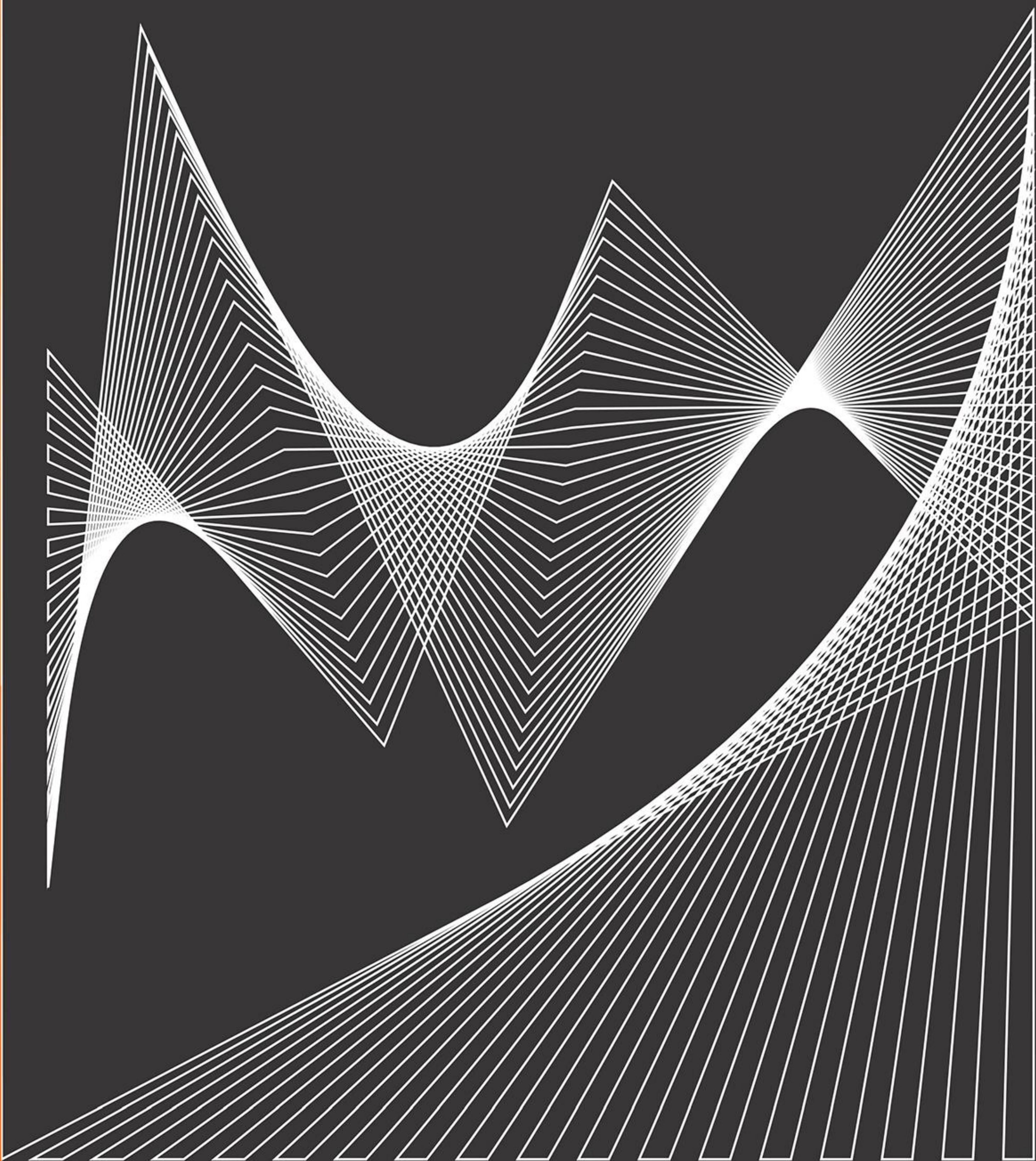
MSOR connections

Articles, case studies and opinion pieces relating to innovative learning, teaching, assessment and support in Mathematics, Statistics and Operational Research in HE.

Volume 17 No. 3



Contents

For information about how to submit an article, notifications of new issues and further information relating to *MSOR Connections*, please visit https://journals.gre.ac.uk/index.php/msor.

This journal is published with the support of the **sigma** network and the Greenwich Maths Centre.

Editorial

Tony Mann, School of Computing and Mathematical Sciences, University of Greenwich, U.K.
Email: a.mann@gre.ac.uk

This issue includes some papers from the recent international conference on E-Assessment in Mathematical Sciences (EAMS), held at Newcastle University on 28 – 30 August 2018. We are grateful to the organisers of this conference, Chris Graham and Christian Lawson-Perfect, for their support with this issue. The case studies by Casey and Crowley and by Kawazoe et al, and the opinion piece by Greenhow, are based on material presented at the EAMS conference.

In addition we present in this issue three research papers. Chingchitnan presents an analysis of an electronic preparatory test for mathematics undergraduates, Lignos writes about the use of data in examining student engagement and planning maths support interventions, and Deshpande's paper looks at using a questionnaire to evaluate financial computing literacy.

We also include an opinion piece by Pfeiffer on a sociolocigal perspective on maths support centres.

We hope that you will enjoy reading this issue of *MSOR Connections* and that you will find much to interest you.

It is a great pleasure to report that Robert Wilson, one of the editors of *MSOR* Connections, has been awarded a National Teaching Fellowship by Advance HE. I am sure all readers of *Connections* will join me in congratulating Rob on this well-deserved award. More information is available at https://www.cardiff.ac.uk/news/view/1547678-national-recognition-for-outstanding-cardiff-lecturers.

Lastly, *MSOR Connections* can only function if the community it serves continues to provide content, so I strongly encourage you to consider writing case studies about your practice, accounts of your research into teaching, learning, assessment and support, and your opinions on issues you face in your work.

Another important way readers can help with the functioning of the journal is by volunteering as a peer reviewer. When you register with the journal website, there is an option to tick to register as a reviewer. It is very helpful if you write something in the 'reviewing interests' box, so that when we are selecting reviewers for a paper we can know what sorts of articles you feel comfortable reviewing. To submit an article or register as a reviewer, just go to http://journals.gre.ac.uk/ and look for *MSOR Connections*.

RESEARCH ARTICLE

Electronic preparatory test for mathematics undergraduates: implementation, results and correlations

Siri Chongchitnan, Department of Physics and Mathematics, University of Hull, United Kingdom.
Email: s.chongchitnan@hull.ac.uk

Abstract

We present a study of the implementation of the Electronic Preparatory Test for beginning undergraduates reading mathematics at the University of Hull. The Test comprises two elements: *diagnostic* and *self-learning*. The diagnostic element identifies gaps in the background knowledge, whilst the self-learning element guides students through an upcoming material. The Test lends itself to an early identification of weak and strong students coming from a wide range of background, allowing follow-ups to be made on a topic-specific basis. The results from the Tests, collected over three years, correlate positively with end-of-year examination results. We show that such a Preparatory Test can be a better predictor of success in the first-year examination in comparison with university entry qualifications alone.

Keywords: Diagnostic; Electronic assessments; Transition.

1. Introduction

Mathematics students starting at most universities come from a wide range of high-school qualifications and grades. In the UK, these qualifications range from A-levels to Scottish Highers and International Baccalaureates, all comprising different syllabi. Therefore, one will naturally find, in beginning undergraduates, a melting pot of abilities and backgrounds in pre-university mathematics. How, then, could university educators ensure that all students have the necessary knowledge to begin the mathematics degree?

In this paper, we will describe our work on the *Electronic Preparatory Test* (EPT) that were designed and used at the University of Hull for undergraduates reading mathematics. The EPT comprises a *diagnostic* element and a *self-learning* element. The diagnostic element ensures that students have a common baseline of background knowledge in mathematics, whilst the *self-learning* element gives a preview of the upcoming new material. Using the EPT, we were able to identify who the weaker and stronger students are, and who might need an early intervention, all within the first week of starting at university.

Pre-semester tests are being used more widely within the HE sector (see for example, Appleby, Samuels & Treasure-Jones (1997), Edwards (1997), Gillard, Levi & Wilson (2010), Sangwin (2015)). Despite many tests being implemented electronically, they often involve simple question types, or are developed on specialist electronic platforms that are not widely used and supported.

