EXPERIMENTS WITH ELECTRONIC EXAMINATIONS OVER THE INTERNET

Pete Thomas, Blaine Price, Marian Petre, Linda Carswell, Mike Richards

Experiments with Electronic Examinations over the Internet

Pete Thomas, Blaine Price, Marian Petre, Linda Carswell, Mike Richards
Centre for Informatics Education Research (CIER)
Department of Computing
Open University
Milton Keynes
P.G.Thomas@open.ac.uk

Abstract

The UK's Open University (OU) has been using the Internet on a regular basis for transporting student assignments (homework) between student, tutor and the university. Tutor marked assignments are a major part of the OU's assessment system, but all courses also have an examinable component that often takes the form of a three-hour closed-book examination taken under supervised conditions. Our aim has been to investigate the extent to which we could use the Internet to automate the examination process. In this paper we discuss the results of two experiments that we have carried out so far, and discuss how examinations might be carried out electronically in a distance education setting.

The first experiment was performed in 1997 when students sat a formal supervised examination in which the examination paper and the students' answers were transmitted between the OU and the remote examination sites using electronic means.

The second experiment, conducted early in 1999, enabled students to take a 'mock exam' accessed via a web page as part of their revision. The experiment was designed to test out the technical feasibility of offering an unsupervised home examination. Students accessed the paper via a web site and submitted their answers in a similar way. Whilst cheating: is an important issue, there are good pedagogic reasons for providing an unsupervised home exam that will be discussed in this paper.

The experiments have shown the feasibility of conducting conventional examinations electronically and, in principle, such examinations can be low-cost and be taken securely in the home. In addition:

- Most students found answering the exam at home a positive experience.
- There are real advantages, in a distance education setting, of sitting an exam at home.
- The provision of a mock exam adds value to a course.
- Designing an infrastructure for an electronic exam is straightforward, although more effort is required to allow students to submit graphical answers.

- The rigorous time limit placed on an exam (whether conventional or electronic) is an issue for students. The electronic exam increases anxiety over time since the student is also responsible for ensuring that their answers are submitted on time. We expected a similar heightening of anxiety over downloading the question paper, but this was not mentioned by the students

We recently completed a third experiment which included fully automated grading of all exam questions (none of which were multiple choice or fill-in-the-blank) and we report our initial findings here.

Keywords: electronic examination, automatic assessment, remote examination, automated grading

Introduction

At the UK's Open University (OU) we have been using the Internet on a regular basis for transporting student assignments (homework) between student, tutor and the university (Petre and Price, 1997)(Thomas et al, 1998). Tutor marked assignments are just part of the OU's assessment system: all courses have an examinable component that often takes the form of a three-hour closed-book examination taken under strictly supervised conditions. Our aim has been to see the extent to which it would be possible to use the Internet to automate the examination process. In this paper we discuss the results of two experiments that we have carried out so far, and postulate about how examinations might be carried out electronically in a distance education setting.

The first experiment was performed in 1997 when students sat a formal supervised examination in which the examination paper and the students' answers were transmitted between the OU and the remote examination sites using electronic means. The second experiment, conducted early in 1999, enabled students on a regular course to take a 'mock exam' accessed via a web page as part of their revision. All of our work on examinations and testing has focussed on free-text entry style answers as opposed to the multiple choice, or fill-in-the-blank style of testing most commonly used (e.g. (Bocij, 1999)).

The First Experiment: transmitting papers and answers via the Internet

Whilst the majority of Open University's 150,000 students are resident in the UK, a significant number study outside the UK (primarily because of work commitments). The OU has traditionally made arrangements for these students to take their examinations in their current location by using local facilities and qualified local invigilators (supervisors). In this way, all students on a particular course will take the same examination at approximately the same time, no matter where in the world the examination centre happens to be.

In the traditional process for remote centres, the examination paper would be sent by post to the invigilator who would have to keep it secure until the start of the examination.

