THE INTRODUCTION OF ELECTRONIC ASSESSMENT INTO ENGINEERING DEGREE PROGRAMMES: OUR EXPERIENCES AND LESSONS LEARNT

A.J. Stocker*, and E.M. Warrington

Department of Engineering, University of Leicester, U.K.

Abstract: The Department of Engineering at the University of Leicester has successfully introduced automatically marked electronic assessment methods into the first and second years of its undergraduate degree programmes. This paper provides a description of the system as it is currently implemented and some of the problems encountered in developing the system. The outcomes from the electronic assessments are comparable to those from written examinations and also allow faster feedback.

Keywords; electronic assessment

**Correspondence to: A.J. Stocker, Department of Engineering, University of Leicester, University Road, Leicester, LE1 7RH, U.K. E-mail: sto@le.ac.uk*

1. INTRODUCTION

Over the last five years, the Department of Engineering at the University of Leicester has implemented a major curriculum review. A significant part of that review has been to introduce the use of automated electronic assessment (e-assessment) to provide

1. faster feedback of marks to the students; and
2. faster, and more consistent, marking

in formal assessments set against a background of a two-fold increase in the number of students over the last few years. As well as fulfilling the above aims, automated e-assessment also enables staff time to be redistributed from marking written examinations to setting assessments and providing feedback. In this paper, automated e-assessment is defined as a 'test' that the students undertake sat at a computer and that, on submission, is automatically marked rather than, for example, a conventional written exam that is then delivered to the human markers electronically. While we have used the Blackboard Learn VLE to implement this assessment, there are a number of alternatives that could also be used (e.g. Moodle).

In changing from traditional written examinations to automated e-assessment, we have encountered a number of challenges, e.g. using appropriate and varied question types, implementation and presentation of the tests, and operational difficulties. This paper provides descriptions of some of our experiences, the lessons we have learnt, and our current position rather than a strict pedagogical approach (although, being engineers, we note that a number of our conclusions are well supported by quantitative analysis).

	Mid 1st semester	End 1st semester	Mid 2nd semester	End 2nd semester
1st year (30 credits)	E/10%/1 hour	E/10%/1 hour	E/10%/1 hour	W/70%/3 hours
2nd year (20 credits)	–	E/30%/2 hours	–	W/70%/3 hours
2nd year variant (20 credits)	–	E/50%/3 hours	–	E/50%/3 hours

Table 1. Assessment pattern in 1st and 2nd year modules. Assessment type/fraction of module marks/duration. E=e-assessment, and W=written examination

2. STRUCTURE OF OUR FIRST AND SECOND YEAR MODULES

The assessment patterns of our first and second year modules are given in Table 1. In both years, the modules run over the entire academic year. We have designed these assessment patterns to give regular summative feedback on performance in the first year, while in the second year there is a reduced need for this and a more substantial assessment is set at the end of the first semester. While we ran two pilot schemes using e-assessment in third year modules, we decided to retain a significant proportion of written assessment since we were initially uncertain as to how well e-assessment would work when deployed across a wider range of modules. The second year variant (taught by the authors), where e-assessment is the only form of assessment, represents a possible future development for all the modules given that we now have significantly more confidence in the outcomes from e-assessment.

The duration of the e-assessments, which is higher per credit than the written examinations, was set because the format of e-assessment, even with what should be a highly computer literate cohort of students, remains more problematic than for a written examination. For example, it is harder to read information off of screen, it is harder to get an overview of the assessment on screen than on paper, and there are more, but smaller scoped, questions in the e-assessment, so more reading time is required. In addition, there is a strong argument for ensuring that the students have plenty of time to complete the assessment since there is merit in being able to work out the answers and not just to rely on knowledge. Although setting the duration initially involved extrapolation from the limited experience provided by the third year modules, we appear to have got this about right since most students complete their attempt of the e-assessments before the time has elapsed.

