

ICT in Assessment and Learning: The Evolving Role of an External Examinations Board

Robert Harding

Director of Interactive Technologies in Assessment and Learning (ITAL Unit)

Email: r.d.harding@ucles-red.cam.ac.uk

Nicholas Raikes

Research Officer

Email: N.Raikes@ucles-red.cam.ac.uk

University of Cambridge Local Examinations Syndicate (UCLES)

1 Hills Road, Cambridge CB1 2EU

1. Introduction

In the world of ICT-led innovation, new styles of learning have an exciting image but assessment usually fails to raise much enthusiasm. And yet it is a vital component of education as a whole. In a paper given at a special QCA conference a year ago [1, Craven and Harding 2001], the interaction between assessment and learning was likened to a three-legged race, in which neither partner can make much progress without the other's contribution. We therefore welcome the focus on assessment in this second SCROLLA symposium as recognition of this fact. A recent OECD report [2, OECD 2001] pays special attention to the interaction between assessment requirements and the undeveloped potential of ICT in learning.

The ITAL Unit (Interactive Technologies in Assessment and Learning) in UCLES (University of Cambridge Local Examinations Syndicate) was set up over 4 years ago, though UCLES was already working on computer-assisted assessment. The formation of the ITAL Unit marked a growing internal realization that significant changes in learning styles were on their way, and that assessment must also change. We take a 'systems approach', in the belief that the whole is greater than the sum of the parts. In order to design a good component we need to understand not just the functioning of the component but its role in the whole system. Our range of activities has therefore been wider than "Interactive Technologies in Assessment". The words "and Learning" in the Unit's name are intended to convey this. We interpret our brief to cover not only the examination itself, but also any related activity that interacts with the assessment process. Such related activities include the use of ICT for test administration, its use in setting questions, in manual and automated marking, support for teachers using ICT related materials in the classroom and the role, use and provision of electronic content.

In a paper of this length we cannot cover every aspect of UCLES' work in ICT related topics. Rather than list many projects sketchily we have chosen to describe a strand of work that covers setting and marking electronically. This has thrown up some central issues and influenced our thinking in a number of key areas.

2. The Enigma Project

2.1 Scope of the Trials

The Enigma Project was a set of trials of online examinations conducted in Singapore in September 1997 and October 1998. The experience gained from these trials has influenced thinking in UCLES on administrative and marking issues as well as on the computer based tests themselves. These trials were initiated in response to the expressed desire of a major customer in Singapore to move towards a more ICT based assessment system as part of their general IT Policy for assessment. The main project implementation was done under contract by the New Zealand based consultancy Aotea Interactive Media (AIM) (See Section 6) In passing, we note that the name 'Enigma' was chosen because initially security issues were seen as a critical problem and it was thought important to demonstrate the use of encipherment. As we will explain, this aspect though important was not found to be the most critical issue in the use of electronic examinations.

In the first trial, the papers of an IGCSE Physics examination of a kind similar to the O Level examinations routinely taken in Singapore were transferred directly into a computer form and so the emphasis of the trial was to consider whether a traditional UCLES paper and pencil test could be administered through a computer. IGCSE is of the same standard as the 16+ GCSEs taken in England, and the question types consist of multiple choice and short answer. Marking was done off-line, with multiple choice items being scored automatically and human marking for anything else. A full evaluation of the trials was carried out [3, Dexter and Massey 1998], comparing performance of paper-based tests with computer-based tests, and exploring administrative and technical issues that arose. Broadly speaking, the first trials showed (as is now widely known) that computer based examining in this context is feasible although there were many technical, administrative and educational issues of detail to be addressed. At this stage there was also a limited trial of questions that involved the marking or manipulation of graphics, etc.

It was also clear that there would be no particular advantage in conducting examinations in this way unless use was made of the new styles of testing that ICT allows. A second set of trials was therefore designed to consist of two components, 'conceptual' and 'analytical'. Questions were at IGCSE level in Physics, Chemistry and Biology. The 'conceptual' component contained a mixture of multiple choice, open ended and manipulation of graphics questions (as in the first trials). The 'analytical' component offered questions that were similar in nature to a Science practical examination except that the "practical" was achieved by expecting the candidates to operate a simulation. This also allowed scientific events to be staged that would not be practicable in a school laboratory in the time (for example, chemical reactions at molecular level, or long term breeding programmes). The second trial did not attempt a comparison with a paper-based form of the tests but it did include a computer based marking interface. Again, a full evaluation was carried out [4, Dexter 1999], and this compared the scores in the Enigma trial with those obtained by each candidate in the corresponding 'real' IGCSE paper that the candidates took in the November 1998. At this second stage, a screen based marking system was introduced, tested and evaluated.

2.2 User Interface

Figure 1 (Appendix A) shows a typical screen from a conceptual component. The interface menu at the bottom allows candidates to move easily between each separate question screen. Where a question runs to more than one screen, this is shown using for instance the convention 2a, 2b, There is a button for each screen with a colour-coded edging to indicate whether the question is current, has been answered, or has yet to be answered. Navigation buttons with arrows work as you would expect to move to the previous or next question, and there are indicators for "Time Remaining" and "Marks Attempted". In the second pilot there were buttons to allow candidates to flag the degree of confidence that they had in their answers, and hence to revisit questions if they had time later in the examination (not shown in the figures). It is not obvious from the screendump, but if the cursor is rolled over any text on the screen, an audio recording is played on the computer's sound system (earphones were used in the trials). It should also be noted that each computer-based 'paper' began with a short audio-visual introduction to the interface.