At the end of the exam, the invigilator would be responsible for gathering the students' answers and posting them back to the UK for grading. Clearly, using the post (or even special couriers) means that the examination paper has to be available well in advance of the examination date, if it is to be delivered to the remote site in time, and there will be another significant delay before the answers are received by the markers. Often, marking is done in teams who begin their task immediately the examination has finished with the aim of finishing the task as quickly in order that the final results can be published as soon as possible. Examination scripts that come from remote locations can be received too late to take advantage of the team marking and suffer further delays as a result.

It is quite clear, therefore, that the electronic transmission of both the examination paper and the students' answers can reduce the transmission time dramatically and obviate the difficulties mentioned above. In the 1997 experiment we arranged for the examination paper on the post-graduate course MZX867, Human Computer Interaction, to be downloaded as an encrypted document via a secure (password protected) web page by the invigilator. Three students participated in the pilot and were each provided with a PC at the examination centres (in three different European Centres). Since some of the students were native English speaker and the native keyboard in the country in which they were taking the exam was not English, each was allowed to bring their own keyboard if they wished. Just before the examination began, the invigilator used their password to access the secure server containing the exam paper that they downloaded and printed. Once the examination had started, students were free to work from the printed paper or the screen, although their Internet connection was disabled during the exam. The students keyed their answers into an ordinary word processor, the results of which were encrypted and returned via a secure web page by the invigilator at the end of the examination period (Carswell et al, 1999).

In this model, it is the invigilator's task to ensure that a clean machine (one without files that reference or pertain to course materials) with a network connection is available for each student.

In the experiment, the downloading of the examination paper and the uploading of the students' answers went smoothly. On receipt of the scripts at the OU, the students' answers were printed out and added to the conventional scripts for grading in the usual way. From this point onwards, the 'electronic' scripts were handled in the same way as the 'conventional' scripts (despite the fact that the scripts were obviously word processed, this was not a distinguishing feature since it is not uncommon for disabled students to use word processors in their examinations).

There were two particular points of interest for us in this experiment:

- Did the students find using a word processor under supervised examination conditions a help or a hindrance?
- Did the process materially affect the examination results of the students?

In post examination discussions with the students, it was generally felt that using a word-processor did not affect their performance and the fact that several people in one room were typing was not disturbing to others. In fact, the students were very positive about the use of word processors, with one commenting that she was able to write a lot more than if she had been restricted to using ordinary pen and paper. It must be said, however, that the nature of the course, a post-graduate course in Human-Computer Interaction, meant that students were used to word processing their tutor marked assignments and were sufficiently skilled to answer the questions in the time available. The examination markers commented that the Internet examination papers were much easier to grade because they were word-processed and thus they had a positive attitude to grading them. For some of the questions on the exam, an ideal answer would have included a diagram.

As far as we were able to tell, the results of these students were typical of the class as a whole, although two of the students had exam scores above the average for the course. Again, in post examination discussions, students did not feel that their final grade was affected by their experience although one student commented that she felt she had more time to answer the questions fully because she could type them (she was an experienced touch-typist). There was an element of choice in the questions answered on this exam and two of the three students avoided the question which involved a diagram, but the third student managed to draw an adequate 'box and arrows diagram' using a simple drawing tool. In the post-exam interview this student did report that it was harder to draw the diagram than it would have been with pen and paper but he said that the extra time required did not affect his performance.

The Second Experiment: an examination at home

Our second experiment was designed to test out the technical feasibility of offering an examination taken at home unsupervised. That is, a student would access an examination paper via a web site and submit their answers in a similar way. The main issue is cheating: using someone else to answer the questions on your behalf. We shall return to this issue later.

However, there are good pedagogic reasons for providing an unsupervised exam that can be tackled in the home. For example, in revising, students often use past examination papers to gauge their progress and to obtain a feel for the kinds of questions they are likely to face in their own exam. On some courses, particularly access courses where students are returning to study, mock examinations are provided. Such examinations do not count for credit but provide a means whereby a student can gain experience of sitting an examination under controlled conditions. The fact that a mock examination does not count for credit frees the teacher to relax the controls and be imaginative about what actually happens during the examination, perhaps providing help on request. A mock examination is there solely for the benefit of the student: it is

up to the student to use the opportunity as they see fit. A mock exam is an ideal vehicle for testing ideas and systems: the student benefits as do the researchers.