3. SETTING ELECTRONIC ASSESSMENT

An important point to note is that e-assessment should not be considered synonymous with multiple-choice questioning. While there may be a place for multiple-choice questions, care must be taken to ensure that students cannot readily guess their way to a pass mark (40% in our case). For example, in test comprising 10 multiple-choice questions each with four possible answers, statistically more than 20% of students will be able to obtain a mark of more than 40% by guesswork. The fraction of students able to pass in this way can be reduced by increasing the number of possible answers (e.g. with 10 questions and six possible answers, fewer than 10% of students will obtain a mark of 40% or more by guesswork), by increasing the number of questions (e.g. with four possible answers, increasing the number of questions from 10 to 20 decreases the fraction of students obtaining a mark of 40% by guesswork from over 20% to about 10%), or by allocating negative marks for incorrect answers. For example, if the desired outcome of the test is that a student who guesses should, on average, end up with a mark of zero, then each incorrect answer should

carry a penalty of M/N marks, where M is the mark available for the question and N is the number of possible answers. We have set guidelines that a range of question types must be used and discourage tests that consist solely of multiple-choice questions. Any multiple-choice questions that are included should have at least 5 (preferably 6) possible answers. In this paper, we will not discuss approaches to setting multiple choice questions since this is widely covered elsewhere (e.g. Dickinson, 2012).

Like all assessments, e-assessment still needs to address the intended learning outcomes. However, in many cases, electronic assessment will replace a written examination (although it can also be considered in assessing learning outcomes from laboratory work), so beginning with written examinations previously employed on the module may form a sensible starting point. It is noteworthy that little formal guidance exists at Leicester in setting written examination questions (Bommer and Zdravkovic, 2010 appears to be a rare example elsewhere) so the precise format of written examination questions will vary by module, department, and institution. However, a written examination question in Engineering might typically consist of the following elements:

- Bookwork: This part of a question might ask the students to derive an equation, formally define variables or describe concepts, etc. (i.e. this is a test of knowledge).
- Calculation/design: This part of a question might ask the student to undertake a calculation, optimise a parameter in a system (e.g. the power), i.e. this is a test of mathematical skills, the ability to use appropriate mathematical tools, and design skills.
- Discussion/'sting in the tail': This part of a question will ask the student to consider what the consequence of their calculation or design might be. Examples include, what does the result mean? What effect does this have on your design? How could you improve performance? etc. (i.e. this is a test of whether students can put their knowledge and calculation skills into context and is usually designed to challenge the students).

In the following sections, a variety of approaches to setting electronic questions that mimic or replace the written examination question types are discussed. While we have implemented our tests in Blackboard Learn and have used the naming conventions of that particular VLE, it should be relatively straightforward to apply the principles to other testing environments.

3.1 Bookwork questions

A number of question forms are suitable for testing background knowledge and understanding of material. These include multiple answer, hotspot, matching, and jumbled sentence. In multiple answer type questions, students are given a number of possible answers to choose from where one or more are true. This type of question also allows the assumptions made in deriving an equation, the names and units of variables, the formulae appropriate to different scenarios, etc. to be tested.

Example multiple answer question: For an ideal resistor which of the following is/are true (choose all that apply)?

1. The power dissipated in the resistor does not depend on its resistance
2. The current flowing through the resistor is proportion to the voltage applied across it
3. The power dissipated in the resistor is proportional to the voltage squared
4. The resistance of the resistor will vary with voltage

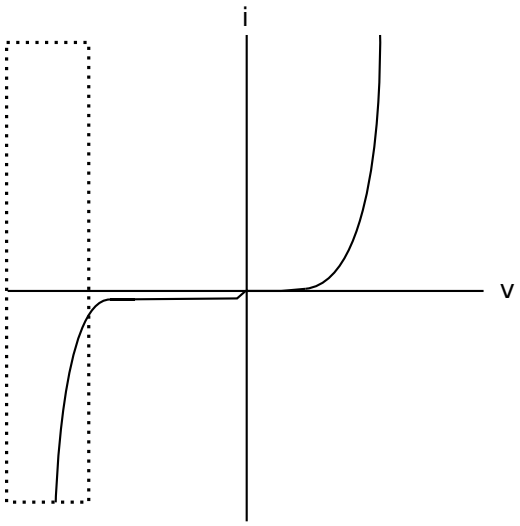

Figure 1. Voltage current characteristics of a diode (note the negative voltage and current scales are not the same as their respective positive scales).

