EVALUATION OF A SMALL-SCALE IMPLEMENTATION OF COMPUTER ASSISTED ASSESSMENT

Carole McLennay

Evaluation of a small-scale implementation of Computer-assisted Assessment

Carole McLennay
Department of Computer Science & Information Systems
Chester College of HE
mclennay@btinternet.com

Abstract

Due to the sheer number of students who are enrolled for the "Introduction to Computer Applications" module in the Department of Computer Science and Information Systems at Chester College of Higher Education, the use of computer-assisted assessment (CAA) was investigated during 1999/2000. This investigation began as an MSc project of a member of staff to implement and evaluate the introduction of CAA into the Department. The Department is keen on encouraging an e-Learning environment and seized the opportunity to use the research to evaluate the use of computer-assisted assessment on the "Computer Applications" module in the first semester of 2000/2001.

The purpose of the introduction of CAA into the Department was to evaluate the benefits and/or difficulties likely to be experienced on such a small scale. The purpose of the evaluation of CAA on this module was to measure the students' attitude towards CAA and their perceived confidence pre and post treatment.

A database of appropriately coded questions was developed, some of the questions having been tested (for discrimination index and facility value). CAA software was then investigated and quizzes were designed in Coursebuilder, a plug-in to Macromedia's web authoring application, Dreamweaver. These quizzes were uploaded onto the department web server and administered to one of three sets of students taking the first year module.

Since the number of students taking the module was expected to be high it was appropriate to select a between-subjects design for the evaluation of the CAA use, using a 'treatment' group and a 'non-treatment' group, the latter group receiving the computer-assisted assessment. Data collection methods used included pre- and post-test confidence logs, together with an attitude survey and comparison of the final marks of the two groups. To further improve validity, a focus group interview was carried out with some of the students in the treatment group and the tutor was interviewed separately.

This paper outlines the difficulties encountered during the implementation of the technology, describes the methodology and analyses the results obtained.

Keywords

Computer-assisted Assessment, e-learning, e-evaluation, Coursebuilder.

Introduction

The Department employs twelve lecturers, of which three are part time, together with five support staff. Currently enrolled are 503 students of which 153 are post-graduates.

The purpose of the investigation into the implementation of CAA was two-fold: (1) to examine the attitude of the staff to CAA and evaluate the benefits and/or difficulties likely to be experienced by a small-scale implementation of CAA within the department, and (2) to measure the students' attitude towards CAA and their perceived confidence pre and post treatment. The CAA was designed to test the students' knowledge and also to help the students assess their own knowledge.

A glossary of acronyms and terms used is provided at appendix 1 for information and clarification. For ease of reading, in this paper the lecturer, who delivered the module and administered the CAA to the students in the tutorials, has been named the tutor.

Research Methods

Prior to commencement of the implementation, staff in the Department completed a staff questionnaire - an attitude survey, designed to elicit their knowledge of, and attitude towards CAA, and the tutor delivering the module also completed the same questionnaire post-intervention

The implementation of CAA was carried out using the Introduction to Computer Applications module that comprised three groups of students, one of which was selected by the tutor as the treatment group, the remainder forming the baseline. The treatment group was exposed to two formative quizzes during the module and one summative quiz at the end of the module. All students completed pre-confidence logs at the start of the module and post-confidence logs at the end of the module and an attitude survey at the start of the module; the treatment group also completed an attitude survey at the end of the module. A comparison was made between the final coursework marks of the treatment group and the non-treatment group as well as the computer confidence of both these groups. Some members of the treatment group met as a Focus Group, the tutor delivering the module was interviewed at the end of the module and a record was kept of the experience of the attempted implementation.

The results of the staff questionnaires were analysed, as were the results of the student pre- and post-treatment questionnaire and confidence logs, and also the responses to the questions in the quizzes. The statistical analyses comprised examination of descriptive statistics, tests of correlation and a Mann-Whitney U-test on the post-treatment responses of both the Treatment Group and the Non-Treatment Group to the students' attitude to their confidence in computer use, together with a comparison of these students' final coursework marks.