The implementation of the EPT at Hull is based on Möbius Assessment (formerly MapleTA). The Tests are online, compatible with mobile devices and graded by the Maple algebraic engine. The EPT comprises an array of interesting question types (including graph sketching and those with infinitely many correct responses). The diagnostic element is partitioned into key topics, involving a wide range of question types. The self-learning element includes an extension of school mathematics, as well as questions on university mathematics (e.g. sets and logic) that are not part of the usual school syllabi. Instead, they examine the students' ability to absorb new ideas and apply them in a self-guided way. Our method of post-Test follow-up by personal tutors will also be of interest to other HE educators.

In this work, we will begin by describing our implementation and pedagogical rationale for the EPT in detail. We will give a sample of various questions in the Test, as well as some useful electronic grading strategies. Finally, we present the results of the Tests, collected over 3 years, and study the correlation with first-year examination results.

2. The Electronic Preparatory Test

2.1 Context

The University of Hull runs a relatively small mathematics department taking in around 40 new undergraduates of mixed abilities each year. Transition from school to university mathematics was identified as a significant problem hindering progression from first to second year, and we therefore looked towards implementing a set of preparatory tests. As a result, the EPT was developed with the help of some of our own students working on creating and refining the problems during summer internships each year.

2.2 Question types

The EPT currently features many question types, including:

- Algebraic response;
- Free numerical response with multiple possible answers;
- Drop-down list (select the correct option);
- Tick boxes (select all correct options);
- Graph sketching.

The EPT comprises a diagnostic element and a self-learning element, which are described in detail below.

2.2 Diagnostic element

Students were asked to complete 5 sets of tests covering the following topics:

- **Algebra**: Binomial expansion, partial fractions, quadratic equations, simultaneous equations;
- **Numbers**: Rational numbers, integer sequences and series, inequalities;
- **Geometry**: Lines and circles in $\mathbb{R}^2$, vectors and geometric transformations;
- **Functions**: Functions and equations involving trigonometry, logarithms and the modulus;
- **Calculus**: Differentiation and integration techniques (up to integration by parts).

The diagnostic element consists of questions that probe the students' background knowledge on these topics. Here are some examples:

Example A (from the Algebra test): Expand $(a-1)^4$.

Comments: To grade the answer, contrary to common grading strategy, it is not enough to ensure that the difference between the input and the answer evaluates to zero (otherwise, the answer $(a-1)^4$ would be marked as correct). Our grading code ensures that only $a^4-4a^3+6a^2-4a+1$ or the permutation of these terms are marked as correct.

The numbers in this and other questions can be randomised, meaning that another student will not necessarily see an identical question.

Example B (from the Calculus test): Integrate the following function with respect to x. Use C for the constant of integration where needed.

$$\frac{4}{x} + 1 + 3x + (2x - 1)^3 + e^{5x} + \cos 2x.$$

Comments: All elements of the expression can be randomised. There is a small penalty for forgetting the constant of integration. The grading code can deal with answers involving permuted terms and alternative trigonometric identities.

Example C (from the Geometry test): Write down the equation (in the form $y = mx + c$) of a line going through $(-9,7)$ and $(5,-7)$. Sketch the line below.

Comments: A set of axes spanning $[-3,3] \times [-3,3]$ is given and sketching can be done by selecting two points on the line. The points can be moved with the cursor to adjust the gradient. The graphics are HTML based (there is no need to install Java or Flash) and the tolerance to the input accuracy can adjusted. Note that neither of the given points can be identified on the given axes, forcing the student to find alternative points on the lines.

2.3 Self-learning element

The self-learning element gives a gentle introduction to new material coming up at university. It comprises the following.

- Extension questions which are part of the five topics discussed in §2.2, but not necessarily part of school mathematics;
- A set of questions on introductory **Logic and Sets** including definitions and simple examples of sets and set operations (∩, ∪), logical implications (⟹, ⟸, ⟺), truth tables (AND, OR, NOT), logical quantifiers (∀, ∃).

Each question has self-learning material, followed by basic questions, and comments on the relevance to the upcoming modules.

Here are some sample questions.

Example D (from the Numbers test): A rational number is any real number that can be expressed as a fraction, a/b, where a and b are integers. A number which is not rational is called irrational.

In your first semester, you will learn a very important proof that $\sqrt{2}$ is an irrational number.

In the list below, which ones are rational numbers?

☐ $\sqrt{8}$	☐ -10	☐ 0.3333 ... (recurring)
☐ $1+\sqrt{2}$	☐ $\frac{3}{\sqrt{2}}$	☐ Any rational number squared

Comments: Irrational numbers are not usually taught in school mathematics. Students will have to deduce the irrationality from the given fact about $\sqrt{2}$. The feedback for this question links to a YouTube video on rational numbers, and mentions the connection to the upcoming module on real analysis. The feedback also alludes to the idea of proof by contradiction, which will also be covered in an upcoming lecture.

Example E (from the Logic and Sets test): Logical implications are statements like:

IF (statement A) THEN (statement B).

For example,

IF (Today is Tuesday) THEN (Tomorrow is Wednesday).

IF (I live in Hull) THEN (I live in England).

We could use an arrow (⟹) to say the same thing:

Today is Tuesday ⟹ Tomorrow is Wednesday.

I live in Hull ⟹ I live in England.

The arrow (⟹) reads “implies that". You could also say it backwards as follows:

Tomorrow is Wednesday ⟸ Today is Tuesday.

I live in England ⟸ I live in Hull.

The arrows could go both ways if the two statements are identical (*logically equivalent*).

Tomorrow is Wednesday ⟺ Today is Tuesday.

But of course, you can live in England without living in Hull…

Now it's your turn. In each pair of statements below, choose the symbol (⟹, ⟸, ⟺ from the drop-down menu) that best relates the two statements.

Let x be a real number.

$x > 12$ _______ $x > 6$

$x \leq 0$ _______ $x^3 \leq 0$

$x^2 = 4$ _______ $x = 2$

$a > 0$ and $b > 0$ _______ $a + b > 0$

$x^2 > 4$ _______ $x > 2$

Comments: Logical implications are not part of school mathematics, but are taught within the first few weeks of the semester. It is important for students to have an early exposure to such a fundamental concept in university mathematics.

Using self-guided questions to introduce important concepts such as logical implication may seem challenging. Nevertheless, we found that the average mark for the Logic and Sets test (all of which are outside the A-level syllabus) is the highest amongst the test topics.

Example C (from the Algebra test): The system of equations

$$3x + 6y = 2$$
$$x + Cy = D$$

has no solution. Give values of C and D which ensure this is the case.

Comments: The grading code accepts infinitely many correct answers. The feedback for this question (regardless of whether the answers are correct) explains the interpretation of the equations as a pair of parallel lines. The feedback also explains the link to the upcoming first-year linear algebra course, in which this situation will be generalised to $\mathbb{R}^3$.

2.4 Rubric and follow-up

At the end of each sub-test, the students are told which answers are correct (with additional feedback, further reading, online resources and links to specific modules). The wrong answers are flagged up along with hints and explanation of common misconceptions. However, the correct answers are not shown in this case.

Students must pass all tests (the pass mark being 75%) by the end of the first week of the semester. They can repeat the tests as many times as they like (correct answers are kept and do not need to be repeated). There is no time limit for each attempt.

After the deadline, results are shared with all lecturers, and failures are flagged up to each student's personal tutor. A student who fails a particular topic will be asked to meet with their tutor to discuss the difficulties they had. They will also need to complete handwritten exercises on that topic to be submitted to the tutor by an agreed date.

3. Pedagogical rationale

Why do we use the EPT?

- The diagnostic element helps identify gaps in the students' background knowledge by the end of the first week of term. This means that early intervention can be administered in the form of electronic feedback and follow-up by personal tutors. Post-test follow-ups and support are essential for the success of such tests (Edwards, 1997).
- The self-learning element serves as a preview of the first-year material, hence giving students the first early exposure to university-level mathematics. It also introduces the student to the practice of pre-lecture reading used by some lecturers as part of the flipped classroom. Learning before lecture has been shown to significantly increase learning gains (Freeman et. al. (2007), Dobson (2008), Moravec et. al. (2017)).
- The EPT helps to sort students of mixed abilities by identifying who have stronger or weaker mathematics backgrounds. This helps tutors to come up with a composition of small tutorial classes that suits their teaching style (e.g. wider or narrower spread of abilities).
- In terms of confidence building, the EPT is the students' first assignment at university. Passing the EPT goes in a long way in reassuring the students that they have come to the right place, are on the right track, and are ready to start the degree. Having increased self-confidence has been associated with better examination performance (Parsons, Croft & Harrison, 2009).
- The EPT clearly sets out the rules of engagement at university. They send a clear message to the students that at university, there are strict deadlines to adhere to, as well as consequences to failing, non-completion and non-engagement.