Therefore, in March 1999 we began an experiment to provide a mock examination to students who were studying on a postgraduate Computing course. The final examination for the course was scheduled for late April 1999 so providing students with the opportunity to take a mock exam in the weeks immediately prior to the final exam was an attractive proposition. We also decided that students would benefit most if we not only marked the student answers but also gave feedback (comments) on a student's performance.

The basic scheme was similar to the First Experiment in which students volunteered to do the mock exam and, as a result, were provided with a password to access a secure web site from which they could access the examination questions. The question paper was a web page containing edit boxes where the student typed their answers. Having finished the exam, the student simply pressed a 'submit' button for their answers to be stored in a database for later retrieval by a marker.

Web browser technology is such that students do not have to be physically connected to the web site for the duration of their examination: the web page is cached on their local PC and the telephone line can be dropped until the time comes to submit the completed paper. (Even with current low-cost ISPs, telephone costs can be a deterrent to the effective use of a web site, and the guarantee that telephone costs will be minimal is often a prerequisite to persuading students to use an on-line facility.)

The basic scheme was enhanced in two significant ways:

1. The time between a student first accessing the exam paper and submitting their answers was recorded. Under normal examination conditions the mock exam would have to be completed within three hours. Although students were told that they could use the mock exam in any way that they chose, we encouraged them to sit the exam in the conventional 3-hour slot by promising feedback on their answers. However, if they chose to take longer, we would only provide sample solutions. In the event, both schemes were used.

2. The mock exam was accessible for a fixed period of two weeks. The provision of a fixed cut-off date enabled us to provide feedback and solutions before the actual exam and without the solutions being available whilst students were still attempting the exam.

At the end of the two-week examination period, a tutor was employed to mark and comment on those papers that were answered within 3-hours, and all students who tried the exam were e-mailed with the solutions.

This scheme attempted to cater for the needs of a diverse student body by providing a relatively lengthy period over which the exam could be taken and allowing students to choose how they used the facility. The mock exam was a model of a final examination having been constructed to match previous real examinations.

The fact that Open University students are adults having other full-time commitments (to families and jobs) means that specifically timed activities such as face-to-face tutorials cannot be scheduled to meet the needs of everyone. Therefore, providing a relatively long period of time over which an activity can take place provides access to a wider group of students. Even then there can be students who are disadvantaged, perhaps because of work commitments. Therefore, we offered all students on the course the opportunity to request a paper copy of the mock exam and its solutions.

The whole experiment was on a voluntary basis and, in the event, approximately one-third of all students took advantage of the mock exam in one of its forms.

The experiment was specifically designed to test three aspects of an electronic examination system:

- the technology of providing secure access to an examination, including timing issues that might apply in a real examination;
- how an electronic examination system might operate;
- the administrative system that would have to support electronic examinations.

The outcome of the experiment was extremely positive. The technology worked well. We put in place a system that provided additional benefits for the student. We developed a system that can form the basis of an operational system that has implications for teachers and administrators.

As a bonus, we have kept the student attempts at the mock exam and will be able to compare their performance to their performance on the final exam to see whether there is any correlation between them. This will enable us to design a large-scale experiment in which we can investigate the pedagogic issues of electronic examinations that we intend to perform later this year. However, our longer-term aim is to study how a more conventional examination might be conducted electronically.

The mock exam structure

The mock exam was based on previous actual examinations. That is, questions were taken from past exam papers and amended in two ways:

- to provide a source of questions that were similar to but sufficiently different from typical actual examination questions;
- to fit the format of the web page used for the experiment.

In so doing, we strove to produce a paper that, as a whole, was a good representative of actual exams and gave coverage of the course overall. Appendix 1 shows the first page of the exam.

The changes we made to fit the web page format were essentially to remove the necessity for students to draw diagrams. Whilst the exam paper itself contained diagrams, the answers had to be entirely in textual format. This was simply a way of producing a web-based paper quickly. In principle, there is no reason why a paper should not require diagrams, certainly not from a technical point of view. There may be an argument to the effect that using a drawing tool in an examination would be too time consuming but this could be minimised in several ways:

- familiarisation with the tool as part of the normal assessment;
- requirement for simple line-diagrams;
- provision of templates.