In implementing these questions, we have applied a partial credit system such that students receive 1 mark for each correct answer, but have one mark subtracted for each incorrect answer with the lowest overall mark for the question being set to zero. Since in this case answers 2 and 3 are correct, a student choosing answers 2 and 3 would receive 2 marks (assuming this is the maximum mark for this question), a student choosing just answer three would receive 1 mark, while a student choosing answers 3 and 4 would receive no marks (+1 for answer 3, and -1 for answer 4). This ensures that students guessing at the answers or just ticking all of the answers will, on average, end up with a mark of zero (provided, in the latter case, there are the same or more incorrect answers than correct ones).

In hotspot questions, students are asked to click on the correct location in a graph or picture. For example, in Figure 1, students could be asked to click in the breakdown region of the diode (i.e. the area in the dotted line – this would be defined, but not visible to the students).

In matching questions, students are given a number of statements together with a number of possible matching answers. Note that the same answer can be applied to one or more question parts and that partial credit can be applied.

Example matching question: Match the question and answer parts so that all statements are true.

Question part:
A carrier signal amplitude modulated with a sinusoidal information signal comprises
A carrier signal frequency modulated with a sinusoidal information signal comprises
A carrier signal SSB modulated with a sinusoidal information signal comprises
A carrier signal DSB modulated with a sinusoidal information signal comprises

Answer part:
no frequency components one frequency component two frequency components.
three frequency components more than three frequency components.

In jumbled sentence questions, choices from a list can be inserted at different places in a sentence. Items from the list can be used more than once and partial credit can be applied.

Example jumbled sentence question: Ohm's law states that the [a] flowing through a conductor is [b] to the [c] across its ends.

Possible choices for [a], [b] and [c]:

conductance	current	equal	inversely proportional
potential difference	power	proportional	resistance

3.2 Calculation type questions

Calculation type questions, i.e. one for which there is a numerical answer are relatively straightforward to convert from written examination to e-assessment form. In e-assessment form, the question should clearly state what units the answer are required in and to what precision it should be written (e.g. two decimal places) and the students type their answer into a box. A tolerance can be specified in the answer to allow for rounding errors. Careful thought needs to be given to breaking this type of question down into parts of the appropriate length to replace the 'method' marks that are generally awarded in written examinations. For example, realistically, few students will work their way through to the correct answer in a calculation requiring more than a few stages of algebraic manipulation. While in a written examination marks can be awarded for completing various stages of the solution, in e-assessment this is not possible in a single calculation question and hence the need to break the calculation down into various stages. However, breaking calculations up in too straightforward a way, can 'lead' the students through the calculation and therefore a compromise needs to be found between setting a question that is too easy and setting one that is too challenging. To avoid this, questions can also be asked on various aspects of what the student needs to know in order to undertake a calculation (e.g. as a multiple answer question) as well as the calculation itself.

Example calculation question: A current of 10 mA flows through a 300 Ω resistor, calculate the applied voltage (your answer should be in volts and expressed to two decimal places)
Answer: 3 (±0.01)

3.3 Discussion type questions

While open-ended discussion type questions are difficult to set such that they can be automatically marked, some forms of e-assessment questions can be used to address these areas. For example, multiple answer and matching type questions with a large number of possible answers can test the student's ability to put information into context without significantly reducing the open-ended nature of this type of question. Another possible solution is to set an essay type question although, clearly, this will have to be manually marked and therefore increase the time between the assessment being sat and when feedback is issued. An example in multiple answer form is as follows:

The performance of a radio communication system can be improved by increasing the power. What problems are likely to be introduced (tick all that apply)?

1. Increase in capital cost
2. Increase in running costs
3. Decreased interference with other systems
4. Reduction in data rate
5. Increase in the number of stations required to cover a region

4 STUDENT PREPARATION, ADMINISTRATION, AND FEEDBACK

As previously mentioned, running e-assessment for all of our first and second year students has been a learning experience and one in which we have improved the testing regime with time. In addition, some changes have also been made in response to student feedback (sought formally as free-text comments and through focus groups). There follows a description of the final stage of the system that has evolved.

