OPINION

Embedding e-assessment effectively

Martin Greenhow, Mathematics, Brunel University, Uxbridge, UK. Email: mastmmg@brunel.ac.uk

Abstract

This opinion piece presents my experiences of using e-assessment for foundation year mathematics, analysing data from a 230-strong cohort. Strengths and weaknesses of using e-assessment are discussed, and I present two experiments that have been embedded this year to attempt to counteract poor mathematical skills and study attitudes amongst some students. Data from the last four years show that learning (as measured by end-of-year traditional exams) has been sustained. It seems probable that this is partly due to the efficacy of these measures, but results from the experiments in using staged tests and in appraising foundationers' mathematical confidence are mixed.

Keywords: e-assessment, blended assessment, student confidence.

1. Introduction

At the 2018 EAMS conference, it was clear that there are several versatile and sophisticated e-assessment systems available, and there is a good deal of activity and sharing amongst the e-assessment community within mathematics. Indeed, it seems probable that mathematics is taking the lead here, and it will be interesting to see the extent to which our developments will influence other STEM subjects, and even social sciences like economics that use mathematics, and *vice versa*. I think mathematics e-assessment is now sufficiently mature to make a real difference, especially for today's largely assessment-driven students, but only if we can embed it properly. I draw on our experience of using *maths e.g.*, see Greenhow and Kamavi (2018), with around 1000 Brunel University students *per annum.*, mainly at foundation or first year level in seven academic departments. I think my comments apply to any e-assessment system, and perhaps not just in mathematics either. E-assessment should not be simply about testing students more because we can do so easily: that route is likely to lead to surface-learning (for marks, not understanding, see below) and student stress, taking all the fun out of a module. Moreover, any assessment-heavy module is likely to force lecturers of other modules to do the same, in order to compete for the attention of students. So we need to be smarter and have specific goals in mind for all assessments, which points to a balanced blend of assessments for all modules; e-assessment can take some of the load, allowing staff to assess more synoptically, for example by essays and reports, as well as traditional exams.

The observations and conjectures below are based on our experiences with a combined foundations cohort; those intending to continue with degrees in mathematics or economics take analytical modules in algebra A or B in term 1 (see below), and calculus, statistics and discrete mathematics in term 2; those wanting computing degrees replace calculus and statistics with computing modules. Much, but not all, of the material is comparable to mathematics A-level. However, I think most of the comments apply to any degree containing 'service' mathematics modules, especially in the first year, see Greenhow (2015) for a detailed study of first year economics students.

I also describe two experiments that seek to embed e-assessment in such a way as to enhance not only students' mathematical skills, but also their understanding and even their attitudes to study: perhaps unsurprisingly, the results are mixed, but I feel that even the negative conclusions are worth reporting.

2. Embedding e-assessment into the curriculum and mitigating the effects of cheating

E-assessment can be embedded into a module in a variety of ways, but I have used a low-stakes summative assessment scenario where between 5 and 15 tests are together worth 20% of the module mark. The student's best ever mark from the first five attempts counts towards the e-assessment component. The remaining 80% comes from a traditional unseen written examination at the end of the module, which must be passed at 40% or more. This goes some way to making cheating less attractive to students who may 'cheat' in a variety of ways. Given the mandatory exam pass, I do not object to students working together; indeed I encourage this since a group of, say, three weaker students must do the test at least three times, logged in as each individual group member. That means they get practice on new *realisations* of each question, or very similar ones from the same topic if the choice of *question space*[*] is also randomised. What I do object to is students using 'illegal' software, such as symbolic manipulators or even spreadsheets or web pages, that were not intended by the question author. Ideally, such tools would confer no advantage, but this is likely to rule out the more standard (easier) questions that often give weaker students at least some marks and thus build their confidence. Given such questions are needed for the foundations cohort, some of whom are almost maths-phobic, we may require invigilated sessions, but that can inhibit students' use of practice tests since they 'do not count'. Even more objectionable is the possibility of aliasing, whereby someone else takes the test (unfortunately we have evidence of this being tried but it was heartening that the students themselves reported this). This sort of cheating is akin to essay mills and rent-a-coder web sites, so it is not unique to e-assessment.