4. Results

In this section, we shall describe the results of the EPT collected over 3 years (2016-2018) and their correlations with examination results.

4.1 Correlations with examination results

Figure 1. Data showing a positive correlation between the EPT results and average first-year exam marks (r=0.49). The data comprise results from 110 first-year students collected over 3 years (2016-2018).

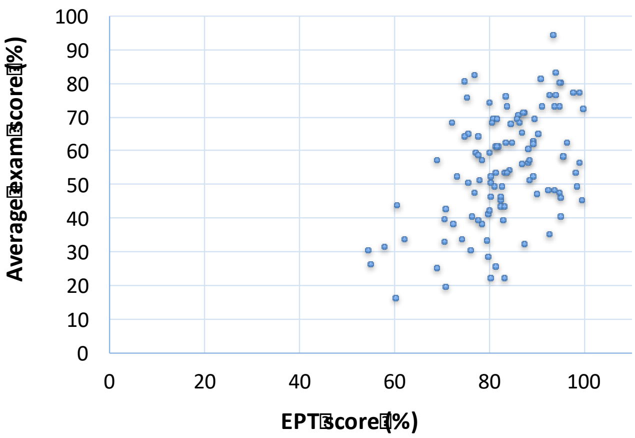


Figure 1 shows a scatter plot of the average end-of-first-year examination scores against the EPT scores for 110 students collected over 3 years (2016-2018). The examination results are calculated by averaging over all mathematics modules. Only first-attempt marks are recorded. We have excluded students who did not take all the examinations (e.g. due to medical reasons or change of course).

We calculate the correlation coefficient, r, defined in the usual way as

$$r_{XY} = \frac{\sum_i (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_i (x_i - \mu_X)^2 \sum_i (y_i - \mu_Y)^2}},$$

where X is the average first-year examination mark and Y is one of the following:

- The EPT score;
- The total entry tariff ('UCAS tariff' points based on the new system introduced in 2017, where grade A*= 56 points etc.);
- The 'maths-only' tariff, calculated using only mathematics and further mathematics grades.

The results of the calculations are shown below.

Table 1. The correlation between the first-year examination results and 3 predictors.

Examination VS	r
EPT score	0.49
Total entry tariff	0.45
`Maths-only'	0.42

4.2 Interpretations

All three indicators in Table 1 correlate positively with the first-year examination results. However, the EPT appears to be the strongest correlator of first-year examination performance, followed by the total entry tariff and, rather surprisingly, the maths-only tariff.

Unlike the EPT score, the total tariff takes into account knowledge of non-mathematical subjects. Points from partial qualifications (e.g. the AS levels) and vocational training also contribute to the total tariffs of many students in our sample, thus weakening its correlation to the mathematical exam performance. Nevertheless, having a high total entry tariff could be an indicator of the ability to handle multiple responsibilities, time management and a good work ethic. Such factors are not taken into account by the maths-only tariffs, and this may explain why it is the worst of the three indicators considered.

Our results echo the findings of Lee, Harrison & Pell (2008) who found that the pre-semester diagnostics were the best predictor for first-year examination results in engineering, as well as those of Yates and James (2006) who found that entry tariffs are an often an unreliable predictor of examination results.

One interesting observation is that, if the EPT were an effective tool in helping students from a wide range of abilities succeed in the examinations, would we not expect to see a *weak* correlation between the EPT results and examination results? In other words, should all students, with the right post-test support, not achieve good exam results regardless of abilities?

To this we argue that it is unreasonable to expect the post-test support to guarantee strong examination results. The factors determining success at university are complex and wide-ranging, and certainly extend beyond background mathematical knowledge alone. Factors such as work ethic, self-confidence, motivation, conducive learning environment, peer support, health and financial stabilities, amongst a myriad of other factors, all play a role in determining success in the examination. The post-test support alone clearly cannot mitigate all these factors.

5. Conclusions and discussion

In this work, we have discussed the use of our Electronic Preparatory Test in mathematics as:

- an important early identification tool for weak and strong abilities, allowing for topic-specific follow-ups;
- a self-guided learn-before-lecture tool to improve learning gains;
- a predictor for first-year success, with a stronger correlation than entry tariffs;
- a first assignment to help students build confidence and understand the importance of engagement at university.

One extension idea is to implement the EPT for all returning students to ensure they retain the necessary knowledge to begin the next stage. In 2017 and 2018, we implemented an advanced version of the EPT for students who were progressing to second year, testing on key first-year ideas in real analysis, calculus, linear algebra and complex numbers. However, the sample size has so far been too small for a meaningful statistical analysis (only one-year's worth of examination results are available at the time of publication). Over the next few years, we will continue to analyse the 2nd-year EPT results and correlations with examination results and, eventually, the final degree classification.

The data for every attempt of the EPT is stored in the Möbius server. The data includes, for instance, how many times each student needed to take each sub-test before they passed and how long each attempt took. This means that it is possible to use such data to produce a better measure of the quality of each student's EPT performance (for example, by weighting the score of each test by the number of attempts needed to pass). Such an aggregate may correlate more strongly with the examination results. Another possible aggregate worth investigating is some weighted combination of the entry tariffs and the EPT results that would maximise the correlation with examination results.

Finally, we are happy to share, upon request, our EPT questions and grading code with other educators in the mathematics HE sector.

Acknowledgements: The author is grateful to Michael Hopton, Elliot Kay and Jake Vinegrad, all of whom are Hull mathematics graduates who helped with the coding and testing of the EPT.

6. References

Appleby, J., Samuels, P. and Treasure-Jones, T., 1997. Diagnosys—a knowledge-based diagnostic test of basic mathematical skills. *Computers & Education*, 28, pp. 113-131. https://doi.org/10.1016/S0360-1315(97)00001-8

Dobson, J. L., 2008. The use of formative online quizzes to enhance class preparation and scores on summative exams. *Advances in Physiology Education*, 32, 4, pp. 297-302. https://doi.org/10.1152/advan.90162.2008

Edwards, P., 1997. Just how effective is the mathematics diagnostic test and follow-up support combination? *Teaching Mathematics and its Applications: An International Journal of the IMA*, 16, pp. 118-121.
https://doi.org/10.1093/teamat/16.3.118

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., Wenderoth, M. P., 2007. Prescribed active learning increases performance in introductory biology. *CBE Life Science Education*, 6, pp. 132-139.
https://doi.org/10.1187/cbe.06-09-0194

Gillard, J., Levi, M. and Wilson, R., 2010. Diagnostic testing at UK universities: an e-mail survey. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 29, pp. 69-75. https://doi.org/10.1093/teamat/hrq004

Lee, S., Harrison, M. C., Pell, G. and Robinson, C. L., 2008. Predicting performance of first year engineering students and the importance of assessment tools therein. *Engineering Education*, 3, pp. 44-51.
https://doi.org/10.11120/ened.2008.03010044

Moravec, M., Williams, A., Aguilar-Roca, A., and O'Dowd, D. K., 2010. Learn before Lecture: A Strategy That Improves Learning Outcomes in a Large Introductory Biology Class. *CBE Life Sciences Education,* 9, 4, pp. 473-481.
https://doi.org/10.1187/cbe.10-04-0063

Parsons, S., Croft, T. and Harrison, M., 2009. Does students' confidence in their ability in mathematics matter? *Teaching Mathematics and its Applications: An International Journal of the IMA*, 28, pp. 53-68.
https://doi.org/10.1093/teamat/hrp010

Sangwin, C., 2015. *Computer Aided Assessment of Mathematics Using STACK*. Cham: Springer International Publishing, pp. 695-713.
https://doi.org/10.1007/978-3-319-17187-6_39

Yates, J. and James, D., 2006. Predicting the 'strugglers': a case-control study of students at Nottingham university medical school. *BMJ*, 332, pp. 1009-1013.
https://doi.org/10.1136/bmj.38730.678310.63

RESEARCH ARTICLE

Using incoming data to monitor engagement and inform mathematics support interventions

Ioannis Lignos, Centre for Student Success, University of East London, United Kingdom.
Email: Lignos@uel.ac.uk

Abstract

Students who do not engage enough with their studies could place themselves at risk of underperforming or failing. Such a risk may be higher for students who are assessed in one or more mathematics modules and lack the appropriate background knowledge, or do not engage enough with related teaching activities. It has been shown for students who engage with mathematics support, there is a significant impact on student performance and progression in the relevant modules. Thus, improving the mechanisms of engagement with mathematics support should be a priority for any student success strategy.