The evaluation studies used questionnaires to identify problems and successes with the interface. Several issues identified in Stage 1 were addressed by Stage 2. Both stages showed high levels of usability except for the tick and "?" feature, which was not generally understood. Generally the questionnaire returns showed that very few candidates (under 5%) said they disliked computer based examinations, the rest dividing evenly between "OK" and "liked it". Regarding candidates' perceptions of fairness, the most commonly expressed comments were that some felt they could write faster than they could type, that computer literacy was an issue, and that it was too easy to look at others' screens.

2.3 Evaluation – "Traditional" questions

In the first pilot, a sample of students sat a computerized version of an examination, whilst another sample of students took the same examination in paper form. The evaluators concluded that for multiple choice questions there were "no obvious differences between the groups and it would seem that performance on MCQs set in computer and paper based forms might be quite similar" [3, Dexter and Massey 1998]. The evidence from open-ended questions was ambiguous. Students who took the examination on paper did better on open-ended theory questions than did those who took the examination on computer, but performances were roughly similar for open-ended questions about practical skills. The evaluators noted evidence that the students who answered

on computer ran out of time when answering the theory questions, and speculated that their improvement in the practical skills questions was due to them becoming practiced at answering on computer, since those questions were administered last. In the second pilot, no paper-based examination was used, and students sat a specially commissioned computer-based examination, consisting of a “conceptual” component containing traditional style questions and an innovative “analytical” paper (discussed later). Correlations were sought between students’ scores in the pilot on the traditional style conceptual items and their scores on the corresponding ‘real’ O Level paper that the candidates took in November 1998, approximately four weeks after the trial. Correlations between items in the Enigma conceptual paper and the Enigma test total were much higher than correlations between these items and the component total in GCE O level; suggesting that the Enigma questions and O level Science may have been measuring traits which differed to a significant extent. There are several possible explanations for this, including the possibility of differential motivation when comparing a trial to a live examination and the relatively narrow sample in the trial. But it is quite possible that a third variable - perhaps computer aptitude - may have intervened.

Although this is a major issue it did not cause serious concerns because of the many possible non-technological causes of the differences, plus the reassuring results from the first pilot. When the trials are to be developed into a project to convert high stakes IGCSE examinations to computer based forms, then further research will need to be done to follow this up, but no such project has yet been agreed. The reasons for this have more to do with other factors, as the following sections on administrative, technical and general issues explain.

2.4 Evaluation - Analytical questions

Analytical questions were tested in the second pilot in order to see how pupils reacted to new and specifically ICT oriented assessment styles. The range of interaction styles used was not meant to be exhaustive. Certainly at that time the idea of a simulation being used in a test was a novel one, and clearly it was relevant to the ways that ICT might be used to good effect in normal classroom teaching. The trial was meant to provide a starting point for discussion. Six such questions were devised, two each in Physics, Chemistry and Biology. All candidates sat the Physics questions combined with either Chemistry or Biology. The developers, AIM, suggested the basic scenarios, and a panel of examiners devised the question wording.

One analytical question is shown in Figure 2 (Appendix A). A typical feature is that there is a screen for the statement of the question and for the candidate’s responses, and a link to a simulation (in this case the Rabbit Lab, shown in Figure 3 (Appendix A)). In general the candidate’s response might be any of the usual forms of screen based response (such as free text, multiple choice, drag and drop, etc). It is also possible to track what the candidate does with the simulation and present the data gathered to the marker. This is done in this case by generating a short text report that the examiner can scroll through, recording the settings each time the candidate clicks the 'Breed' button.

In the evaluation questionnaire, pupils were asked whether they were able to do the activities requested of them. Despite the unfamiliarity of the test the evaluators were able to report that “generally the results were encouraging”. In fact in only one of the six simulations did large numbers of candidates fail to understand what was expected of them. About 40% of candidates did however report that there was an occasion when they were unsure what to do (this also happens with paper-based tests, of course). The evaluators remarked that this highlighted “the need for examiners to develop question writing skills in writing such analytical style questions”.

The evaluators drew attention to the poor correlation between marks on the analytical questions and marks obtained in the ‘real’ O-Level papers taken later, as happened with the Conceptual papers. There were also poor correlations with the O-Level practical component. As with the Conceptual results, this was not followed up.

2.5 Evaluation - overall student impressions

The evaluators included a number of questions in the questionnaire in order to get some overall impressions. Approximately 40% of students said that answering on a computer slowed them down. This result was interpreted to mean that the students in the trials were unfamiliar with using computers in their schoolwork, and this result highlights the need to consider learning and assessment as an integrated system. A similar percentage felt that the computer’s limitations stopped them answering the way they wanted. One particular

complaint was that some of the answer boxes did not allow enough space to write as much as the candidate wanted. Sometimes the developers had used scrolling text boxes but in some cases scrolling was made not possible in an attempt to limit the amount written. Another common cause of complaint was that in one case it was not allowed to change the first attempt at an experiment.

In the second trials nearly 70% of students said that computer based tests were as fair as conventional tests, and those who disagreed mostly cited typing speed as the main reason, with computer literacy as the next most common reason given. Once again the need for integration between learning and assessment is highlighted.