3. Scheduling tests – experiment one

To counter the above, and to avoid students waiting until, say, week 10 before starting a suite of tests due in week 12, this year we have set up a set of five fortnightly tests that are invigilated (but with practice tests available beforehand, and the test remains active after the deadline for further practice). The main idea here is that students will discover any gaps in their knowledge and can improve their understanding in tutorials, or from maths support staff, while there is still time. This should have made the end-of-term exams far less stressful, although we have not investigated this. However, throughout the term the experiment seemed to be successful, especially as the invigilated tests were rather informal; students could ask the invigilators to explain/discuss any question (although they do answer it for them, of course). These light touch tests are complemented by the formal end-of-term exam, and are justified by the process of getting students to engage consistently with the module as it progresses; staff from concurrent Algebra A and B modules told me that students attendance was better and they were more active in their tutorials this year. (Algebra A students have A-level mathematics, at any grade, on entry and are considerably stronger than Algebra B students who do not, and who usually have done little maths since GCSE.)

The tests' focus was on learning, not testing. We certainly do not need additional marks beyond those from the formal exam to decide if a student should progress to the next level, i.e. foundation to level 1 or level 1 to 2. In these years, the marks do not contribute to their degree classification,

---

[*] A question space is the set of all realisations produced by the underlying code that drops random parameters into a question at runtime. The question designer must therefore ensure that the realisations are algebraically and pedagogically equivalent, for example by ensuring that random parameter choice does not affect the solution method or level of difficulty of the realisation. This is further discussed in Greenhow (2015).

but for levels 2 and 3, the above relaxed approach may not be possible. Since it is clear that more staff time is needed to set up and run the five invigilated tests, it is pertinent to ask if staging the tests has any benefits in terms of students' perceptions of their learning and the effect on their exam marks.

Students' perceptions of their learning, via e-assessment or otherwise, is clearly important. They 'feel' the tests are doing them good and that they are being treated fairly*, whatever that means. Such perceptions are often gauged by questionnaires, often with very poor participation (especially if delivered in the second half of a module), and sometimes with polarised returns (many 'ok' students may not bother to give their views). We are on shaky ground here, especially as a lecturer engaged in new things, like e-assessment, is likely to confound its effect by being engaging in other aspects of his/her teaching too. So this paper adheres to observed effects, especially those revealed by the last two years data in the exam results. (The exams were considered to be of constant difficulty and content for these two years for each module).

We can now ask who benefits from e-assessment? Do students with poor prior mathematical skills avoid failure, or do already-good students become even better, or neither or both? We could shed light on these sorts of questions by exploring the very large databases generated by e-assessment systems, although the ethics of setting up control groups (possibly previous cohorts without access to e-assessment) are problematic. We also need to develop metrics (beyond the raw exam marks used here) to assess the efficacy of using e-assessment more fully (especially in measuring student confidence). With these metrics in place, and noting that e-assessment generates a lot of data that could be analysed further once we know which questions we should be asking, we might be able to alter the quality, as well as the quantity, of student assessment in a robust and meaningful way.

For now, the results from the exam results for about 80 Algebra A students are encouraging, with an increase from last year's 37% to this year's 52% getting grade A but a small rise from 5% to 8% failing. Tentatively we can say that good students become even better as a result of staging the tests.

For Algebra B students, the measured effect is unfortunately rather depressing: the number getting grade A fell from 25% to 16% (a drop of four students), whilst those failing rose from 16% to 43% (a rise of 31 students). Attribution to any specific cause is confounded by having a larger cohort (84 to 103) of weaker students this year, but it does indicate that students do not seem to have benefitted from the staged tests. Possibly leaving the tests until just before the exam left content fresher in their minds, but it can hardly have added to their understanding of what was taught throughout the term. The same students' failure in one or more of the discrete mathematics, calculus and statistics modules in the next term confirms this. On the basis of this two-year comparison, we can conclude that staging the tests, or not, has little effect on the weaker students' understanding of algebra and has done nothing to alter their study attitudes. I do not have a solution to this, so the next three sections only apply to students with stronger mathematical skills and more mature study attitudes i.e. those who engage with e-assessment.

* Fairness is certainly not obvious since the chosen marking scheme depends on the purpose of the assessment, so that a valid schema may be regarded as unfair: for example, the use of negative marking, described below, can be valid, but worries many students.