We also took the opportunity to reword some questions to elicit simpler answers as a prelude to thinking about short answer questions that might be used in the future as the basis of automatic marking.

The Questionnaire

At the end of the examination, students simply clicked on a single ‘Submit’ button. This caused their answers to be uploaded and took them to another page in order to answer a brief questionnaire (on a voluntary basis). The questions asked are shown in Appendix 2.

One student lost all of his answers when he re-connected to submit them after working for 2.5 hours and so reported a very negative experience. About half of the students reported completing the exam and about half though that there was enough time (although these groups do not completely intersect). Most students reported being at least 2-4 finger (10 wpm) typists with 3 of the 9 reporting over 40 wpm. Most thought the experience was comparable with taking a conventional exam with a third feeling it was better. One of those who thought it was worse was a poor typist, did not finish, and did not thing there was enough time allowed. Other than the student with the technical problem, all of the others would take another electronic mock exam with a third of students also saying they would be willing to take a real exam electronically. Only one student complained about not being able to use diagrams and one student noted that being able to use books (cheat) made the exam slightly unrealistic. All the students reported volunteering to help with their exam revision and a third said that they found the congenial environment, not having to travel, and the timing as being positive aspects of the experience.

Marking the Mock exam

The students' answers were stored in a database for later retrieval by a marker who also gave additional feedback on the solutions. The marking was done 'on-screen' and the feedback was sent back to students via e-mail. See Appendix 3 for a discussion of marking issues.

Future work

The experiments we have conducted to-date have convinced us that it is possible to put in place an operational electronic mock examination system that provides additional benefits to students at relatively low cost. Such a system will enable us to validate many of the components of a real examination system via the Internet. Of course, the final hurdle will be to allow students to take an examination in their own home unsupervised. The major obstacle is the need to ensure that cheating does not take place.

There seems to be two ways in which technology can help tackle this issue. One way is to perform a style analysis of a student's assessed homework and compare this with a similar analysis of their examination answers. If a discrepancy is detected, this might indicate that the individual who sat the exam was not the same as the one who did the homework, and steps could be taken to investigate the reason for the discrepancy. Research in this area is required to prove the robustness of such a system (a system that detects discrepancies on a large scale would not be cost effective) and a great deal of work would need to be done to convince the public that the system is reliable.

The other approach is to raise awareness of detection in the students' minds by asking them to negotiate a number of hurdles, each one of which could detect cheating at different levels of certainty. For example, immediately after an examination a student could be phoned and asked a number of simple questions related to the examination paper or even the student's own answers (this might even be done automatically using email). Incorrect or dubious answers would ensure that a second phase of questioning would be invoked. In addition, a proportion of those who correctly answered the first phase of questions would also go on to the second phase. Typically, the second phase would be similar to a viva voce examination in which a student would meet face to face with an examiner who would determine whether or not the student's examination results would be allowed to stand. In other words, the examination system would become an honour system in which the student was aware that measures would be taken to detect cheating and that there was a reasonable chance of being caught out.

However, it can be argued (see Appendix 4) that the only real use of a conventional supervised and timed examination is to verify the identity of the person who submitted the coursework, that is, the exam is there to detect cheating in the coursework. If this argument is accepted, perhaps the need for conventional examinations should be revisited. After all, there are other ways of assessing students. The only merit in conventional examinations is that they are a relatively inexpensive way of assessing large numbers of students, but many would argue that this form of assessment is

inherently flawed. Some of the potential drawbacks of electronic examinations are security and system problems, which we discuss in Appendices 5 and 6.

Our most recent work in this area has just been completed and although we have not fully analyzed the data the results are promising. Although it was again a mock exam, it more closely simulated exam conditions by having all students take the exam during the same 3 hour period with some attempt at remote invigilation. All the students ran Microsoft NetMeeting connected to our server. All had 2 way audio connections and 2 of the 11 students had cameras. During the exam we interrogated each of the students twice at random intervals in an attempt to verify that they were who they claimed to be. In addition, we could invoke the remote desktop sharing facility in NetMeeting to see exactly what each student saw on their screen and what the were doing with their mouse. In the event of technical difficulties, we could take over their mouse (with their permission) and remote control their machine (we had to do this for one student who thought that she had lost all her answers). In addition, we provided a facility for students to save their work to our server at any time.