2.6 Administrative issues

One of the most useful aspects of the Enigma trials was the opportunity to observe how a school's normal IT system interacts with the needs of a computer-based examination. All the schools in the trial had good networked systems but even so there were problems. On a few occasions individual computers crashed (but AIM had provided an incremental answer dumping system that allowed the session to be resumed). There were unpredictable interactions between the Enigma software and installed software: in one or two cases screensaver software prevented access to the Enigma login screen. A high level of expert attention was needed to install and start the examination software and ensure that it ran smoothly throughout the session: certainly computer based examinations will be more demanding on staff than paper based testing.

One security issue that was noted by candidates was the problem that most users could easily see others' screens. If a school has a tightly packed computer room there is no easy answer to this. If computer based testing were to become the norm then it would not be economic to leave alternate computers unused and some physical changes (such as redesigning the rooms or introducing booths or panels) would seem necessary. There are other security issues such as what other software or data is accessible to candidates. Again, there is no simple answer. A working school cannot be expected to uninstall software that is being used for regular teaching or to dedicate large amounts of equipment for assessment purposes.

2.7 Electronic marking

The second Enigma trial generated electronic answer scripts, and included a prototype trial system for returning these scripts electronically and then for assigning them to markers who then marked the scripts on screen. This trial was also evaluated at the time [4, Dexter 1999], but as the experience led to further design and development we will go on to report the most recent trials of the most recent system. The only points to note here from the second Enigma trial are concerned with Internet problems. Firstly, the developer's plan to use a dedicated Internet port to transmit scripts to a remote server could not operate because of the security measures installed in the customer's education network. The lesson here is that such controls cannot be changed lightly. Secondly, only a small fraction of our examiners were sufficiently well connected at home to be able to download scripts over home modems (in these cases the laptops were brought into the ITAL Unit office for script loading). This situation has already changed for the better, but there are still issues of IT support for examiners working at home and of security.

3. Developments

The range of work described indicates the 'holistic' nature of the changes that must take place if ICT is to play a greater role in school education. Looking back from today, we have not seen a surge of computer-delivered examinations in schools because taken altogether the material circumstances in most schools would not permit even a minority of public examinations to take place fairly and securely and still perform their essential function of meaningfully validating pupils' achievements. We can however note that in many countries there has been significant progress towards meeting the conditions that will allow greater use of ICT in assessment. For example, there is a greater level of technical support for IT in schools, of equipment and training for teachers, and in the volume and quality of online learning resources available.

Even satisfying material conditions will not on its own be sufficient to allow ICT to fulfil its potential. Figure 4 (Appendix A) shows how a full cast of influences, processes and agents contribute to 'the learning process'. This model shows that with traditional learning practices it appears easy to box off the summative assessment. The appearance is a little deceptive because there are many traditional links between assessment activities and other components. For example, it is not really Examination Boards that set questions but

examiners, most of whom are or have been practising teachers. Traditional learning methods have been stable, so that although syllabuses and examination structures have changed, the interaction between assessment and learning is treated as part of the background. The introduction of ICT based teaching and learning styles requires a rapid evolution of nearly every aspect of the learning process and therefore requires stronger and more rapid linkages to ensure consistency. The central point here is that for summative assessment to be valid, it must be conducted within the same educational ambience as is the learning. Were it not, the feedback built into the system will either influence the learning process to move towards meeting requirements for success in the final assessment, or put pressure on the assessment system to move towards changed learning styles, or both.

In the case of the influence of ICT, an obvious evolutionary step is for formative and summative assessment processes (boxes P2 and P3 in Figure 4) to use similar techniques and technology, although the roles of these processes remain distinct. (The role of a summative examiner is to provide arm's length external certification of performance at the end of a learning stage, whereas that of the formative test writer is to help during learning.) It is also possible for data collected during the formative stage to be used for certification, as happens already with coursework and more extensively with vocational qualifications. ICT will act as an agent for change and innovation in assessment.

3.1 ICT based resources and educational ambience

In late 1997, with the preliminary results of the first Enigma pilot trials as a guide, the newly formed ITAL Unit had to decide how best to proceed in order to inform thinking within UCLES on these issues. We decided that although the centre of gravity of our activities should be close to the examination processes themselves, we should also initiate activities in related areas. The main related areas we chose were:

- Generic support for teachers in the choice and use of ICT materials in the classroom.
This belongs to box R2 in Figure 4, something that influences almost everything else in the diagram to some degree. We helped sponsor the TEEM Project (Teachers Evaluating Educational Multimedia) by hosting their website [5, TEEM], and advising on and initially implementing the site itself. TEEM has been extremely successful, attracting a substantial DfES grant for its continuation. It lists over 150 evaluated CD-Rom titles and over 100 evaluated websites. All the evaluations are carried out by practising teachers, and the site also gives case studies showing how the materials have been used in class. Our link with TEEM helps to keep us in touch with professional reaction to ICT in the classroom.
- Specific support for teachers relating to UCLES syllabuses.
Syllabuses belong to the resources box R1 in Figure 4, and support is another member of box R2. We have set up and evaluated the use of e-mail lists to build electronic communities of teachers around specific syllabuses. The evaluations show that teachers use the e-lists to exchange answers to frequently asked questions, exchange ideas on approaches to syllabus materials and suitable teaching resources. Although none of this need relate specifically to computer based assessment it helps to raise the general level of ease with new technologies and in due course will become the most obvious and effective way to support the introduction of computer based examinations. In the case of the MEI Mathematics A-Level modules where candidates do use a computer in a high-stakes examination (see below), the use of an e-list proved very helpful.
- Provision and use of electronic learning resources
This belongs to box R1 in Figure 4, and it is hoped directly influences resource authors (A4) and test writers (A2). The amount of time needed to write good simulations means that if simulations are to be used in examinations, then they must be re-usable. There needs to be a range of simulations, around each of which a variety of questions may be set. Candidates also need to be familiar with the use of simulations as a normal part of their learning. We therefore decided to extract the simulations from the original Enigma pilot examinations and set them up as free-standing simulations. CIE (Cambridge International Examinations) decided to run a pilot project with these simulations as learning support materials in relation to CIE's IGCSE examinations in science. There are now six simulations, three from the original Enigma pilots and three newly commissioned ones. Each is accompanied with a variety of suggested activities base on the simulation. [6, CIE]

