple mechanisms, etc.).
- More emphasis should be placed on the quality rather than the quantity of data returned when designing questionnaires, tasks and guidelines.

User Testing

- Better instructions and guidelines need developing for conducting the User Tests. Many of the qualitative responses were of little use because of the frequency of one-word responses. The open questions in the rating tool should be phrased so as not to allow the return of yes/no answers or 1-word replies. Keywords used should be: *discuss.., explain.., elaborate..,* etc.

General

- Efforts must be made to improve the ease with which the methodology can be applied. More specific tools and templates should be created to facilitate the work of experts, users, novices, and test supervisors, such as more detailed guidelines as to the scope of the review.
- The possibility of combining a wider range of tools and processes to apply to any software in the same generic category should be explored.
- A cost/benefit analysis of the methodology needs to be conducted, as well as a test against outcomes of the case study. Again these are possible avenues for future research.

6.4. Summary

While the preceding assessment of the evaluation methodology has identified a number of important weaknesses, it has also clearly demonstrated that the program nonetheless displays most of the features of a good evaluation instrument. Information returned has generally been accurate, sufficient and useful, and mechanisms are in place to ensure the flexible adaptation of the methodology to other informational/educational systems. No significant problems were encountered in the application of the methodology to the case study.

Chapter 7. Recommendations for improving EduKit2000

As a result of applying the evaluation program developed by this research to the EduKit2000 CD produced by Edith Cowan University, a set of recommendations was produced to improve the product in future incarnations. These recommendations are listed in five categories, representing the five areas of inquiry evaluated, and are a synthesis of all the data returned from the five evaluation processes applied:

7.1. Interface Design

- The icons for the Print facility, the Volume Control device, the Live Connection, and the Zoom feature should be improved or replaced.
- Some text fonts and colours need to be revised.
- The backgrounds are often too strong, affecting readability, and should be toned down
- The quality of graphics and photos can be improved since the product is CD based, not web based.
- Windows metaphors such as minimise buttons should be replaced
- Better video compression can be achieved, to avoid slow, jerky viewing. VC video is poor and chunky, frame rate is jerky, and audio is substandard. Should be redone running 25FPS 320x240 with 16bit mono off a standard CDROM
- Step-by-step elements need to be enhanced with pictures and a clearer/less text oriented description of each step… this is especially true of the remote-access section that a novice student would have difficulty to follow.
- Cursor changes on rollover should be rectified in the Help section
- Text should be anti-aliased
- Packaging graphic design incompatible with CD graphics and should be redesigned
- A richer multimedia mix, including demos and animations is required to effectively transfer information to the user
- A Skip video button should replace the hypertext
- Minimise and quit buttons should be labelled
- Remove inactive "back" and "forward" arrows
- Campus maps don't work… clicking on buildings should tell user what the building is.

7.2. Interactivity/Navigational Structure

- Cross-referencing may be confusing and should be improved or removed
- Enhancing the nav path and maybe making it clickable could probably be used whilst page indicators should be removed

- The quit screen is confusing. There should be a message saying "Are you sure you want to quit?"
- The text menus used are often indistinguishable from the main body text
- Back function (next to find) means back to a previous text content section but does not include the navigation from the home page … back should eventually get you to the main menu.
- The search facility should be improved to return accurate results to any keywords
- The Scrolling mechanism is too slow and should be improved

7.3. Content

- Games in Software section need upgrading
- More information on specific units is needed, including previous class averages and exams, assessment details, etc.
- More descriptions of clubs and societies, services such as the cafeteria, bar, etc. are required.
- More information in the Lifestyle section
- Better instructions in Modem Pool Access section
- Rectify omissions in Credits page
- More synthesising, reducing and reformatting of text for readability.
- Poor image quality (low level JPEGs) needs improving
- To a void violation of both the Quicktime and the Macromedia licensing agreements, a screen is needed which is displayed for at least four seconds on quit showing respective logos.

7.4. Scope

- Reduce overall volume of content
- Increase scope of Search facility
- Continue developing EduKit2000 by students
- Indicate length of time video runs
- Design a more intuitive paging system
- More multimedia content should be used for instructional purposes, e.g. demos, animations, interactions
- A more effective Help facility is required

7.5. Overall functionality

- Improve scrolling

- Some games do not install properly (e.g. Solitaire)
- Video should play correctly on all platforms
- Slow start-up and scrolling need to be improved
- Menu options should work on mouse up rather than mouse down
- FAQs need reformatting for more efficient access

7.6. Summary

The recommendations presented above represent all the changes that this research has determined are necessary to improve EduKit2000, and to enable it to better achieve its primary objectives. Only those improvements presented in the Interface Design section are likely to incur large costs. Most other development changes can be performed in the course of an annual upgrade of the product, easily and cost-effectively. The evaluation has determined that while the tested product (EduKit2000) is of a generally high quality, a number of improvements are necessary in several major areas to produce a truly useful informational/educational system.