We used a Lotus Domino server to serve the exam so that we had secure data locally and secure data transmission, although the students used a web browser interface. We used colour more effectively in presenting the questions, but we again required text only answers. Although Domino allows use to have rich text data entry and diagram entry via web pages, we did not activate this feature because the main point of this experiment was automated grading.

We designed a system to encode an answer/grading tree for each question (all questions were either short or long answer essay type). Since our testing domain, computer science, has a limited scope for essay answers, we did not have to employ the more complicated essay analysis techniques such as those used by Whittington (Whittington, 1999) or We were able to send students a provisional grade within 24 hours from the automated grading system. Two tutors were also employed to hand mark each paper, so students also received personalized feedback within a week. We appear to have a very good correlation between the grades automatically assigned and those given by the two tutors, but this will be fully reported in a later paper.

Conclusions

Conventional examinations can be provided via the Internet in such a way that security of the process can be guaranteed (this is ultimately an issue of fairness for each student). It is possible to put in place a low-cost process of mock examination that provides educational benefit to students.

The experiments we have conducted illustrate the feasibility of conducting conventional examinations electronically and, in principle, such examinations can be taken in the home. The biggest challenge in providing a conventional-type real exam via the Internet will always be the authentication and integrity of the student. If the quality assurance

aspects of examinations can be put in place, examinations taken at home via the Internet could solve a number of educational issues.

References

Bocij, P. (1999). *Can computer based testing achieve quality and efficiency in Assessment*. International Journal of Educational Technology (IJET), **1**(1), 108-115.

Carswell, L., Thomas, P.G., Petre, M., Price, B., Richards, M., (1999) *Understanding the 'electronic' student: Analysis of functional requirements for distributed education*. Journal of Asynchronous Learning Networks **3**(1): 7-18.

Thomas,P., Carswell, L., Price, B., and Petre, M. (1998) *A holistic approach to supporting distance learning using the Internet: transformation, not translation.* British Journal of Educational Technology, **29** (3)

Petre, M. and Price, B. (1997). *Teaching programming through paperless assignments: an empirical evaluation of instructor feedback*. In proceedings of: ACM SIGCSE/SIGCUE Conference on Introducing Technology into Computer Science Education (ITiCSE '97), Uppsala, June. SIGCSE Bulletin.

Whittington, D. (1999). *Technical and security issues*. In S. Brown, P. Race, & J. Bull (Eds.), Computer Assisted Assessment in Higher Education (pp. 21-29). London: Kogan Page.

Appendix 1: The Mock Exam

M861 Mock
Postgraduate Level Course Examination
COMPUTER ARCHITECTURE AND OPERATING SYSTEMS
Mock Examination Paper

Time allowed: 3 hours

This question paper consists of two parts. Part 1 is compulsory and consists of 10 short questions, each of which is worth four marks. Part 1 carries 40% of the total marks; you should answer all the 10 questions. You should attempt three from the five questions in Part II, each of which carries 20% of the total marks. The marks allocated to each part of a question are shown on the right hand side of the page in square brackets. **NOTE THAT DIAGRAMS CANNOT BE USED IN ANY ANSWER**. You must answer all questions using plain text.

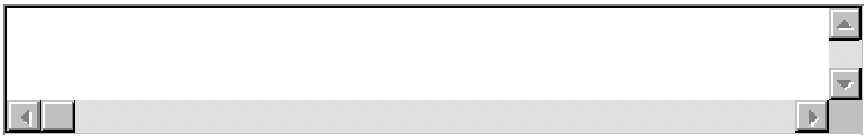
At the end of the examination

Submit your answers via the Electronic Examination web page **on or before 11 April 1999**

Part I
Answer all questions. Brief answers only are required.