We discuss the monitoring of attendance and performance data of first-year engineering students, as it becomes available, in order to inform interventions which suit the observed student behaviour best. Specifically, the method described was used with first-year engineering students at the University of East London (UEL) during the 2017-8 academic year. We find that when monitoring processes are applied to an already tailored support package, they can often help maintain engagement levels, understand why some students do not engage, and prompt us to differentiate support further.

Keywords: student engagement; data-driven interventions; mathematics support; retention

1 Introduction

One of the most widely discussed issues in Higher Education (HE) is student engagement (Vuori, 2014). Research suggests that student engagement has a causal relationship with student academic performance (Fuller et al., 2011). Student engagement refers to attendance and participation in learning and teaching activities, such as lectures, use of online resources, and academic support sessions. In this article, engagement concerns attendance in mathematics support sessions of any type, from one-to-one appointments and drop-in sessions to small group tutorials and workshops. Although the levels of participation between the attendees may differ, when students attend optional activities, they demonstrate a certain level of active attitude towards their studies, and thus we consider mathematics support attendance as a form of student engagement.

Engagement with mathematics support sessions is increasingly important, given that many undergraduate students nowadays lack the appropriate level of mathematical background in relation to their subject (Pell and Croft, 2008). This often makes mathematics support a positive contribution to student progression and retention (Berry et al., 2015) and academic performance (Mac an Bhaird et al., 2009; Jacob and Ní Fhloinn, 2018). Students who engage with mathematics support tend to do better than those who do not, especially in subjects which include mathematics summative assessments, such as engineering (Lee et al., 2008; Rylands and Shearman, 2018).

A variety of reasons affects engagement with mathematics support. Students being aware of the service and deciding to invest their time are additional challenges for optional teaching activities. Social and emotional reasons and different motivations towards study also affect engagement with mathematics support (Symonds et al., 2008; Mac an Bhaird et al., 2013; Grehan et al., 2016). Another difficulty drawn from our experience is that most students would not commute to the campus to exclusively attend co-curricular activities or use the facilities, unless combined with their academic timetable. This poses an additional challenge in finding appropriate times to offer support sessions.

In this paper, we explore how engagement with mathematics support could be maintained and even enhanced by monitoring the attendance and performance of a cohort of first-year engineering students at UEL in the academic year 2017-8. We are able to identify students who are predicted to be at risk of failing more accurately. Based on emerging trends, we pace and tailor our communications to the cohort and differentiate our support offer to match the circumstances of students who are predicted to struggle with the mathematics module but do not seek for mathematics support.

2 Mathematics Support for First-Year Engineering Students

Mathematics support at UEL is provided by the mathematics tutors of the Centre for Student Success, a student service which offers career coaching, employability and academic tutoring services. The mathematics tutors provide workshops, drop-in sessions and one-to-one appointments across two campuses. When appropriate, part of the provision is a diagnostic test tailored to the academic requirements of the course and often a post-diagnostic workshop, where the results of the test and associated feedback is provided to the students.

The first-year engineering students at UEL are required to pass a credit-bearing engineering mathematics module. In this paper, the cohort under study is comprised of two of the possible programme routes available at UEL. The mathematics component of the module for the specific cohort is taught in compulsory lectures and seminars. During both terms of the academic year (twenty-four weeks), the students attend a mathematics lecture every two weeks, and the associated seminar on the week following the lecture. In 2017-8, the module summative assessments included three mathematics written exams in Weeks 9, 21, and 24. Since 2016 the course leaders have opted in for timetabled, but optional, mathematics support workshops, in addition to the types of mathematics support available by default, such as drop-in sessions and one-to-one appointments.

In 2017-8, the support for this cohort was provided by one mathematics tutor, the author of this paper. A diagnostic test took place during the (compulsory) seminar in Week 2, in order to maximise participation. The questions test prior knowledge of the student at that point in time and without preparation, covering the basics of a broad range of topics of foundation mathematics (for engineers). A series of ten optional timetabled workshops was offered in all even weeks from Week 4 to 20, and one in Week 3, planned with the helpful insight of the academic leaders of the cohort. The workshops appear on the online timetable of every student in the cohort, regardless of how the student performed in the diagnostic test.

In the first workshop (post-diagnostic), the answers and techniques seen in the diagnostic test are discussed. An expected outcome of the diagnostic test and the post-diagnostic workshop is that students have identified gaps in their knowledge and understood the importance of engaging with mathematics support. For the rest of the workshops, the tutor sets the topic to be relevant to the previous lectures, reinforcing relevant skills and highlighting key aspects of the lecture material, but the students present have a say on how the workshop evolves.

Students are encouraged to utilise the rest of the support package, such as drop-in sessions, and individual appointments when they need additional help. Drop-in sessions and individual appointments are the same in structure, where one or more students bring a question to discuss with the tutor at a designated library space. Their difference is that drop-ins run regularly at specific times and days of the week without prior booking, while appointments are bookable and allow for more tutoring time (at least 30 minutes).

2.1. Addressing lack of awareness and availability of the service

It is important to demonstrate how we setup mathematics support such that two important practical reasons for non-engagement are contained as much as possible: lack of awareness of the service (Symonds et al., 2008; Patel and Rossiter, 2009), and support being offered at times not suitable for some students (Mac an Bhaird, 2013).

To raise awareness about our service we visit the students in one of their first lectures, and follow other common marketing methods, such as putting up posters at popular points on campus and sending university-wide emails about our service. In these communications, there is a clear, positive message that mathematics support is for every student, improves understanding and performance, and that friendly and patient tutors are happy to address any question. For the specific cohort of first-year engineering students, perhaps the most effective publicity was that the optional support workshops are timetabled, and thus equally visible to normal lectures and seminars. In addition, a dedicated section on the module VLE page describes the different types of support available.

We also make every effort that our scheduled times for any type of mathematics support are reasonably suitable. Our experience has shown that offering optional timetabled workshops on days when there is no other compulsory lecture or seminar, does not work for most of the students. Besides, that would not have been inclusive towards students who wish to attend but have other commitments allocated for that time. Thus, we look carefully at the student timetable and schedule the workshops on days students must attend other lectures or seminars. Since the cohort schedule in those days is busy, we identify the day with the longest possible break for each of the existing (seminar) groups of the cohort. We timetable more than one iteration of the same workshop such that every seminar group has at least one choice conveniently available. The logic for scheduling drop-in sessions is similar. Since these are open to all students, lunchtime is usually the most appropriate time for most programmes at UEL. In general, we make effort that there is at least one drop-in session which overlaps with one of the (lunch) breaks of every cohort. An easy way to do this is to just extend the duration of the existing drop-in sessions accordingly.

2.2. Data collection

We collect data directly from students who use the service, and merge this with student data held centrally from the student registry. We use the data to investigate which students use the service and how they use it, explore what types of support are requested from which courses, and other similar comparisons which can lead to the improvement of the service. After the introduction of the monitoring process, described in the next section, it is required to collate additional performance data related to module assessments, whether formative or summative assessment (SA), or informal assessment based on interactions with the mathematics tutor.

The data available centrally includes student number, name, programme, and route. Diagnostic testing and SA grades are collected as soon as they become available from the faculty. Furthermore, we gather systematically mathematics support data for every student who attends a support session of any type. This data consists of: student number, type of session, date and main topic of the session. We also note other relevant information or points discussed. For some cohorts, this may include the level of independence of the student, which the tutor evaluates using a three-point rating scale. More details on this type of evaluation are found in the next section, as it was implemented for the engineering cohort.

3 Monitoring Engagement

Monitoring student attendance (or participation) and performance happens unavoidably as part of teaching, for example via assessment and feedback processes. Thus, systematising the monitoring of student engagement with the intention to carry out interventions to increase it, builds on common existing practices in teaching and learning. At the same time, classifying and predicting student behaviour at a detailed level is something which may only be possible using advanced techniques, something which the learning analytics community works on (McKie, 2018).