Issues of Implemenation

o   Question Bank Issues

A paper-based pilot study of questions was undertaken in the previous academic year, the responses to tests being analysed for Facility value and Discrimination Index. These questions were input to a question bank/database developed to house a bank of questions.

One of the advantages of using CAA is said to be the ability to statistically analyse results, [9], [11], [3], and in the TLTP3 Update [12] it was reported that the establishment of question banks at a national level was one of the most cited future developments that educational technologists wished to see. Analysis of the question items in quiz 1 showed five questions falling within the recommended limits for Facility Value and Discrimination Index whereas the results of analysis in respect of quiz 2 showed only 2 questions falling within the recommended limits for Facility Value and Discrimination Index; no results being available for analysis with quiz 3. Significantly, the tutor who designed and delivered the module designed the questions himself for quiz 1, and the person carrying out the evaluation designed all other questions, as the tutor himself did not have the available time to spend on learning how to design appropriate multiple choice questions or to actually set all of the questions. Questions are always likely to be more appropriate if the person responsible for the teaching on the module is closely involved in the design or adaptation/selection of questions from a question bank. It would be sensible to insist that the tutor delivering the module is the person who selects the items to be included in quizzes when carrying out evaluations in the future, as this would enhance the likelihood of suitable questions being used.

Concern is currently being expressed regarding the unreliability of discrimination indices if they are only based on a few hundred examinees, the validity of published tables of criterion values, and that where guessing of answers is not discouraged (for example by negative marking) the indices are even less reliable [4]. These concerns should be taken into account when carrying out analyses on the student responses to tests. It should be ensured that such item analyses are appropriate and that, when such items are subsequently coded, this is made clear in some way.

Software Issues

Dreamweaver is Macromedia's professional Web authoring environment that facilitates good web site management. The CourseBuilder extension for Dreamweaver works with Dreamweaver 4 and is said to enable quick creation of compelling Web-based instructional content that works across multiple platforms and browsers. CourseBuilder for Dreamweaver includes a gallery of over 40 pre-built interactions such as multiple-choice, drag-and-drop, fill-in-the blank and more, a visual manager (wizard) for creating complex interactivity without having to know JavaScript, the ability to send student information and results to any AICC-complaint Learning Management System (LMS). Coursebuilder was investigated and

considered to be suitable, especially in terms of its range of question types, ease of use through wizards, multi-media provision, variety of types of feedback, suitability for formative and summative assessment, promise of data-tracking facilities which would enable results to be output to a Learning Management System (LMS), Instruction Management System (IMS) or a table on a secure server for subsequent analysis. This software, it was thought, would certainly go some way to "providing the seamless integration with existing user databases at campus level" desired by educational technologists in the CAA Centre TLTP3 Update. [12] Coursebuilder is suitable for delivery across the WWW, and is not thought to present problems of interoperability since it is AICC (Aviation Industry CBT Committee) compliant [1]. Coursebuilder proved to live up to almost all of its promises except one important issue – that of data tracking. It certainly is possible to utilise its data-tracking facilities provided an appropriate (and very costly) LMS is *in situ* and/or that programming expertise is available to utilise/modify Course builder's in-built data-tracking template files which, incidentally, provide only the transfer of a final mark (suitable for summative assessment) but not the responses to individual questions (suitable for formative assessment and subsequent analysis). It should be mentioned that Macromedia supply other learning products, which do provide built-in data tracking and can be seen at www.macromedia.com/elearning [13] such as the Authorware courseware, used by Tripartite Interactive Assessment Delivery System (TRIADS) [14]. However, this was not investigated at this time since it was thought to be more a CBT authoring application than a CAA application, and due to time constraints *vis a vis* what was thought to involve a steep learning curve.

Quiz Issues

The first two quizzes were intended for formative use and the third designed for summative use. They were administered at weeks 5, 9 and 12 respectively throughout the twelve-week teaching period.