4.1 Student preparation

Preparation of the students for this type of test is a key aspect of running e-assessment successfully. For example, providing example tests ('quizzes') with questions of a similar nature to those in the formal assessment ensures that students (and academics) obtain experience with the software, the question types, and what happens when things go wrong, e.g. students sometimes submit the assessment before they are ready to since the submit button is next to the save button and changes to the VLE software or browser versions has meant that equations or pictures were not displayed (in one case, the absence of 3-letter extensions from picture files containing equations meant that they were not displayed in Internet Explorer – unfortunately, the assessment had been tested using Firefox). As with any examination, students should also be given appropriate information about the assessment beforehand (e.g. duration, number of questions, choice of questions, balance of subject material in the test, etc.).

4.2 Assessment operation and presentation

Given that our e-assessments form significant portions of a module these are formally invigilated in the same way as written examinations. Our first year assessments are 'open-book' (lecture notes are allowed, but not text books) since they are scheduled at 6-week intervals and we wish to test the students' grasp of the concepts and their ability to undertake calculations rather than recall. The second year assessments are closed book since these are essentially replacing one of the written examinations and this is departmental practice.

The questions are presented in a single scrollable page in a web-browser (strongly preferred by our students) since this allows the students to easily find the questions they want to answer first, or later in the test, those they have yet to answer. The questions can be presented a page at a time if this is required (e.g. some medical conditions may make this a more suitable form). The question order and the answer order within questions are not randomised since this means that the 'paper-trail' associated with the test is more straightforward and errors within the test are easier to detect and correct with Blackboard. However, this does mean that invigilation does have to be stronger since tests with a random order of questions make it harder for students to collude. Although Blackboard allows it, it is also clear from feedback that having different tests for each student (e.g. a different set of values in a calculation – again, this can reduce the effect of collusion) is not popular with the students since they feel that each student has not necessarily had a test of the same difficulty.

One significant problem that we have faced following the significant growth in the number of students during the period of our curriculum review is finding sufficient computers for all of the students to take the test at the same time. Currently, we are one of a small number of departments using e-assessment at the University of Leicester and this problem would become more acute should other departments begin to do so.

4.3 Feedback

One aspect of e-assessment that became clear after widespread implementation was that it generally takes longer to set than a more traditional written examination and therefore it is advantageous if a significant number of questions can be recycled from year to year, although there are benefits to developing a large 'pool' of question variants over time. A consequence of this question re-use policy is that the students cannot be shown their answers to individual questions since this would compromise the reuse of the questions in subsequent years. Students are given their overall mark immediately they submit their test satisfying the aim of providing fast feedback. However, formal written feedback from students has indicated that they would like to receive more information than just the mark. Therefore, we are experimenting with providing a more detailed transcript that can be delivered by email soon after the completion of the test following some post-processing of the results. An example detailed (anonymised) transcript from an actual test is given below.

```
EGXXXX. A.N Other: Overall Mark = 51.1%
                                        Topic     NumQ  Topic Marks
                          Parity and Hamming codes        6         13.9
                            Variable length coding        8         11.7
 Pulse code and pulse amplitude modulation                8          4.2
                             Transmission lines           9         15.0
                                     Waveguides           5          6.4
NumQ is the number of questions you attempted in topic (out of 10)
```

5 RESULTS

5.1 A comparison of performance in e-assessment and written examinations

A comparison of the second year student performance in e-assessment and traditional written examinations is presented in Figure 2 (left hand panel). For each module, every student has two marks – the e-assessment mark (sat in January and worth 30% of the module marks) and a written examination mark (sat in June and worth 70% of the module mark). Having aggregated the written examination marks into bins 5 percentage points wide (i.e. marks of 0–5%, 5–10%, … 95–100%) the associated e-assessment marks have been averaged and plotted against the examination mark. The error bars represent the standard deviations of the