Perhaps the work most related to a rigorous approach to monitoring engagement involving a mathematics support service is that of Burke et al. (2013). The authors describe the delivery of weekly tutorials to first-year students who are assessed in a mathematics module but not studying towards a mathematics degree. The students were given a set of exercises to work on in groups, asking tutors for advice. They submitted work done in relation to the tutorial, receiving a mark which contributed to the overall module grade. In addition to that, the authors had a dedicated person to monitor who is attending and submitting work. Students who did not engage to a satisfactory level were contacted via email in a multi-stage process, where every emailing round was more urgent in tone, signed by an academic of the department in a respective senior role. The authors report a significant impact of the intervention on the number of students submitting their work and grades.

Although the purpose of the monitoring process we carry out is similar, our process differs in scope and staff perspective. Like Burke et al. (2013), we intervene to engage students who will benefit from attending workshops or other types of support, but our monitoring process does not focus on a specific teaching activity. It tracks the overall engagement of the student with the support service, as an indicator of whether they have addressed their gaps in knowledge. Moreover, the mathematics support tutor did not have the authority to assign work as part of module assessment, and consequently the type and tone of communication interventions differs.

A possible outcome of mapping attendance of support sessions against performance could be better targeted interventions, such as adjusting the timing of service reminders when engagement is low, or sending emails to specific groups when they would be most impactful. Finally, particular student circumstances may become known, which can lead to adapting the type of support offered. In the rest of Section 3, we first outline the basic procedures of our monitoring process, and then we show how we applied those during the academic year 2017-8, by describing specific updates of incoming data and how these led to interventions with the intention to maintain or increase engagement.

3.1. Setting-up the Monitoring Process

We call *support attendance* the level of attendance of any type of mathematics support and *cohort performance* the distribution of students (count) into different performance groups. Each assessment, including the diagnostic test, can generate three performance groups, namely *at-risk*, *not-at-risk*, and *unknown-risk.* If a certain threshold grade is achieved, the student is not-at-risk, otherwise they are at-risk. If the student did not participate in the assessment, and the latter is not summative, then they are placed in the *unknown-risk* group.

The cohort performance can contain more than the three performance groups mentioned, when the results of two or more assessments are combined. When the first SA is available, we subdivide the performance groups by creating four (two-by-two) performance groups, based both on the diagnostic and the SA. Namely, students who remained in the at-risk or not-at-risk group for both the diagnostic and first SA were placed in the *twice-at-risk (T-R)* group or *twice-not-at-risk (T-NR)* groups, respectively. Students who moved from at-risk in the diagnostic to not-at-risk in the first SA or vice versa, they were placed to the *at-risk-to-not-at-risk (R--NR)* or *not-at-risk-to-at-risk (NR-R)* groups, respectively. The cohort performance is initialised at the start of the year based on the diagnostic test. Although the threshold for the diagnostic test would usually be close to the standard pass grade for any assessment (40%), its exact value is determined using heuristics based on some practical reason or observation. If a student does not participate in the diagnostic test, they are classified as unknown-risk, but at-risk if they are repeating the module.

For 2017-8, the diagnostic and first SA were the only major benchmarks, since other assessment data was not available during the teaching period. There was no student who scored between 43% and 50% in the diagnostic test, although the questions were assigned the same weight allowing for a uniform set of possible grades from 0 to 100. Thus, we set the threshold at 45%, following the emerged clustering. There was no obvious clustering for the first SA. We set it at 45%, mainly to allow for a small 'buffer' between the actual pass grade (40%) and the threshold (and for simplicity being equal to the diagnostic test threshold).

We also collect performance data in relation to specific competencies practiced in a session, when possible. For example, is the student able to carry out matrix multiplication? This is done informally through a discussion and observation on a specific competency during any type of a session. The outcome is measured on a simple 3-item scale based on whether the student is able to complete the task independently: 'yes', 'not sure', and 'no'. 'Not sure' means that they needed help in carrying out the steps during the task but show some understanding. This kind of evaluation is not always possible to carry out objectively for every student during a workshop or small group. That is why, we would update the performance status of the student only if it is somewhat clear that what we observed reflects the actual ability of the student at that moment and, for example, has not been based on input from other students. This rule is followed, even if there is a set task with solutions submitted in a systematic way. Since we wish to encourage group work and not to impose exam conditions, it may be the case that most of the solutions on a given task are a result of a collective effort, where knowledgeable students share parts of the solution. Consequently, most of the tutor evaluations are confirmed with an informal discussion with and observation of the student attempting the task on their own.

These evaluations contribute to updating the status of unknown-risk students, or confirm the accuracy of a previous classification. A 'no' evaluation on a competency at a level required by the diagnostic test material (e.g. how to expand brackets is required when solving simultaneous equations), would signal that a student should be in the at-risk group. A 'yes' evaluation on a competency at a level higher or equal than it is required by the diagnostic test material would signal the opposite. (In practice, this means that not knowing lecture material does not necessarily classify a student as at-risk). We shall note that competency-based monitoring of students has also been used by Gallimore and Stewart (2014) in devising 'individualised learning plans' for students.

Thus, for each student there is a record (in the form of a unique row of a spreadsheet) which is updated weekly, capturing both performance and support attendance data, and performance group classification for each assessment.

3.2. Monitoring Incoming Data and Data-driven Interventions

We describe how incoming data in 2017-8 informed our decisions on the type of interventions to carry out, for which students and when. All interventions we list here are data-driven, a result of monitoring the available data. What makes them different than other interactions with students is not their actual composition, but their evidence-based trigger and scope.

We provide a timeline of incoming data, updates on cohort performance and the resulting interventions, as they happened week by week. Figure 1 and Table 1 aid in following the roughly chronological description. A 2-week period starts on an even week, for example the 2-week period labelled as Week 4 includes Week 5. Figure 1 summarises support attendance of the cohort attending at any 2-week period during the year. Specifically, the ‘Any Mathematics Support’ series is support attendance of any type of session, ‘First Time Engaged’ is support attendance by students who have not attended earlier in the year, and ‘Workshop’ is support attendance at the workshop of that week. We call *unique support attendance* the support attendance levels of unique students. Table 1 shows the weekly updates of cohort performance and cumulative unique support attendance by performance group. The cohort performance was initially based on the diagnostic test (Week 2), then it was updated based on the results of the online quiz given during the first workshop (Week 3). Students who did not take the diagnostic test, and we had no other evaluation of their performance during the first workshop were placed in the unknown-risk group.

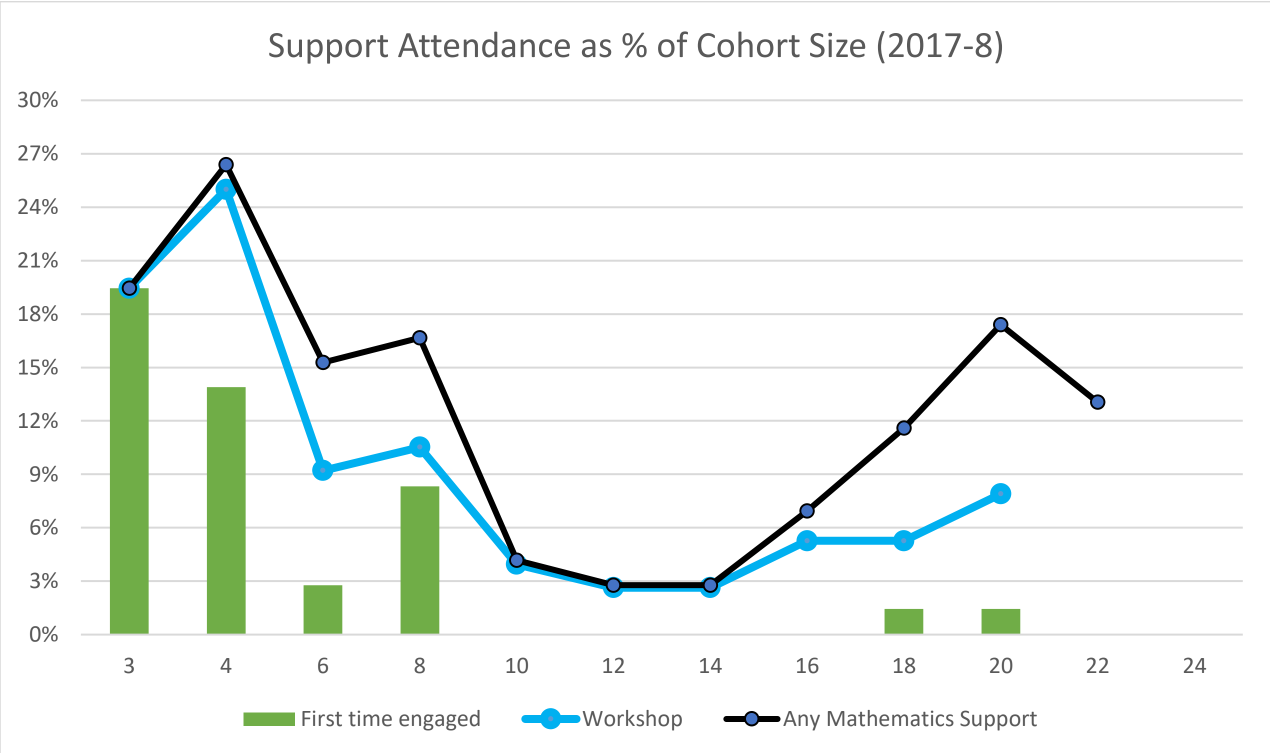


Figure 1. Percentage of the cohort who attended at least once during a 2-week period (starting on the even week) mathematics support, as well as the percentage of unique students engaged for the first time in the year.