3.2 Specific Assessment Developments

The educational ambience is important but we focussed most of our resources on continued development of assessment technologies. Some activities began in parallel with the second stage of the Enigma pilots, others were started later.

- **Mathematics Assessment in collaboration with Heriot-Watt**
This activity relates to both formative and summative assessment (boxes P2 and P3 in Figure 4), as well as to the provision of learning resources (box R1) and the skills of test writing (box A2). The CALM project at Heriot-Watt [7, Beevers et al 1999] had successfully developed an online mathematics course for first year engineering students and software that could handle mathematical answers. We entered into a research contract with them to explore the use of this system in the context of UCLES examinations and to develop the system for wider use.
- **Use of computers in MEI Mathematics A-Level modules**
This project relates to nearly every component in Figure 4. The MEI project team wanted to make the syllabus for two topics ('Numerical Analysis' and 'Decision and Discrete Mathematics') reflect the way that mathematics is used in practice, which means that standard mathematical software should be used both during study and then in a "mathematical practical" examination. Only small numbers of candidates enter for these modules (although for the MEI course as a whole the numbers are large) and a teacher prepared to teach for these topics is very likely to be conversant with the software aspects. This meant that provided a suitable syllabus was put in place and that the examining body had suitable procedures for allowing computers to be used in the examination room, the trial could go ahead. With support from the QCA, the new syllabuses were examined for high stakes for the first time in the summer of 2000. This was not a computer-based test because answers were written on paper, and marked in the usual way, but the trials were a very useful test of the practicability of allowing computers into the examination room.
- **Standards for conduct of electronic examinations**
This work relates to the administration of assessments (and so belongs to box P3 in Figure 4). The Enigma project and the experiences just described informed our contribution to the British Standards working party on BS7988 for which the ITAL Unit has provided a representative on behalf of UCLES. The first draft of this standard "A Code of Practice for the use of information technology for the delivery of assessments" was released for public comment in December 2001. [8, BSI 7988]
- **International standards work: IMS and the "Question Test Interoperability" (QTI) standards**
An examinations body is not primarily a software provider and indeed if it were to write its own test delivery system it might well find that many centres would not want to use it because they have already chosen different systems for learning management and formative assessment. An award authority and their examiners are more interested in having a way to set questions that all potential candidates can take. The ITAL Unit has therefore played a full and leading part in the QTI working party, part of the IMS global consortium [9, IMS]. In terms of Figure 4, this activity relates to the technical side of the assessment process (boxes P2 and P3, both formative and summative), and to the provision of e-Learning materials (box R1).
- **On screen marking**
This activity relates to the technical and administrative sides of the assessment process, mainly summative (box P3), but perhaps also formative (box P2) as teachers may well want to mark class work electronically. This work has wider significance than might first be thought because on-screen marking of scanned paper scripts or of electronic scripts provides a bridge technology between the traditional examination style and newer ICT based styles. Because this aspect is rarely discussed in the literature we decided to describe it more fully, and this is done in the next section.

4. On Screen Marking of Scanned Paper Scripts

On screen marking of scanned paper scripts is the process by which paper scripts are scanned (at UCLES or at regional bureaus) and the images transmitted to an image server at UCLES. Copies of these images are then distributed electronically and marked on screen by examiners. Question-level marks and examiners' annotations are captured by UCLES throughout the marking process, without manual intervention.

4.1 What do we want it for?

The system allows UCLES to enhance service to our customers in a number of ways:

- **Faster and more flexible marking and script management processes**

Two factors will work together to help us ensure that scripts are marked in the shortest time possible.

- Improved monitoring lets us track the status of every script throughout the process, allowing tighter management and rapid identification of bottlenecks.
- Dynamic apportionment of scripts to offsite examiners – scripts are only allocated and distributed to an examiner when he or she is ready for them, ensuring that no examiner is waiting for scripts while there are scripts waiting to be marked.
- **More efficient quality control.**
We can more effectively monitor examiner reliability while marking is underway, allowing us to identify and investigate problems at an early stage, when interventions can be made most efficiently. A conceptual diagram of a prototype workflow for marker standardisation and quality control for a typical examination is attached as Appendix B.
- **More detailed feedback to centres about how their students performed.**
Since every candidate's mark on every question is recorded, we may provide information to centres about how well their candidates did on each question compared with other candidates. This may help centres identify strengths and weaknesses in their coverage of a syllabus. Some individual candidates (for example, those considering re-taking an examination) might also be interested in similar information about their individual performances.
- **Smooth transition to online assessment**
Our customers are likely to want a mixture of paper-based and computerised examinations for some time. The paper scanning approach allows us to use the same basic infrastructure to process both types of assessment, facilitating a smooth transition and allowing us maximum flexibility to meet our customers' needs.