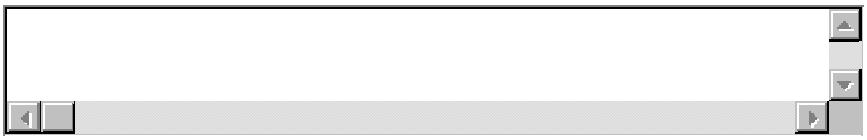
Question 1

(a) Convert 0010 1111 1010 0011 to decimal and hexadecimal [2]



(b) Write the hexadecimal code for the following characters using leading even parity for each byte:

MZX3 [2]



Appendix 2: The Questionnaire

Your exam has been received and accepted, thank you.

We plan to e-mail personalized feedback on each of your answers within 7 days. We want to thank you for participating in this trial mock exam. We hope that it has been useful to you, but you should also know that you are helping us with a research project into electronic examinations.

As such, we would be very grateful if you would take a further 2 minutes to answer the questions below and pressing the SUBMIT button at the bottom when you are finished.

Thanks again from the MZX Team!

1. Did you complete the exam?
2. How much time did you have for the exam?
3. How competent do you feel you are as a typist?
4. Compared with a conventional paper exam, what was your overall impression of taking an electronic exam?
5. Would you take another electronic exam as:
   - a mock exam, a real exam, both, neither
6. What, in your opinion, was lacking from this trial?
7. Why did you decide to take part in this trial?
8. Do you see any advantages in taking an electronic exam (whether as a mock or real exam)? Explain and provide any other comments you wish:

Appendix 3: Marking of electronic examinations

There are two approaches:

- manual
- automatic

For automatic marking, multiple choice type questions including, ‘fill in the missing gaps in an argument using words and phrases taken from a supplied list’. Note that the OU has had Computer Marked Assignments (CMAs) using mark sense cards since 1971.

Manual marking is used where long answers are expected because some interpretation is required.

It is possible to have a mixed-mode that includes both short answer questions (for automatic marking) and long-answer questions (with manual marking).

On course M353 we began with a mixed-mode exam (all manually marked) in which 50% were short answer and 50% were long answer. Some analysis showed that the students' results on the short answer questions were a very good predictor of the students' final grades. Therefore, we changed the weighting to 70% short and 30% long and attempted to make the long answer questions more demanding. The outcome was still that the short answers predicted the final grade extremely well. Ultimately, it is the design of the examination questions that is the main issue.

Appendix 4: Invigilation

The main issue seems to be: is the exam to be done at home or under supervised conditions? If supervised exams are essential then it is necessary to ensure that the machines that are used are 'clean' or that the design of the exam is such that access to materials is not significant (so why not allow students to do the exam at home?).

Whose machines will the students use? Their own, or will they be provided at a centre (with significant cost, possibly)? Providing machines is not a scalable solution.

Is a timed exam essential? If so why? Is the date on which the exam taken significant? If so, why?

One approach to home exams is to build up a large database of questions with the ability to randomise parts of each question. When a student sits the exam, the paper is constructed for that student by a random selection from the database. Perhaps one should consider allowing students access to the database (for revision purposes). Care is needed here because a student could use a computer to store answers to all the questions and simply use the machine to select the appropriate answer. Some way of randomising the individual questions is required to make this process difficult.

Appendix 5: Security

Security of a timed invigilated exam is not a problem provided that the exam is downloaded just before the start time. The transmission of the paper and the students' scripts is not a security issue (it may be a privacy issue, however). Security is as good as the security on the machine from which the exam is downloaded.

How many times should an individual be allowed to submit an answer to an exam? If only one, what happens if the transmission fails in some way and only a partial script is received? An invigilator might have privileges to re-submit.

Appendix 6: System problems

What approach is to be adopted to student queries such as:

- I couldn't download the exam because my computer/modem/telephone line was broken/misbehaving/too heavily loaded or the traffic was too heavy to download in time.

- I could not submit my answers because of some system fault.

One approach to alleviate some of these problems is to get the student to register on-line and to get them to do a mock exam to prove that they and their system work. Avoiding the use of a tightly timed exam is another approach.