4. Flipped learning

I think e-assessment is not just quantitatively different (allowing students repeated practice attempts for example) but, more importantly, *qualitatively* different for the student experience. We should seek to develop their fundamental skills by providing instant, targeted and fulsome feedback that students actually engage with. I used to think that my students would emerge from my lectures 'fully formed' and it would just be a matter of applying what I had taught to the e-assessment. Of course I *taught* it; the problem was that students were unable to *learn* it by just listening to me, or anyone else. Mathematics is, after all, not a spectator sport. The learning happens when they *do* it and, for many, this means doing tests, rather than self-study from other sources (rarely books nowadays, I fear). It is not just about rote practice, although this is beneficial to most, just as practising scales is in music. The learning goes beyond merely that. To explain, I note the comment of a student who had done several (different) realisations of the same question (on completing the square) and suddenly remarked "But I have done this one before!". That student had moved beyond moving the numbers in the question around, to a more abstract understanding of the structure of the solution method. It is now a much shorter step to understanding the general theory, probably from a teacher, but also possibly from the question feedback where it is often given alongside the worked solution of the realisation given (i.e. with the numbers). This is surely where we want students to be, but instead of the usual ordering of "theory then example", more learning may take place when students reverse this and move from the specific to the general. Although it is possible to do such flipped learning by other means, e-assessment greatly facilitates this by allowing students to make the jump when they are ready, although neither the jump nor a feeling of being ready seem to be consciously decided by the student. So this contrasts to a flipped classroom approach where timing is dictated.

5. Learning and teaching processes using e-assessment

Many students take e-assessments primarily for the marks, but in doing so, they learn from the questions and especially the feedback. Most feel comfortable working in groups where they explain to each other, and even end up arguing about mathematics. This probably contrasts with the use of time spent privately doing problem sheets, where the temptation may be to wait for the answers, and then kid themselves they could have done it had they tried. We have often seen students spending a lot of time on feedback screens, going through it line by line to identify where they made their mistake(s). This is in stark contrast to the inevitably-delayed human markers' feedback, where some, perhaps most, students are simply not interested in anything but their mark, and frequently even fail to collect their marked work at all, still less read the feedback. If the learning process described here and in the previous section are more universally true, then why do we bother to set problem sheets at all? The answer may be primarily historical, and it is interesting to ask if paper-based problem sheets could make a case for their introduction if we were already widely using e-assessment. I am not ready to give up on problem sheets yet, but I am happy to test the more mechanistic aspect of 'getting an answer' by e-assessment only. That is not to say that e-assessment is limited to such aspects; indeed it can, and should, go far beyond that. However, I sometimes just teach the basic ideas, some theory and perhaps one example in class, using the saved time for more 'interesting' material e.g. modelling and applications, in other words telling them why I am teaching this material in the first place.

Paradoxically, e-assessment sometimes gives rise a 'reverse' problem of a student not stopping until he/she gets 100%, presumably to gain 'grand wizard' status. This is fine if he/she continues to learn throughout the process, but not if it causes them to neglect other aspects of their studies. I have sometimes intervened to tell them to **stop** doing maths, something I never thought I would do! Another surprise was the student comment "e-assessment provides a bridge between me and the lecturer." We may think face-to-face contact is the gold standard, but this may not be true, especially if it is one lecturer to 200 students and not one to one. How much real contact do we have in traditional settings?

E-assessment can therefore be used to establish accuracy, speed and confidence in students by doing more routine calculations i.e. those for which there is an algorithm and hence those that can be tested objectively. These skills are *necessary*, but not *sufficient*. A particular issue is the use of correct notation; for example, inputting a sequence 1,2,3… and not a set {1,2,3…} as required. E-assessment therefore frees up staff from marking routine material so that they can use their skills to assess higher-level and more subjective processes like modelling, proof, interpretation (and also, for the present, curve sketching, but this may change in the future). My favourite replacement activity is simply to 'drift' round the lab when students are doing e-assessments and talk to students. They are all busy, which gives me time to ask individuals general questions, such as what their strategy is for doing a question, or what the question is asking him/her to do, etc. Unlike in traditional tutorials, here there is nowhere for the student to hide, but at least you are on their side *against it* ("*it*" being the question or even the computer), metaphorically and even physically. You can also ask them to explain the feedback screen, again allowing identification of misconceptions that sometimes do not occur even to an experienced teacher. This doesn't have to last long with each student, but it **is** real contact. Even better, I stand back and get students to explain the feedback to each other; usually the explainer learns more than the other student, and my intervention is rarely required.