4.2 Does it work?

UCLES' first major test of on screen marking of scanned paper scripts was conducted in winter 2000 (though several small scale trials were conducted before this). The main aims of this study were to prove the practical possibility of the scanning and electronic movement of scripts, to provide research data concerning examiner reliability, and to uncover issues, both technical and human, to be investigated in later stages of the development of a production system.

Three November 2000 components were chosen, one each from O Level Mathematics, A Level Geography and A Level English Literature. These components were selected because they covered a variety of question types and script types. Scripts were scanned in Cambridge. After scanning, the paper originals were distributed and marked conventionally¹, and the images were distributed over the Internet to examiners' homes for on screen marking. Two types of on screen marking were investigated: whole script marking, where examiners marked all of a script, and individual question marking, where scripts were split by question and examiners specialised in particular questions². Five examiners for each subject took part in the study, all chosen for their proficiency as examiners and **not** for their IT literacy. They all marked the relevant component operationally (on paper) immediately before doing their screen marking³.

An annotated screenshot of the marking software used by examiners is attached as Appendix C.

In general, the scanning process implemented for the trial worked and the practical viability of scanning scripts and distributing them electronically was demonstrated. Many opportunities for improvement were identified in an issues register, and this has informed subsequent developments. Examiners' impressions and suggestions were captured in a questionnaire, the main findings of which were:

- Examiners could generally schedule downloads such that they had work available when they were ready to mark;
- Most scripts were at least as legible on screen as on paper;

¹ All scripts were marked conventionally, and these were the marks used operationally, since the reliability of on screen marking had not been proven.

² All three components were marked on screen as whole scripts, but only the Mathematics and English Literature components were marked as individual questions.

³ As far as possible, individual examiners only marked a particular script once (i.e. either on paper, or in one of the screen marking modes).

- Most examiners felt that they ended up with the same marks on screen as on paper. Of the three who didn't, two marked English Literature and one marked Geography. The main concerns related to the trial nature of the facilities provided in the prototype software for annotations and scrolling, and to a perceived need to know a script's total mark;
- Mathematics examiners tended to find question marking boring or less rewarding than marking whole scripts;
- All the examiners who took part in the study would mark on screen again.

Examiners' responses to the questionnaire also contained many valuable insights and suggestions for improvements.

Analysis of marks data suggested that:

- For the Mathematics component, examiners applied similar standards and were similarly consistent across the three marking methods (on paper, whole scripts on screen, and individual question marking on screen);
- For Geography, although most of the marking was satisfactory, one examiner was a little more severe when marking on screen and one examiner, whose paper based marking was reasonably consistent, was inconsistent when marking on screen;
- For English Literature, two examiners were a little more severe on screen than on paper (it made little difference whether they were marking whole scripts on screen or individual questions). Examiners tended to be most consistent when marking on paper and least consistent when marking question apportionments (on screen). This may have been because examiners were unable to be influenced by a candidate's performance on other questions when the scripts were split by question.

The results indicated that with suitable modifications to the software used by examiners, screen based marking of whole scanned paper scripts would be likely to be as reliable as conventional marking. Individual question marking required more investigation, particularly for English Literature.

5. Discussion and Conclusion

The three-legged race analogy that we used to introduce this paper is appropriate: we observe the steady increase in the volume of online learning resources coupled with the steady increase in computer hardware and connectivity in schools, and note the resulting pressure on the assessment process to change. We have also noted that assessment is an integral part of the multi-component system that comprises 'the learning process', and that indeed assessment must change if ICT is to be used to its full potential.

As the above experiences show, there may be more than two legs that are tied together in this race. The model of the learning process that we used (Figure 4) shows just how many links exist between different components in the system. Greater use of ICT will affect every one, will change the strength of linkages between components, and may even permit entirely new linkages to form. For example, formative and summative assessment may not remain so distinct, and it is now possible for an award authority to gather data to be taken into account when at the end of a course a certificate of achievement is issued.

One feature of the learning process remains strikingly constant, and that is the central role of teachers. The existence of content in any form does not weaken this. (When was a school or university ever successful by merely telling its students to go and study in the library?). As Figure 4 shows, teachers are the hub of both the learner's experience and the net of influences on nearly every part of the learning system as a whole. We think that all work on ICT in learning should take account of this observation.

Computer based learning is clearly set to play an increasingly important role in education, but this is turning out to be a much more complex matter than many of the pioneers of ICT in learning envisaged. UCLES recognizes the potential, and is playing a full part in assisting and encouraging progress in every part of this process.

6. Acknowledgements

Aotea Interactive Media (AIM) of Auckland, New Zealand, was responsible for the design and implementation of the two Enigma trials, and for supervising the conduct of the trials. We would like to pay a special tribute to the contribution of the late Barry Fenn who was the head of AIM. Barry's creativeness and energy allowed this relatively small-scale project to produce a great deal of useful lessons.

We wish to record our thanks to staff of the Ministry of Education and Schools in Singapore whose willing cooperation made the Enigma trials possible, and to the team of Examiners who adapted their experience with paper examinations so imaginatively.

We also wish to acknowledge the contributions of many colleagues in all branches of UCLES.