- A multi-page quiz (a single question on each page plus a summary page). The summary page was a table giving the score for each individual response together with the time taken to answer each question and the number of attempts on each question.

- A one-page quiz which provided, on each occasion that the user submitted a response, remedial information as well as a correct/incorrect response.

- A one-page quiz, which provided a final score only, in a pop-up window.

The quizzes can be viewed at *www.195.195.128.170/student/95023016* [5]. It was not possible to implement the data-tracking facilities described in Coursebuilder, and the students were requested to print the results of their quizzes and hand them to the tutor. This, they were happy to do with the first two quizzes, the results of which resulted in only one or two pages of printing. However, the students refused to print out the result of the third quiz (resulting in five pages) due to the cost – they are required to use print credits at their own expense. The consequence of this was that only the results of the first two quizzes were available for analysis. This loss of data for analysis could have been avoided with a little foresight or perhaps the inclusion

of a reminder to check for printing facilities in, for example, the Pre-test audits at 11.3.1 of the Blueprint for Computer-assisted Assessment [7].

Staff Questionnaire Issues

The staff questionnaire, aiming to find out what knowledge/experience the staff had of CAA and their attitude towards CAA, was administered over the summer period prior to the implementation. It had been assumed that student numbers on this module would be high and that at least three staff members would have been engaged in delivering the module, and therefore actively involved in the implementation of the CAA. These involved staff members would have completed a post-intervention staff questionnaire so that comparisons could be made between their attitudes pre and post-intervention. This assumption was wrong and only immediately prior to commencement of the implementation did it become apparent that numbers on this module were considerably reduced. This decrease in student numbers meant that only one lecturer was employed to deliver the module across the three student sets, and hence, appropriate to complete the post-intervention staff questionnaire.

Results

Staff Questionnaire

There was a 69% response rate to the staff questionnaire comprising 8 responses from academics (2 with previous experience of objective testing) and 3 from support staff. All respondents agreed that students benefit from CAA, 80% of respondents considered that use of CAA provides a fair method of assessment, and 91% that CAA can add validity to the testing procedure. There was wide agreement that it is difficult to design good multiple-choice questions (90%), that CAA should be an optional component, that it should be piloted prior to summative use (75%) and that, to be effective, CAA requires continuous evaluation (91%). Only 9% of respondents consider that learning CAA puts unnecessary pressure on students, 27% of respondents consider that the introduction of CAA will increase demands on lecturer time and only 11% agree with the statement that "Lecturers find that when CAA is implemented, their workload will increase, the 2 academics with previous experience of objective testing, disagreeing. 34% of respondents believe that there is a problem with CAA in respect of security.

Surprisingly, of the 6 academic respondents, half agreed that CAA is NOT suitable for summative assessment (the 2 academics with previous experience of objective testing, disagreeing). However, 70% of all respondents agree that CAA is suitable for both formative and summative assessment.

Every subject responded to the statement 'CAA enables movement away from memory testing and it is interesting to note that of the two respondents with previous CAA experience, both agreed (one strongly), and of the remaining 6 academic respondents with no experience of CAA, 3 disagreed. There was uncertainty over the view that CAA enables wider coverage of the syllabus (55%). There were mixed

views as to whether CAA provides a more reliable method of assessment, and that CAA is less rigorous than traditional forms of assessment with 45% agreeing, the 2 academics with previous experience of objective testing disagreeing

Of the 8 respondents who considered that CAA would be useful for the department:

- 8 considered that it would save time with marking
- 5 considered that it would facilitate evaluation of our teaching methods
- 5 considered that it would facilitate increased reliability
- 5 considered that it would facilitate increased validity
- 7 considered that it would facilitate immediate feedback
- 4 considered that it would facilitate appropriate feedback.

Tutor post-treatment questionnaire

In the pre-implementation staff questionnaire the tutor who delivered the module, agreed with the statement "CAA is NOT suitable for summative assessment" and disagreed with the same statement post-implementation. He changed his views from Disagree to Agree on the following statements: "Cost-savings will be made when CAA is introduced into a department", "The introduction of CAA will increase demands on lecturer time", and "CAA results in saving of lecturer time".