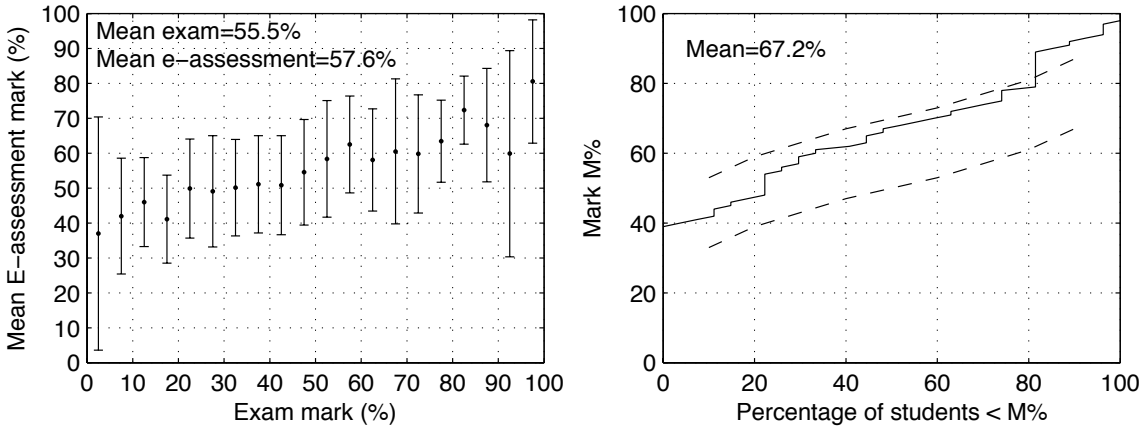

Figure 2. (left panel) A comparison of e-assessment and written examination performance in all second year modules in 2012/13, and (right panel) a cumulative mark distribution for a module assessed solely by e-assessment (the dotted lines represent the recent historical variation in marks for our second year modules).

e-assessment marks in each bin. There are a number of points to note from this figure. Firstly, the e-assessment marks are broadly correlated with the written examination mark and the mean marks of the e-assessment and written papers are within a few percentage points of each other. Secondly, students at the lower end of the class tend to do better in the e-assessment than the exam while the opposite is true for the students towards the top end of the cohort. The first of these points provides strong evidence that our e-assessments are successfully producing outcomes that are comparable with those from our written examinations. However, we currently have no strong evidence to explain the second point, although it may be that the weaker students do better in the e-assessment because they are encouraged by the test format (i.e. no question choice) to undertake revision on the entire module content and that they can then successfully answer the more straightforward questions on each topic. Conversely, while top end students are able to perform excellently on a narrow range of topics in a traditional written examination that includes question choice, they cannot extend this excellence to cover all of the topics examined in the e-assessments.

5.2 Results for a module assessed solely electronically

One of our second year modules is currently assessed entirely using the e-assessment methods described in this paper. The cumulative distribution of marks obtained in 2013/14 is presented in Figure 2 (right hand panel). We note that the students have been given a choice of questions (essentially 30 out of 50) in the two e-assessments undertaken for this module, whereas in the other modules no choice is offered in the e-assessment, although choice is offered in the written examinations. In general, the mark distribution is satisfactorily within the historical variation of marks for our second year modules, although students at the top end (i.e. towards the right hand side of the figure) have performed somewhat better than historically. This may indicate that we have still not quite perfected introducing the more challenging sting in the tail type questions into e-assessment.

6 CONCLUDING REMARKS

The Department of Engineering at the University of Leicester has successfully introduced a significant element of automatically marked e-assessment into the first and second years of its degree programmes that provides outcomes comparable with those from written examinations. While this form of assessment requires some adjustment on the part of both staff and students, it fulfils the aims of providing faster feedback and more consistent marking. While setting e-assessment is more time consuming than the equivalent written examination, this time is offset by considerable reductions in the time spent marking, this being critical in an era when cohort sizes have doubled. The authors and the Department of Engineering would like to acknowledge the patience of our students while we have developed our e-assessment methods and their help in improving these methods.

7 REFERENCES

Bommer J. and Zdravkovic L, Code of practice on setting and marking undergraduate examinations, Imperial College London, 2010.

Dickinson, M., The thing about multiple choice tests, *Learning Solutions Magazine*, March 2012 (available at http://www.learningsolutionsmag.com/articles/861/the-thing-about-multiple-choice-tests-).