7. References

1. "ICT in Assessment: a three-legged race" Patrick Craven and Robert Harding, February 2001, <http://www.ocr.org.uk/news/artcl34.htm>
2. "Schooling for Tomorrow. Learning to Change: ICT in Schools" OECD 2001 (<http://www.oecd.org>)
3. "Evaluation of Enigma computer based examination pilot (7-9 October 1997)", Trevor Dexter and Alf Massey (UCLES/RED internal report, January 1998)
4. "Evaluation of the 2nd Enigma computer based examination pilot (5-8 October 1998)", Trevor Dexter (UCLES/RED internal report, June 1999)
5. The TEEM Project: <http://www.teem.org.uk/>
6. CIE on-line learning support: http://www.cie.org.uk/learning/science_practicals/
7. "The Emerging Philosophy Behind Computer Based Assessment", C E Beevers, D J Fiddes, G R McGuire and M A Youngson IMA J of Teaching Maths and Applics 18 147 - 149 (1999)
8. BS 7988 "'Code of practice for the use of information technology for the delivery of assessments," see <http://www.bsi-global.com/About+BSI/News+Room/exams.xalter>
9. IMS (IMS Global Learning Consortium Inc): <http://www.imsglobal.org/>

Appendix A

Some of the Sun's radiation passes through a prism. The diagram shows the spectrum of this radiation.

Which point on the screen does the infra-red radiation reach?

The diagram shows a blue triangular prism dispersing radiation from the Sun. The resulting spectrum on a screen is labeled with 'not visible violet not visible', 'violet', 'visible', and 'not visible red'. Points A, B, C, and D are marked on the screen with red arrows. Point A is at the red end, B is slightly past it, C is at the violet end, and D is further past the violet end.

Physics | Chemistry

1 2 3 4 5 6 7 8 9 10 11a 11b 12 13 14

Time Remaining 1 hr 29 min 1 mark 4 of 60 Marks Attempted

Figure 1: MCQ screendump from computer-based pilot, 2nd pilot, Conceptual Physics Q1

Design a breeding program that will yield approximately 50% white rabbits.

Use this laboratory to breed at least fifty offspring, about half of which are white.

Arrange a Punnet square with the required gametes, using the letters below, to show the desired proportion of genotypes. You can do this by either clicking and typing, or by just dragging them into place.

The screenshot shows a 'Rabbit Lab' simulation interface. It includes a laboratory setup with rabbits and a Punnet square. The Punnet square has 'b' and 'B' in the top row and 'b' and 'Bb' in the left column. A 'Go to Rabbit Lab' button is visible.