Chapter 8. Conclusion

This project has meandered through the process of evaluating informational/educational systems. It has explored the work already conducted by theorists and previous researchers in the evaluation of multimedia software, with particular attention to a study of the tools and processes most appropriate to informational/educational systems, which are seen to be increasingly used by universities and other educational institutions, as the use of computer technology in education becomes more widespread.

A methodology was designed and developed with the objective of providing the best possible evaluations of such systems, in order to enable their improvement. This methodology consists of several independent components, each intended to test or evaluate specific features or attributes, using the minimum possible resources yet achieving the best possible results.

Once this methodology was developed, it was tested on a case study, the EduKit2000 CD, in order to assess its performance and to identify any problematic areas in its application. The accuracy of the results cannot be established without comparing the results with a test on outcomes, which has not as yet been conducted.

A by-product of testing the methodology on EduKit2000 has been the production of a comprehensive summative evaluation of the product, accompanied by a valuable list of recommendations for improving it in future incarnations.

Finally the test enabled weaknesses and faults in the methodology to be identified. This will enable the improvement of the methodology, if further research was conducted in the area.

The research conducted for the purpose of this thesis may have opened a Pandora's box rather than put the lid on a method. Rather than answer all questions it has actually raised a lot more, but has at least highlighted the paucity of available developed resources for evaluating this type of software, and pointed to the direction future research should take to remedy this.

Chapter 9. References

Alessi, S.M., and Trollip, S.R. (1991). *Computer based instruction: methods and development*. Englewood Cliffs, NJ: Prentice Hall.

Barker P.G., and King T.R. (1993). Evaluating interactive multimedia courseware-A methodology. *Computer Education*. *21*(4), 307-319.

Bennett, S., Priest, A., and Macpherson, C. (1999). Learning about online learning. *Australian Journal of Educational Technology*, *15* (3), 207.

Brown, M.I., Doughty, G.F., Draper, S.W., Henderson, F.P. and McAteer, E. (1996). Measuring learning resource use. *Computers and Education, 27* 103-113.

Callison, D., and Haycock, G. (1988). A Methodology for student evaluation of educational microcomputer software. *Educational Technology*, *28* (1), 157-182.

Clark, R.E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53* (4), 445-459.

Collings, P.A., Kleenan, D., Richards-Smith, A., and Walker, D.W., (1997). Developing new group work practices: an evaluation of the design and use of groupware-based work systems for a graduate student course. In Information Systems, Gable G.G., and Weber (eds) The Confluence of Theory and Practice: *3rd Pacific-Asia Conference on Information Systems Proceedings,* QUT Brisbane, 185-192.

Collings, P.A., Pitt, C., and Walker, D.W. (1998). Enriching information technology courses through flexible delivery in Goodell J.E., (ed) *Proceedings of the Australasian Joint Regional Conference of GASAT and IOSTE* Curtin University, 283-289.

Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J. and Young, R. M. (1995). Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. In: Nordby, K., Helmersen, P., Gilmore, D.J. and Arnesen, S. A. (eds), *Human-Computer Interaction: Interact '95.* Champing and Hall.

Desurvire, H. W., Kondziela, J. M., and Atwood, M. E. (1992). What is gained and lost when using evaluation methods other than empirical testing. In: *People and Computers* VII, Monk, A., Diaper, D., and Harrison, M. D., (eds). Cambridge: Cambridge University Press. 89-102.

Draper, S.W., Brown, M.I., Edgerton, E., Henderson, F.P., McAteer, E., Smith, E.D., and Watt, H.D. (1994). *Observing and measuring the performance of educational technology* TILT project report no.1, Robert Clark Centre, University of Glasgow.

Duchastel, P.C., (1987). Structures and methodologies for the evaluation of educational software, *Studies in Educational Evaluation*, *13*, 111-117.

Erickson, T. D. (1990). Working with Interface Metaphors. In Laurel, B. *The Art of Human-Computer interface Design*., B. (ed), Addison Wesley.

Ewing, J.M., Dowling, J.D., and Coutts, N., (1998). Learning using the World Wide Web: a collaborative learning event. *Journal of Educational Multimedia and Hypermedia* 81, 3-22,

Faraday, P., and Sutcliffe, A. (1995). Evaluating Multimedia Presentations for Comprehension. *Technical Report, Centre for Human Computer Interface Design, School of Informatics*, London: City University.

Faraday, P., and Sutcliffe, A. (1995b). *Walkthrough Evaluations for Multimedia Comprehension.* In IMMI-1, Edinburgh, UK.

Fowell, S., (1998). Applying constructivist principles to the information curriculum. Unpublished, *Proceedings of SITIS98*, UNSW.

Gould, J. D., and Lewis, C. (1985). Designing for usability - Key principles and what designers think. *Communications of the ACM*, 28, 300-311.

Grudin, J., (1990). Groupware and cooperative work. *The Art of Human-Computer Interface Design*. New York: Addison-Wesley.