6. Using mal-rules

So e-assessment is confined to testing the routine – correct? Well no, but even if it were, there would be nothing wrong with that. It is a reasonable hypothesis that most mistakes are made because of clearly-defined and commonly-used incorrect, but structured, thought processes, that can be encapsulated in *mal-rules*, widely used in e-assessment, see e.g. Walker *et al* (2015). Programming the effects of these mal-rules acting on the question's randomised parameters can pick up, and therefore specifically help, the majority of students; this can run from the trivial ("Your calculator is set to degrees, not radians!") to the more revealing of students' understanding ("You are illegally commuting matrix multiplication." or "This Laplace transform does not exist."). Whilst mal-rules are clearly helpful in designing multi-choice question distracters, they can also be used in responsive numerical input questions, where the values of random parameters can be fed into as many mal-rules as one cares to code; the bottom line (literally) is that if no anticipated mal-rule leads to the students answer, the feedback "I think you may be guessing!" is given, which often surprises students who did guess and wonder how 'it knew'. The whole area of mal-rules is important: if we know which mistakes students make, we might be able to teach better. E-assessment allows the rapid collection and analysis of this data, but the principal sticking point is that we have no adequate taxonomy to group similar errors together in a meaningful way so that they can be acted on. Either these are so general that we cannot act, let alone code algorithms, on them, or they are so specific that they apply to only that particular question or mathematical subtopic. Further research is needed here, but the rewards could be substantial.

Students do not object to such 'traps' being set, and rarely argue about their mark or complain about their treatment. Deadlines are accepted as absolute, and they also accept the need for accuracy – if it asks for two decimal places, then six will not do in some cases (most questions just issue a warning and allow students to correct their input); if it asks for a set, a sequence **is** wrong (as above). A notable exception is if they feel they have been marked wrongly, in which case they will certainly challenge the result, often robustly, even rudely, which I suppose shows some passion for the marks, if not the subject! Sometimes they are correct, since mistakes in coding and ambiguity in question wording can occur. Usually, however, they have not read the question carefully enough, or are certain they are correct when they are in fact wrong i.e. they are deluded. This gives the chance to correct the student, sometimes by re-iteration of a point in a whole-class setting. In any event, students, teachers and question authors all benefit from the insights gained.

7. Confidence questions – experiment 2

Greenhow (2015) discusses different question types and suggests that the humble multi-choice question (MCQ) still has a part to play in building the confidence of students, particularly the maths-phobic. Taking this further, this year I asked the foundations students how confident they are in tackling basic maths problems, see Fig. 1. In the third fortnightly e-assessment, one section was for confidence appraisal, the other for traditional testing of the same types of skills. The idea was to see if the two results correlated, but it became clear that many students simply said they were 'confident' or 'very confident' to everything (even though they knew no marks would be awarded for any response). However, the other test component showed that in many cases the students had no cause to be so confident: perhaps they were deluded or perhaps most students simply saw it as a waste of time, even though their responses were to form an appendix to a related Study Skills module task where a mathematical stock take was required, together with a plan of action. This curious group response was not universal, and some students did consider their abilities more sensibly, as in Fig. 1. This student produced an excellent diagnostic and action plan that he could discuss with the maths support tutor and/or lecturer, so the basic idea is probably sound, but its implementation by some students clearly was not. Although the original experiment backfired, students may have still benefitted from the *process* rather than the *product*.