Physics | Biology

3a 3b 3c 3d 4a 4b 4c 4d 4e

Time Remaining 1 hr 29 min 3 marks 3 of 40 Marks Attempted

Figure 2: Analytical question screendump, 2nd pilot, Analytical Biology Q3c

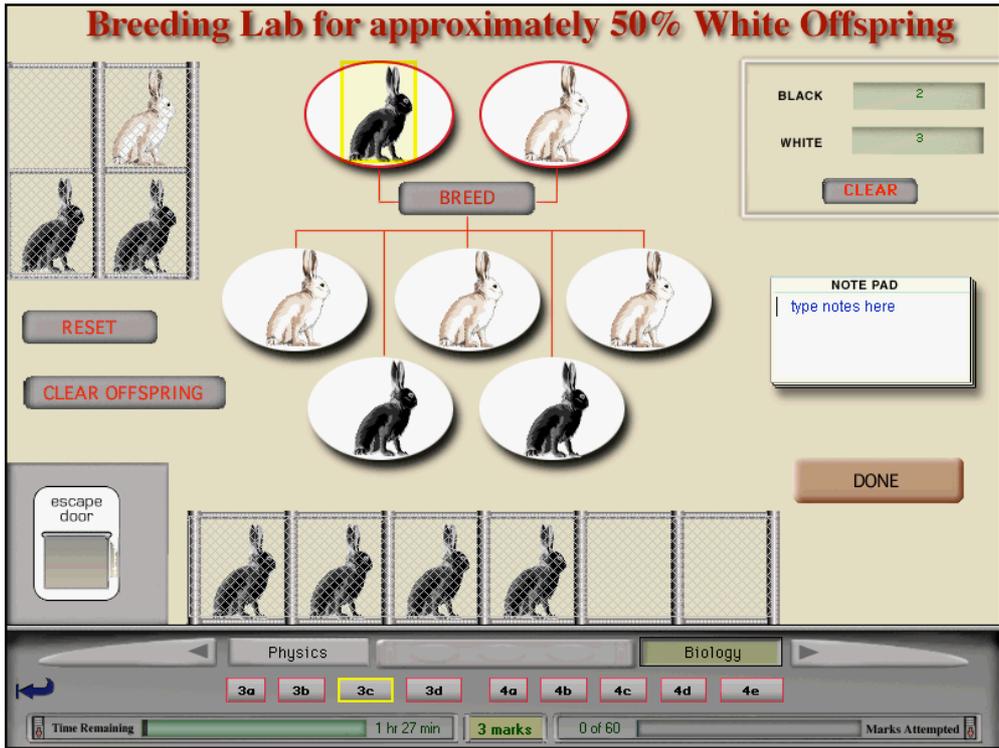


Figure 3: the Rabbit Lab simulation

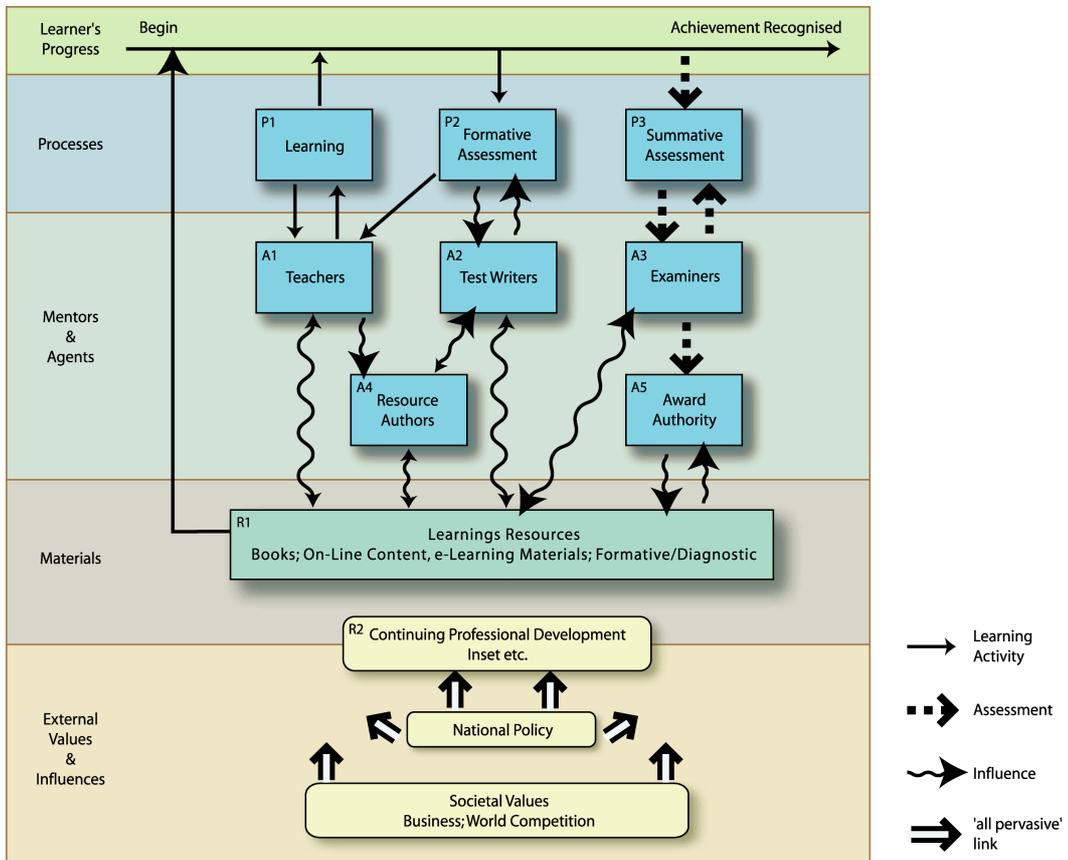
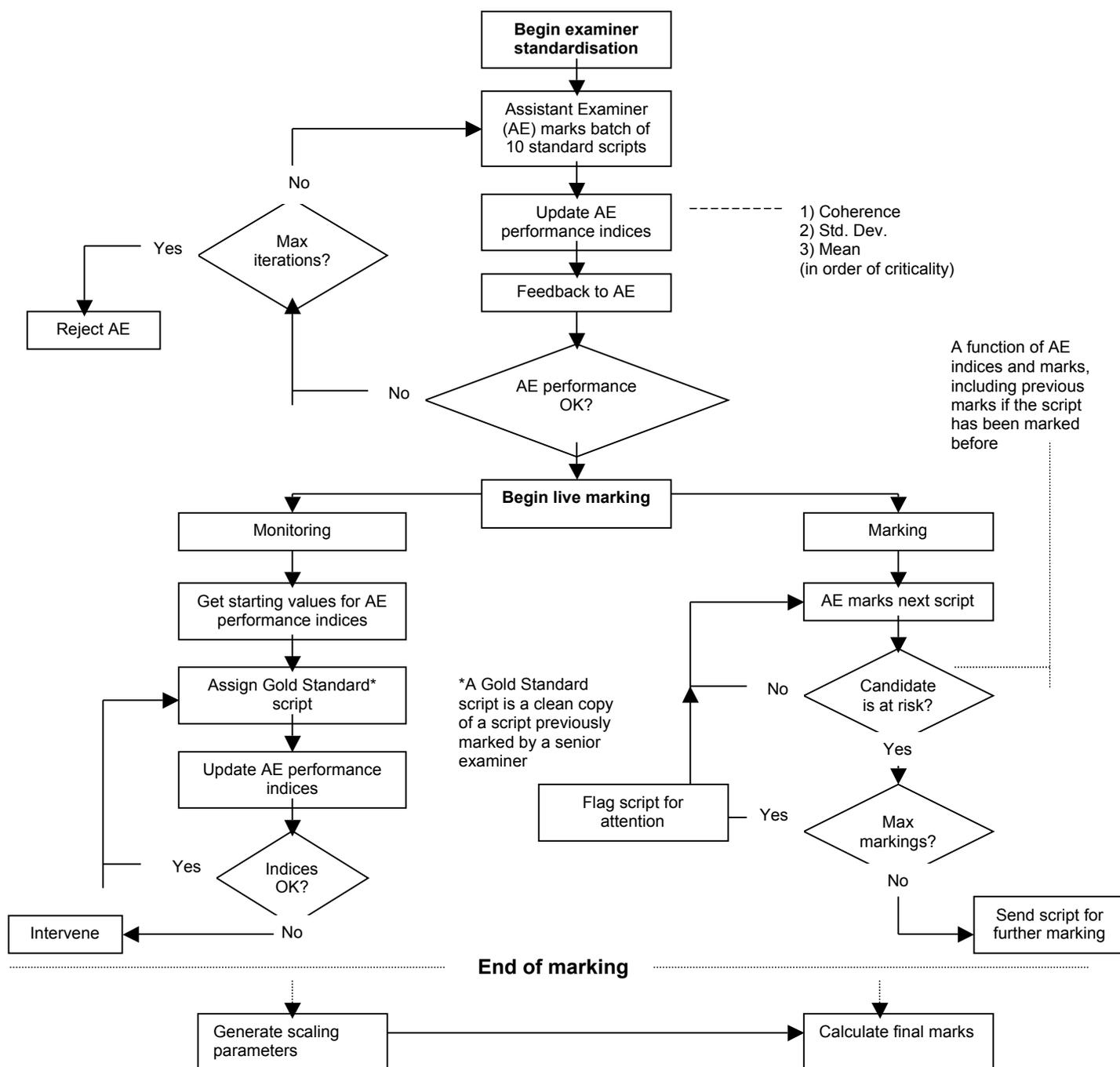