He previously disagreed that "Learning how to use CAA software puts unnecessary pressure on students", that "CAA is ONLY suitable for formative assessment" and that "Students do not gain any benefit from CAA" and now strongly disagrees. On the following six statements the lecturer's views strengthened from Agree to Strongly Agree.

- "CAA should be an optional component of course design."
- "CAA provides a fair method of assessment."
- "CAA is suitable for first year students."
- "CAA can add to the validity of the assessment procedure."
- "Questions used in CAA should be piloted before use in summative exams."
- "To be effective CAA requires continuous evaluation."

In addition the tutor gave a further reason why he considers that CAA would be useful for the department. This is that CAA could facilitate evaluation of our teaching methods.

Student Pre and Post-treatment questionnaire

None of the students had experienced computer-assisted assessment previously but 21 indicated that they had experienced paper-based objective assessment and students favoured, firstly a mix of objective and traditional assessment and secondly objective assessment. All students indicated that they would like to take a computer-assisted examination. There were no significant responses to the open-ended questions designed to elicit their views on the benefits and/or weaknesses of traditional assessment.

Feedback

All students pre-treatment agreed that feedback during their studies is important to them, 90% agreed that how quickly they receive feedback is important to them. 68% would be prepared to do more testing in order to receive more feedback and 91% would undertake more training in order to receive more feedback. From the results of the post-treatment questionnaire, the treatment group agreed that feedback during their studies is important to them and how quickly they receive feedback is important to them. The treatment group were also asked whether they would undertake more testing in the use of CAA in order to receive more feedback both pre and post treatment and their agreement rose from 63% to 89%. The treatment group were also asked whether they would undertake training in the use of CAA in order to receive more feedback both pre and post treatment and their agreement rose from 75% to 89%.

The question "Which of the following aspects of feedback do you consider to be more important, Appropriate or Immediate"? was asked of all students both pre and post treatment. All students responded and indicated that 'Appropriate' feedback was more important. The actual results were 92.5% pre-treatment and 100% post-treatment. Interestingly, as shown in the table at appendix 3, two-thirds of the students expressed a preference for quiz 2. This quiz provided remedial feedback and information on whether the response was correct or incorrect, via a pop-up window which appeared immediately the student submitted an answer The numbers of responses and reasons given for the preference for quiz 2 were as follows:

Question	No. of responses	Reason given for preference for quiz 2
Which quiz did you like best?	4	You know where you are
	2	Answer given so more learnt

Table 1 – Student Responses

The student preference for quiz 2 was confirmed during the focus group interview and during the interview with the Tutor administering the quizzes.

Computer Confidence Logs

The responses to the question "How confident are you about using a computer?" for both the treatment and non-treatment groups from the student post-treatment confidence logs were analysed using a Mann-Whitney U Test for independent samples, to see whether the treatment group were significantly more confident. As expected the results showed no significant difference between the results of the two groups and the average final mark for coursework was 62 for each group.

Qualitative analysis

Both a Focus Group interview (7 members) and an interview with the tutor delivering the module were carried out in order to provide qualitative data in support of the quantitative analysis of the student questionnaire and confidence logs. It was also intended, for further triangulation, to include participant observation but this was not possible. The table, illustrating the results of discussion during the Focus Group interview together with the responses to the same questions in the interview with the tutor who delivered the module and administered the quizzes, is shown in appendix 3.

Tutor Interview Results

From the interview with the tutor the following views were elicited. .

- Because of the unfamiliarity with the system, it took time out of the tutorial to get the students to access the quiz and print off their results.

- It was costing the students money to print off the results as the cost was deducted from their own print credits and, in the case of the third quiz, the printing ran to five A4 pages.

- The tutor did not approve of some of the questions provided in some of the quizzes and wished to insert questions that he considered to be more appropriate.