Our communication interventions in 2017-8 were messages sent to students’ email inbox, either directly from the email account of the author or via the VLE. Of these, standard reminders were sent to the whole cohort, but the trigger to send one could be the existence of a sizeable at-risk group within. However, more targeted emails were sent to specific performance groups, with the content of the message referring to circumstances and facts only applicable to them. For example, we may have referred to their performance or explained why engaging would have been beneficial, including details of tailored support available. Our teaching-related interventions involved the delivery of types of support which suited a specific group or the whole cohort at a specific time of the academic year, for example working with small groups.

Cumulative Unique Support Attendance by Performance Group until first SA

Performance Group	at-risk	not-at-risk	unknown-risk	Cohort
Week 2 (Diagnostic)	37 students	23 students	12 students	72 students
MS % (Cumulative)				
Week 3*	**21%** (8 / 38)	**25%** (6 / 24)	**0%** (0 / 10)	14, **19%**
Week 4	**39%** (15 / 38)	**38%** (9 / 24)	**0%** (0 / 10)	24, **33%**
Week 5*	**37%** (15 / 41)	**38%** (9 / 24)	**0%** (0 / 7)	24, **33%**
Week 6	**41%** (17 / 41)	**46%** (11 / 24)	**14%** (1 / 7)	29, **40%**
Week 8 (First SA)	**41%** (17 / 41)	**54%** (14 / 24)	**14%** (1 / 7)	31, **43%**
First SA	13 at-risk	56 not-at-risk	0 unknown	^69 students for Term 2
				***performance groups updated ^three students withdrawn**

Table 1. Cumulative unique support attendance by performance group is illustrated for every week. Cohort performance was initialised in Week 2 and updated with the first SA results.

In Week 5, one of the students repeating the module informed us that they did not have the same timetable with first-year students. The support workshop, which they had not attended, was the only session on their timetable on the specific day. Three of the repeating students, previously of uknown-risk, were classified as at-risk. A targetted email was sent to these students to attend a *group study* session, a small group tutorial with time for practice allocated, exclusively arranged for them. Two (25%) of these students responded, attending the session, and engaged with us during the year (7 visits in total). The reason we arranged their own group study, instead of just signposting to the workshops was both to match their availability but also to prevent them from being affected by reasons for non-engagement associated with at-risk students. In Symonds et al. (2008), 10 out of 17 failing, and non-engaged, students felt intimidation for their lack of knowledge.

By the end of Week 5, engagement of at-risk students was at 37%. To increase this rate, we sent one message per week to the cohort in Weeks 5 to 7. Two messages were workshop reminders and one a summary of all the types of support available, including the workshops. As a result, two at-risk and five not-at-risk students engaged for the first time, by the end of Week 8, with engaged at-risk students at 41% of their group.

With the first SA results available, we investigated the performance of non-engaged at-risk students. From the 26 non-engaged at-risk students, 13 students scored at least 60% in the first SA, and all of them below 36% in the diagnostic. The specific group of 13 students demonstrated on average 302% improvement on their performance between diagnostic and summative, while the average for all 60 students who took the diagnostic test improved by 149%. Since the size of the at-risk group is a driver for targetted interventions, we decided it was important to revise the cohort performance based on this evidence, deeming the diagnostic result of the 13 students as mispredicting their actual performance group. Thus, 13 students were classified as not-at-risk for both before and after the SA, and so in the T-NR group. This also revised the rate of engaged at-risk students before the SA to 57%.

Table 2 shows unique students' support attendance (at least one visit) for each of the four performance groups, separately for each of the two periods, but also overall in the year. A useful outcome of this view is to see how much different performance groups engaged before but not after the first SA.

Unique Students' Attendance by Performance Group

Diagnostic Test - First SA	***T-R***	***R-NR***	***T-NR***	***NR-R***	***Cohort***
Up to Week 8*	**9%** (1 / 11) (avg = 3)	**83%** (15 / 18) (avg = 1.9, 2)	**37%** (14 / 38) (avg = 1.5, 1)	**50%** (1 / 2) (avg = 2)	**45%** (31 / 69) (avg = 1.8, 1)
After Week 9	**9%** (1 / 11) ^ (avg = 7)	**39%** (7 / 18) ^ (avg = 5.6, 3)	**13%** (5 / 38) (avg = 1.4, 1)	**0%** (0 / 2)	**17%** (12 / 69) (avg = 3.8, 3)
All Year	**18%** (2 / 11) (avg = 5)	**89%** (16 / 18) (avg = 4.3, 2)	**37%** (14 / 38) (avg = 2, 1.5)	**50%** (1 / 2) (avg = 2)	**48%** (33 / 69) (avg = 3.3, 2)

avg = mean, median visits per engaged student
^ one new student engaged
*using end of year data retrospectively

Table 2. Unique students' engagement by cohort performance group. Each unique student attended at least once in each period separately. The performance groups are based on diagnostic and SA.

The average number of visits per performance group, also shown in Table 2, is what measures how often the engaged students attend. Our goal in maintaining engagement in Term 2 was, practically, that both T-R and R-to-NR students attended regularly, such that they are prepared for the upcoming SAs. Their low diagnostic result had suggested that they may have been potentially at-risk in relation to the upcoming SAs, despite their passing of the first SA. Thus, we sent an email before the first workshop of Term 2 to the whole cohort, referring to the importance of keep attending the support workshops.

One reason for low engagement in Term 2 was the focus on other assessments of the course with mid-term deadlines, happening much earlier than the next mathematics SA of Week 21. We know this from discussions we had with the academic leaders. Until Week 17, there was no new engaged student and general attendance was at the lowest in the year (Figure 1). This prompted us to explore whether the workshop format was still appropriate as the only type of tailored support, or rather encourage students to choose the type of their choosing. Thus, we scheduled bookable group study sessions (groups of maximum 5 - 6 students) to run in addition to the timetabled workshop. The slots offered were chosen according to seminar group availability. Another option would have been to promote appointments, which are also bookable at various times during the week, but this would have limited the number of students attending at their preferred time.

This arrangement seems to have had some distinct contribution to engagement. There was a total of 13 visits from 9 unique students to three booked group study sessions, between Weeks 18 and 23. Four of the students had not attended any workshop in Term 2, with two of them engaging for the first time. In informal feedback they provided they mentioned that they preferred small group or individual type of support to workshops.

In Week 20, we emailed at-risk students (T-R and NR-to-R). We explicitly referred to the fact that they did not pass the first SA, and that we had the intention to help them do better in the next SAs, listing all the available types of support. Within hours, two T-R students booked to attend the upcoming group study session, but they did not show up on the day, despite the automatic reminder sent 1 day before the booking with the option to cancel. Nevertheless, the e-mail intervention dis spark an initial motivation to engage. We sent a further email to the whole cohort after the second SA in Week 22, reminding them of the support available during the period before the final SA took place. Overall attendance improved near the two assessments, reaching the highest rate since Week 4.