Heidler, J., (1993). *Multimedia Development Tools.* Available WWW: http://mime1.marc.gatech.edu/MM_Tools/evaluation.html

Hammontree M., Weiler P., and Nayak, N., (1994). Remote usability testing. *Interactions*

Hannafin, M., and Peck, K. (1988). *The Design, Development, and Evaluation of Instructional Software*. New York: Macmillan Publishing.

Hollnagel, E. (1991). The influence of Artificial Intelligence on Human-Computer Interaction: Much ado about nothing? In *Human-Computer Interaction, research directions in cognitive science: European perspectives.* Rasmussen, J., Andersen, H.B. and Bernsen, N. O. (eds). Volume 3. Lawrence Erlbaum Associated, Publishers, U.K.: Hove.

Holmberg,B., (1983).*The Theory and Practice of Distance Education*. London: Routledge

Jacob, L., and Jones, S. 1994. Formal dialogue specification for hypertext and multimedia systems. Technical Report No. 175, *School of Information Sciences*, University of Hertfordshire, UK.

Jeffries, R., Miller, J. R., Wharton, C., and Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques. *Proceedings ACM CHI'91 Conference (New Orleans, LA, April 28-May 2), 119-124.*

Jonassen, D. (1999). Thinking Technology: Towards a Constructivist Design Model. Available WWW: ftp:// Ithaca.icbl.hw.ac.uk/pub/nato_asi/dhj.txt.gz

Karat, C., Campbell, R. L., and Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. *Proceedings ACM CHI'92 Conference* (Monterey, CA.), 397-404.

Laurel, B., (1990). *The art of human-computer interface design*. Reading, MA: Addison-Wesley.

Laurel, B., Oren, T. and Don, A. (1992). Issues in Multimedia Interface Design. In *Multimedia Interface Design*, Blattner, M. M. and Dannenberg, R. B. (eds). ACM Press, New York: Addison-Wesley Publishing Company.

Lindgaard, G. (1994). *Usability Testing and Systems Evaluation: a Guide for Designing Useful Computer Systems*. London: Chapman and Hall.

Luca, J., (1997). *Project Management for New Media*. Perth: New Media

McAdams, M. (1995). *Information Design and the New Media. Interactions*.

McLaughlan, R., and Kirkpatrick, D. (1999). A decision making simulation using computer mediated communication. *Australian Journal of Educational Technology*, 15 (3), 243-257.

Moore, G.M., and Kearsley, G. (1996). *Distance Education*. Belmont: Wadsworth Publishing Company.

Morrison D. J. (1987). Interactive learning systems and the learner. *Aspects of Educational Technology*, 134-138.

Monk, A., Wright, P., Haber, J., and Davenport, L., (1993). *Improving Your Human-Computer Interface: A Practical Technique*. London: Prentice Hall.

Muller, E.W. (1985). Application of experimental and quasi-experimental research designs to educational software evaluation. *Educational Technology, 25* (10).

Nielsen, J. (1992). Finding useability problems through heuristic evaluation. *Proceedings ACM CHI'92 Conference* (373-380). Monterey, CA.

Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings ACM CHI'90 Conference* (249-256). Seattle, WA.

Oren, T. (1990). Designing a New Medium. In: *The Art of Human-Computer Interface Design*. Laurel, B. (ed), Addison Wesley.

Preece, J., and Jones, A. (1985). Training teachers to select educational computer software: Results of a formative evaluation of an Open University pack. *British Journal of Educational Technology,* 16(1).

Rasmussen, K. L. (1993). The University of West Florida EME 6990, Programming with Web Based Tools.

Reeves, T.C. (1993). Evaluating Technology-Based Learning. In Piskurich, *A.S.T.D. Handbook of Information Technology.*

Reeves, T.C. (1997). Evaluation Tools. Available WWW: http://mime1.marc.gatech.edu/MM_Tools/evaluation.html

Reeves, T.C. (1993). Systematic evaluation procedures for instructional hypermedia/ multimedia. *American Educational Research Association Journal*.

Sims, R. (1999). Interactivity on stage: Strategies for Learner-Designer communication. *Australian Journal of Educational Technology*, 15 (3), 270.

Schueckler, L.M. and Shuell, T.J. (1989). A Comparison of software evaluation forms and reviews. *Journal of Educational Computing Research*, 5(1), 17-33.

Spool, J.M., Scanlon, T., Schroeder, W., Snyder, C., and DeAngelo, T. (1999). *Web Site Usability: A Designer's Guide.* California: Morgan Kaufmann.

Tucker, R., (1989). Software evaluation: Who cares? In Bell. C., et al., (Eds.), *Aspects of educational training and technology* - Volume XXII: Promoting Learning, (8 –16). London: Kogan Page.

Tweddle, S., Avis, P., Wright, J., and Waller, T., (1998). Towards criteria for evaluating websites. *British Journal of Educational Technology*. Oxford: The Alden Press.

Utting, K., and Yankelovitch, N. (1989). Context and orientation in hypermedia networks. *ACM Transactions on Information Systems,* (7), 58-84.