Figure 4: Assessment in Learning diagram

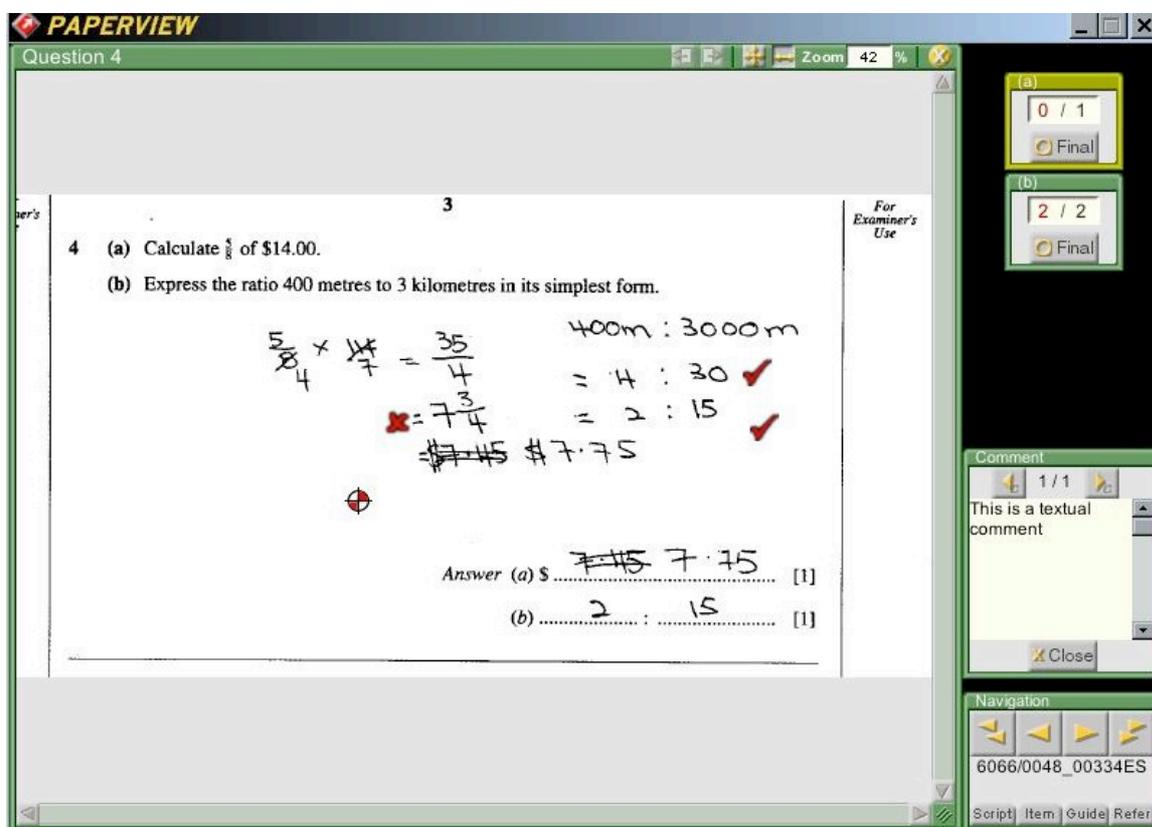
Appendix B

Conceptual diagram of a prototype workflow for examiner standardisation and quality control for a typical examination



Appendix C

Screenshot of the prototype marking software used by examiners



Main window (left hand part of the screen)

This displays the candidate's answer and any ticks and crosses entered by the examiner. The partially shaded circle is the anchor for the examiner-entered textual comment displayed on the right hand side of the screen. The controls in the top right of the window are used for setting display properties and for navigating between pages of multi-page answers.

Navigation box (bottom right of screen)

The outer buttons with the double triangles are used for navigating to the same question in the next or previous script. The inner buttons with the single triangles are used for navigating to the next or previous question in the current script.

The Script button at the left of the bottom row of buttons displays a list of the scripts in the current batch, together with an indication of which ones have been completely marked.

The Item button contains a list of the items in the current script, again with an indication of which ones have been completely marked.

The Guide button displays guidance notes for the examiner on how to mark the currently displayed question.

Examiners use the Refer button to refer a problem script (with a comment) to an administrator or Team Leader.