3.3. Engagement Summary

In 2017-8, 48% of the cohort engaged with mathematics support at least once, attending one or more of any type of mathematics support. Of those 33 students, 73% attended at least one of the first two workshops, and 94% engaged before the first SA in Week 9. The percentage of the cohort who visited at least two, three and four times was 28%, 19%, and 13% respectively. Students who had been at-risk at some point (R, R-NR, and NR-R) attended more than those who were not-at-risk in any of the terms (T-NR). Of the overall mathematics support attendance in the year, 74% (81 visits) was made by at-risk students. The rates of attendance by each of the three at-risk groups and their average number of visits is seen in Table 2. The highest engagement rate in the year was by R-NR students in Term 1 at 83%. On the other hand, only one of the eleven T-R students engaged with us.

4 Discussion

With a monitoring process of support attendance and performance data in place, a concurrent status of the cohort is available in respect to who engages with mathematics support and how often. Informed by this data, we plan reminders, targeted emails, and differentiate the types of support at the most appropriate time, if and when it is needed.

Although everyone should benefit from support and there is always scope to help not-at-risk students to do better, our interventions tend to focus on at-risk students. Students weaker in mathematics may be less likely to seek for support (Mac an Bhaird et al., 2013). At the same time, those who do not seek for support are also less likely to be motivated (Symonds et al., 2008) than other students. This requires support mechanisms and interventions which motivate those students to engage. In relation to this, in 2017-8, from the 29 students at-risk in Term 1, 15 engaged early by attending one of the two first workshops. Three more engaged later in Term 2, with some evidence suggesting that this was because of appropriate interventions (reminders and the group study initiative). Nevertheless, there were 9 students who according to our records were at-risk until the second SA and did not engage. Three were of those repeating the module who also did not participate or respond to any intervention targeted at them until the end of the year.

As a result of data-driven interventions there were students engaging in good time before the exams in Term 2. Support attendance increased much before the week of the second SA, with less students "cramming" their revision (or support engagement for that matter) in the last minute, as for example we observed in 2014-5 when only drop-in support and appointments were available for specific cohorts (Millwood and Lignos, 2015). Especially in 2016-7 the tailoring of support was mostly similar to 2017-8 with optional timetabled workshops running during the year, but with no particular targetted interventions towards the exam period. It seems that this has had a different effect in Term 2 engagement. In 2016-7, 37.5% of students (6 out of 16) classified as at-risk in the diagnostic test, attended support in Term 2 only in the week of the SA taking place, while for 2017-8 the rate was zero. That is, students attended a mathematics support session for the needs of Term 2 SAs more proactively in 2017-8.

Moreover, data-driven interventions seem to have been effective in maintaining a minimum number of visits of at-risk students, while contributing less to cumulative unique support attendance. On the other hand, most of the unique support attendance was driven by already effective support provision, such as timetabled workshops. The pattern of the 'First Time Engaged' series in Figure 1 shows that most of the students who were aware of the service and decided to engage did so early enough. The cumulative rate of engaged students reached 33% after the first two support workshops, with almost no other type of support attended and no data-driven intervention. This suggests that the timetabling of the optional support workshops leading on from the diagnostic test was the main cause. After Week 4, unique support attendance grew to 43% before the first SA. It is not certain whether our email interventions in Term 1 helped increase unique support attendance (the 10% added between Weeks 4 and 8) or this happened because of a fear of failure in the first SA.

According to Rylands and Shearman (2018), our engagement level of 48% of the cohort is above what they had reviewed in the literature for other mathematics support services (all were below 35%), though smaller than their own rate (58%). Like us, they attributed their high rate to the fact that the support tutorials they offered were timetabled and that three of them were compulsory. It is evident that, in 2017-8, the workshops set out a firm foundation for engaging students at least once, with data-driven interventions potentially boosting that result and securing engagement of more than one visit.

In conclusion, interventions which signpost to types of support which have already reached a high rate of engagement may not significantly increase that rate. Nonetheless, data-driven interventions can help maintain regular attendance and engage hard to reach students. Some of them may be affected by known reasons for non-engagement such as managing their studies and other commitments, low motivation and emotional reasons related to low performance (Symonds et al., 2008; Grehan et al., 2016).

5 Conclusions and Future Work

Using performance and support attendance data allowed us to constantly refine the at-risk status of the student. Knowing the latest levels of engagement and personalised performances, we inform targetted interventions, which can potentially engage new students or maintain the already engaged. We identify and target students who are persistently at-risk throughout the year (T-R) or are not deemed as independent as other students, even when they pass the first SA (R-NR). We did so by sending tailored reminders about the support available or by differentiating the types of support to suit student circumstances.

We initiated such data-driven processes as a monitoring tool of checking who engages with our support, so that we can intervene when especially students at-risk do not engage adequately. An additional purpose was to know which group to target with interventions that could possibly increase or regulate engagement, by for example getting at-risk students to revise for an exam earlier. Some insights would not have been possible without analysing the data as it became available, and thus we may not have taken action promptly.

The type of interventions used could have been carried out without a data-driven process in place. For example, a successful communication strategy and the trial of new types of support can be done without detailed knowledge of the cohort, and yet engage new students. Also, the timing and type of interventions could be determined by experience and interaction with students only. The difference with data-driven interventions is that there is extra evidence, often previously unknown, to support the necessity of carrying them out. Certainly, being proactive in motivating as many students as possible at the right time is worth the effort, even if the outcome is saving just one more student likely to fail, as Pell and Croft (2008) conclude.

Further work could be done on how competencies map between diagnostic test and SAs. This may affect how the cohort performance thresholds are defined. Furthermore, it seems that there may be a sizeable group of the cohort (19% for 2017-8) who do not perform accurately in the diagnostic test. We described a heuristic cohort of identifying those students, specific to the 2017-8. This could be developed into a method which can be applied on any cohort.

Different types of interventions could be added to the ones we have carried out. Calling students for attendance and performance issues is the responsibility of the faculty (or specifically the personal tutor) and other services, such as the specialised retention team. Telephone interventions, organised in coordination with these teams, could include direct referrals to mathematics support. Apart from motivating the student, it would be an additional way of confirming that the student has received our intended message, something we do not know with standard email messaing. Moreover, exploring how academic peer-mentoring schemes can be applied successfully at UEL in relation to the improvement of mathematical skills is another way of reaching out to the non-engaged at-risk students, especially in light of success stories elsewhere (Burke et al., 2012). It would be interesting to see how a peer-mentor could not only engage a student with their studies, inclusive of the mathematics component, but also how they could encourage the mentee in attending mathematics support. Finally, extending the monitoring to all cohorts with assessments in mathematics would be of benefit to a mathematics support service in allocating the usually limited support and intervention resources efficiently.

6 References

Berry E., Mac an Bhaird, C. and O'Shea, A., 2015. Investigating the relationships between the usage of Mathematics Learning Support and performance of at-risk students. *Teaching Mathematics and Its Applications,* 34 (4), pp. 194-204. https://doi.org/10.1093/teamat/hrv005

Burke, G., Mac an Bhaird, C., and O'Shea, A., 2012. Peer mentoring for at-risk students. In P. Rowlett and D. Waller, ed. *Proceedings of the CETL-MSOR Conference 2012.* University of Sheffield, 12-13 July. Birmingham: The Maths, Stats & OR Network. pp.9-14. Available at: http://www.sigma-network.ac.uk/cetl-msor/archive-of-cetl-msor-conference-proceedings-2006-2012/

Burke, G., Mac an Bhaird, C., and O'Shea, A., 2013. The Effect of a Monitoring Scheme on Tutorial Attendance and Assignment Submission. *International Journal of Mathematical Education in Science and Technology*, 44 (4), pp. 545-553. https://doi.org/10.1080/0020739X.2012.756553

Dzator M. and Dzator J., 2018. The impact of mathematics and statistics support at the Academic Learning Centre, Central Queensland University. *Teaching Mathematics and Its* Applcations. Available at: https://doi.org/10.1093/teamat/hry016

Fuller, M., Wilson, M., and Tobin R., 2011. The national survey of student engagement as a predictor of undergraduate GPA: a cross-sectional and longitudinal examination. *Assessment & Evaluation in Higher Education*, 36 (6), pp. 735-748. https://doi.org/10.1080/02602938.2010.488791

Gallimore M. and Stewart, J., 2014. Increasing the impact of mathematics support on aiding student transition in higher education. *Teaching Mathematics and Its Applications,* 33 (2), pp. 98-109. https://doi.org/10.1093/teamat/hru008

Grehan, M., Mac an Bhaird, C. and O'Shea, A., 2016. Investigating students' levels of engagement with mathematics: critical events, motivations, and influences on behaviour. *International Journal of Mathematical Education in Science and Technology,* 47 (1), pp. 1-28. https://doi.org/10.1080/0020739X.2015.1050706