- The quizzes were administered in a tutorial taken by the tutor. He considered that it might have been more expedient had the quizzes been delivered in a workshop situation where support staff would have delivered the quizzes and the exercise would not have encroached onto teaching time.

- The tutor welcomed the introduction to CAA and showed a keen interest in using it in the future.

- The tutor plans to use objective testing for, at least part of, future examination papers.

- The tutor recognised that writing effective questions was not as easy as it seemed and enquired about writing such questions effectively.

Discussions and Conclusions

At the design stage of the experiment, the erroneous assumption that numbers would be high and therefore three lecturers would be involved in the implementation, jeopardised the external validity of the design methodology. Due to the low number of students it is not possible for the results from the samples to be generalised to the

whole population. This error would not, it is considered, have occurred had the project been a team exercise or part of a larger strategic implementation.

In general, the staff questionnaire results agree with CAA Centre Survey on, for example, the difficulty of writing good multiple choice questions, concerns over time issues, the perceived low level of student anxiety, and agreement over the suitability of CAA for level one/two. Interestingly, only 34% of respondents in the staff questionnaire believe there is a problem with CAA in respect of security but in the CAA TLTP3 Update [12], security issues such as security of the results, confidence in hardware, software and marking systems were mentioned as concerns and educational technologists cited "improved security" as one of the future developments in CAA they would like to see. In the staff questionnaire uncertainty in attitudes was demonstrated by lecturers (particularly those without previous experience in objective testing) about statements such as CAA enables wider coverage of the syllabus, CAA enables movement away from memory testing, CAA is NOT suitable for summative assessment, and CAA provides a more reliable method of assessment. Perhaps this level of uncertainty over these important issues reveals a pressing need for training of existing staff in CAA issues and that this training should be implemented even before the topic of CAA implementation is introduced in an institution. The tutor who delivered the CAA changed his attitude on a total of twelve questions following his experience of using CAA. Certainly there is little point in evaluating staff attitudes about a topic before they are made aware of its benefits/limitations. "Innovation will necessarily require changes in what exists already, and if this is not acknowledged and accommodated, then the innovation will not succeed." (Laurillard 1993). Lecturers today exist in a fast-changing culture and existing staff particularly, need regular and up-to-date training to enable them to develop coping strategies for the dynamic environment in which they find themselves.

The Focus Group interview and the tutor interview mainly confirm the findings from the analysis of the student pre- and post-treatment questionnaire and confidence logs but the Focus Group members and tutor held opposing views over whether or not the CAA exposure hindered the students in their studies, the students considering that they could have been doing more work rather than filling in questionnaires. Parlett and Hamilton (1987) believe, "the observation phase occupies a central place in illuminative evaluation", and, as they describe, a record was "built up of ongoing events, transactions and informal remarks" which proved a useful tool in understanding attitudes of the lecturers and the students. It was planned to carry out participant observation in a number of the tutorials but due to time-table clashes this was not possible. Information gained by this strategy would have further supported the qualitative data from the Focus Group interview, thus aiding the process of triangulation. It is felt that the opportunity was lost because an individual instead of a team undertook the evaluation, since another team member could have carried out the participant observation. This loss of data for analysis could have been avoided with a little foresight or perhaps the use of a reminder in, say, the Blueprint for Computer-assisted Assessment [7].

One of the advantages of CAA to students appears to be the remedial feedback provided in formative use and one of the main advantages to higher education institutions is said to be time saved in marking in summative use.

The evaluation of the small-scale intervention identified and emphasised the importance of some of the implementation issues likely to be encountered in any future implementation of CAA. It also highlighted the need for more staff awareness of and training in CAA prior to any future implementation. The investigation into students' attitude towards CAA and their perceived confidence pre- and post-treatment confirmed their liking for CAA used summatively and for remedial feedback. "Integrative evaluation" [6] will be an important tool in CAA implementation and help to ensure that CAA fits well into the '"learning milieu"[10] [8].