Topic	Your response
Expand $(-3-x)(4+2x)$	Very Confident
Simplify $3(-3-x)-2(2-3x-x^2-2x^3)$	Very Confident
Expand $(-3-x)(2-3x-x^2-2x^3)$	Very Confident
Expand $(-3-x)^3$	Very Confident
Expand $(-1-2x-3x^2)^2$	Very Confident
Factorise $-12-10x-2x^2$	Confident
Show $(x+4)$ is a factor of $-4-9x-14x^2-3x^3$	Confident
Use polynomial long division on $\dfrac{-2x^3-x^2-3x+2}{-x-3}$	Little Confidence
Complete the square on $-3x^2-2x-1$	Little Confidence
If $f(x)=-3-x$ and $g(x)=-1-2x-3x^2$ find $g(f(x))$	Confident
If $f(x)=4+2x$ sketch the graph of $f^{(2)}(x)$	Some Confidence
State the sum and product of the roots of $-4x^4+x^3-3x^2-x+3=0$	Confident

Figure 1. Confidence appraisal question answered by a foundation student without A-level mathematics. The students were not asked to do the stated question, but rather to think through the strategy and recall knowledge in order to do it, and then state how confident they were that they could carry this through to a solution.

In the spirit of asking students what they think, I did just that to attempt to shed light on the above conundrum. Here are two (slightly edited) responses:

"*The task was useful in the sense of getting direction to where I need to prioritise my revision. I believe it was a good task which helps transition the knowledge gap between my A-level results and the algebra module."*

On following this up, the same student also gave some interesting insights based on his previous A-level Computing project:

"*I found that people don't really choose extreme options on multiple answer-based questions, probably because they do not want to stick out (after all, it seems to be anonymous but obviously the answer files show who is making the responses). Initially I had made the questionnaire Likert-scale responses on paper but with computer delivery I noticed a large shift from negatives to positives - maybe something to think about, probably due to not wanting to stick out in the lab or some form of social manipulation? I did sense some form of 'Survey Fatigue' where I think most people would just end up copying the confident responses – maybe the survey should be given before being taught a specific topic and then after, to compare, or maybe splitting segments of the survey up before each e-assessment test to avoid copying and pasting 'confident' responses. Maybe even swapping the unconfident to confident responses from the top to confident to unconfident? It all rather needs a bit of research."*

If nothing else this shows that this student did think about his mathematical development quite deeply; e-assessment, perhaps done along the lines he suggests might provide a way for staff to understand more about how students think about their studies.

One way of getting students to be more perspicacious may be to use certainty-based questions comprising part a) a normal question, followed by asking them in part b) to give their level of certainty in their part a) answer. A valid marking scheme for students getting part a) correct/incorrect is to increase/decrease the awarded mark according to their level of certainty in that answer, including negative marking for students who are certain that their wrong answer is correct. Anecdotally, students generally dislike, even hate, the idea of negative marking! As mentioned above (footnote 2), the issue of fairness depends on the purpose of the test. Regardless of this, my view is that students should take the time to check that their part a) answer actually 'works' and deserve to be 'punished' if they do not. For example, they sometimes ask staff if their solution to e.g. a differential equation is correct, when they should substitute it into the ODE to see if it really is a solution and really does satisfy the boundary conditions, or, to give another example, to see that an eigenvector $\mathbf{v}$ really is one by evaluating $A\mathbf{v}$. Such action would underline the idea that mathematics has meaning beyond just following an algorithm or procedure to get 'an answer'.

8. Concluding remarks

This opinion piece discusses the potential of e-assessment to enhance the students' learning experiences. Clearly with the large cohort used in this study, any proposals need to be timely and have rapid feedback, which is only feasible using e-assessment. On the basis of the two experiments presented here, it seems students polarise into those who engage with their studies, and those who do not; e-assessment certainly helps the former but does little to affect the latter. That is not to say that e-assessment has no role here, but more seems to be needed to address the somewhat nebulous, but nevertheless extremely important, issues surrounding students' study attitudes. In terms of grades, e-assessment can improve students who are already good, but the above implementation has failed to improve pass rates for the Algebra B module and follow-on modules that rely on understanding its content.

9. References

Greenhow, M. and Kamavi, A. *maths e.g.* student interface Available at: http://www.mathcentre.ac.uk:8081/mathseg/ and *maths e.g.* teacher's interface Available at: http://www.mathcentre.ac.uk:8081/mathsegteacher/teacher.jsp [Accessed 15 Oct 2018].

Greenhow, M., 2015 Effective computer-aided assessment of mathematics; principles, practice and results. Teaching Mathematics and its Applications. 34(3) pp 117-137.

Walker, P., Gwynllyw, D. R. and Henderson, K. L 2015 Diagnosing student errors in e-assessment questions. Teaching Mathematics and its Applications. 34(3) pp 160-170.