Acknowledgements

References

- Brogan, P. (2001) *Using the Web for Interactive Teaching and Learning.* www.macromedia.com/resources/elearning/(14/05/01) (click on More Resources)

- (Ellington, H., Percival, F., and Race, P. (1993) *Handbook of Educational Technology*, Third edition. Kogan Page, London

- Bull, J. (1999), "A Glimpse of the Future", in Brown,S., Bull,J., and Race,P. (Eds. *Computer-Assisted Assessment in Higher Education*, Kogan Page, London.

- Burton R.F (2001) "Do Item Discrimination Indices really help us to improve our tests?" *Assessment & Evaluation in Higher Education* 26(3) June 2001.

- www.195.195.128.170/student/95023016 (14.05.01)

- Draper, S., in *LTDI Implementing Learning Technology* "Observing, measuring or evaluating courseware: A conceptual introduction". www.icbl.hw.ac.uk/ltdi/implementing-it/measure.htm (14/05/01)

- http://www.caacentre.ac.uk/blueprint.shtml

- Laurillard, D., (1993) *Re-thinking University Teaching*, Routledge, London.

- McKenna, C., and Bull, J., (2000) "Quality assurance of computer-assisted assessment: practical and strategic issues", *Quality Assurance in Education* Vol 8 Issue 1, 2000. www.emerald-library.co/brev/12008ac1.htm (7.5.00)

- Parlett, M.R. & Hamilton, D. (1987*) "*Evaluation as illumination: A New Approach to the Study of Innovatory Programmes". In Murphy, R. and Torrance, H. (1987) (Eds) *Evaluating Education: Issues and Methods.* Paul Chapman Publishing Ltd, London

- Stephens, D**.** and Mascia, J. (1995), *Results of a Survey into the Use of Computer Assisted Assessment in Institutions of Higher Education in the UK 1995* , Flexible Learning Initiative, Loughborough University.

- McKenna, C. and Bull, J. (2000) Update on the National TLTP3 Project: The implementation and Evaluation of Computer-assisted Assessment".

- www.macromedia.com/resources/elearning/(14/05/01)

- www.derby.ac.uk/assess/talk/demopage.html (13.05.01)

Appendix 1

Glossary

AICC	Aviation Industry CBT Committee	AICC is an open forum of training professionals that develops guidelines for interoperable learning technology that began a decade ago in the commercial aviation industry.
ASP	Active Server Pages	Middleware
CAA	Computer Assisted Assessment	A term used to describe the use of computers to support the assessment of student learning.
CBT	Computer-based Training	A term used to describe the use of computers to support teaching and learning.
CIT	Communications and Information Technology	
CMI	Computer Managed Instruction	An instruction system managed by computers
CTI	Computers in Teaching Initiative	
Data-tracking	Data-tracking	Data-tracking is the facility to transfer the results of student responses into a table, database or LMS.
DI	Discrimination Index	The results of a formula used for calculating the potential of a question to distinguish between stronger and weaker candidates.
FTP	File Transfer Protocol	
FV	Facility Value	The result of a formula to measure the difficulty of a question by dividing the number of correct responses by the total number of responses.
HE	Higher Education	
HEFCE	Higher Education Funding Council in England	
HEI	Higher Education Institution	
IBIS	Internet Based Information System	
IMS	Instructional Management System	A computer system for managing and integrating aspects of Instruction

ITTI	Information Technology Training Initiative	
LMS	Learning Management System	A computer system for managing and integrating aspects of Learning
LTDI	Learning and Teaching Dissemination Initiative	
LTSN	Learning Technology Support Network.	Network of subject centres based in HE institutions.
MCQ	Multiple Choice Question	
TALENT	Teaching and Learning Using Network Technologies	
TLSS	Teaching and Learning Support System	A service of Chester College of HE that maintains extensive resources for staff and students.
TLTP	Teaching and Learning Technology programme	
TLTSN	Teaching and Learning Technology Support Network	
TRIADS	Tripartite Interactive Assessment Delivery System	Utilises interactive diagrammatic and graphical features to expand question formats.