Jacob M. and Ni Fhloinn, E., 2018. A quantitative, longitudinal analysis of the impact of mathematics support in an Irish University. *Teaching Mathematics and Its* Applcations. Available at: https://doi.org/10.1093/teamat/hry012

Lee S., Harrison M.C., Pell, G. and Robinson, C.L., 2008. Predicting performance of first year engineering students and the importance of assessment tools therein. *Engineering Education*, 3(1), pp. 44-51. https://doi.org/10.11120/ened.2008.03010044

Mac an Bhaird, C., Morgan, T. and O'Shea, A., 2009. The impact of the mathematics support centre on the grades of first year students at the National University of Ireland Maynooth. *Teaching Mathematics and Its* Applcations, 28 (3), pp. 117-122. https://doi.org/10.1093/teamat/hrp014

Mac an Bhaird, C., Fitzmaurice, M., Ni Fhloin, E., and O'Sullivan, C., 2013. Student non-engagement with mathematics learning supports. *Teaching Mathematics and Its* Applcations, 32(4), pp.191-205. https://doi.org/10.1093/teamat/hrt018

McKie, A., 2018. UK gets world-first national learning analytics service. [online] Available at: https://www.timeshighereducation.com/news/uk-gets-world-first-national-learning-analytics-service [Accessed 19 November 2018]

Millwood, G. and Lignos, I., 2015. Re-pacing mathematics support to transcend the propensity for "cramming". In: D. Green, ed. *CETL-MSOR Conference Proceedings.* University of Greenwich, London, 8–9 September. Loughborough University: The sigma Network. pp. 66–73. Available at: http:// www.sigma-network.ac.uk/wp-content/uploads/2016/11/CETL-MSOR-2015-Proceedings.pdf

Patel, R. and Rossiter, A., 2009. Student engagement and non-engagement with maths support. In: D. Green, ed. *CETL-MSOR Conference Proceedings.* Open University, Milton Keynes, 7–8 September. The Maths, Stats, and OR Network. pp. 99–104. https://www.heacademy.ac.uk/system/files/proceedings_2009_upload_0.pdf

Pell G. and Croft T., 2008. Mathematics support – support for all? *Teaching Mathematics and Its* Applcations, 27 (3), pp. 167-173. https://doi.org/10.1093/teamat/hrn015

Rylands, L.J. and Shearman, D., 2018. Mathematics learning support and engagement in first year engineering. *International Journal of Mathematical Education in Science and Technology*, 49 (8), pp. 1133-1147. https://doi.org/10.1080/0020739X.2018.1447699

Symonds, R., Lawson, D. and Robinson, C.L., 2008. Promoting student engagement with mathematics support. *Teaching Mathematics and Its* Applcations, 27 (3), pp. 140-149. https://doi.org/10.1093/teamat/hrn011

Vuori, J., 2014. Student engagement: buzzword or fuzzword. *Journal of Higher Education Policy and Management,* 36 (5), pp. 509-519. https://doi.org/10.1080/1360080X.2014.936094

RESEARCH ARTICLE

Development and validation of a questionnaire to measure success in financial computing literacy

A. Deshpande, Ph.D., FHEA, Department of Mathematical Sciences, University of Liverpool, Liverpool, UK. Email: addeshpa@gmail.com

Abstract

An embedded model of teaching financial computing within a course on numerical analysis in finance has been proposed recently in (Deshpande, 2017). It consists of only 10 steps that are aimed at programming beginners. These steps expect students only to be self-motivated to learn. Hence other attributes like pre-knowledge of programming and cleverness aren't expected to influence the learning outcome. Through qualitative assessment via laboratory observation this was indeed found to hold true. In order to understand the outcome of these 10 steps on a much finer scale, we develop here a questionnaire that measures success in financial computing literacy (SFCL) via quantitative assessment. Four scales were developed: self-efficacy or computing confidence, active learning strategy/pro-activeness, learning environment stimulation and an achievement goal in terms of student satisfaction. Findings of this pilot study confirm construct validity of the questionnaire. Importantly we conclude that self-motivation is not enough and that tenacity is a vital component to keep motivation going. Tenacity can be induced via providing credit for attempting steps.

Keywords: questionnaire, construct validation, success in financial computing literarcy.

1. Introduction

Many universities have prioritized the development and the teaching of financial computing courses in response to the increase in recruitment in the financial services sector.This has been done primarily under the umbrella of financial mathematics/engineering. (Kyng et al, 2013) in fact mention that financial services employers rate software skills as important and would like their employees to be more highly skilled in the use of software. A course on financial computing is generally taught separately, before a course on numerical analysis of financial derivatives (or financial numerics). The logic behind this ordering is that the student can then do the financial computing required within financial numerics. However as argued by (Deshpande, 2017) this lacks inclusiveness in terms of a student's ability. Hence, he proposes an embedded inclusive model of introducing financial computing within financial numerics via only 10 steps. These steps as such do not test any logical construct and hence do not need any prior programming background. It has been claimed therein that these steps achieve "financial computing literacy" akin to what "computer literacy" is. It acknowledges that the success in financial computing via 10 steps is primarily due to:

(i) *self-efficacy or computing confidence,* whereby students believe in their own ability to perform well in financial computing, irrespective of whether they had any prior programming experience cf. (Cretchley et al.,1999), and also refer to (Fogarty et al., 2001) and references therein;

(ii) *active learning strategies/pro-activeness*, in which students take an active role in using a variety of strategies to construct new knowledge based on their previous understanding. Note that this scale does not relate to intrinsic self-motivation;

(iii) *learning environment stimulation,* which *encompasses* the in-class learning environment surrounding students, such as the curriculum, teacher's approach to teaching and pupil interaction influencing students' motivation in learning financial computing and

(iv) *student achievement goal or satisfaction.*

These four scales have been borrowed from the science learning questionnaire study of (Tuan, Chin & Shieh, 2005) and (Fogarty et al., 2001). The main aim of that study had been to assess the construct validity of the questionnaire in terms of these scales, that relate to the structure of the 10 steps with students' attitudinal background and concluding with comments in support of success in financial computing literacy (SFCL).

2. Research Design

2.1 Participants

The participants/students were financial mathematics postgraduate students at a UK-based university. These students came from varied backgrounds like engineering, business and mathematics. The SFCL survey questionnaire detailed at the end of this article was handled online. In total 13 out of 14 students (7 males) enrolled in the course completed the survey at the end of conducting the 10 steps/exercises. Though enrollment in master's programs is expectedly small and consequently so is this sample, for this pilot study it is however well representative with almost equal participation by gender and mature students, enough to expect a robust response validating much of what was observed in the qualitative assessment noted in (Deshpande, 2017).

2.2 Materials

A total of 30 questions were developed with regards to four out of the six scales from (Tuan, Chin & Shieh, 2005).

Broadly speaking these refer to:

Self-efficacy: Students believe in their own belief to perform well in computing tasks.

Active learning strategies**:** Students take an active role in using a variety of strategies to construct new knowledge based on their previous understanding.

Learning environment stimulation: In the class, learning environment surrounding students, such as curriculum and pupil interaction influenced students' motivation in science learning.

Achievement goal**:** Students feel satisfaction as they increase their competence and achievement during the computing laboratory exercise.

We note that we do not incorporate the remaining two of the six scales from (Tuan, Chin & Shieh, 2005), viz. *Science learning value* and *Performance goal,* since these are not specific to our domain of study as it chiefly revolves around *motivation in science learning via competition.* This is so because, in our case, students inarguably are aware of the value of financial computing and the structure of our 10 steps do not involve or encourage learning via competition. Additionally, items under the self-efficacy scale have been borrowed from (Fogarty et al., 2001) as it desirably relates to computing confidence. In the scale relating to active learning strategies, we incorporate questions that consider both constructivists learning and deep learning strategies. The latter is important since the first three of the 10 steps that we propose start with stand and deliver style instruction, which (Ramsden, 2003) argues encourages surface learning. Hence, we expect to know how our remaining

7 steps fare in terms of deep learning strategies. All scales except self-efficacy had a mixture of fairly balanced positive and negatively worded statement items. The majority of questions on the self-efficacy scale have a negative orientation, thus reflecting the primary concern of most educators, which is a possible handicapping effect of negative attitudes towards computers and mathematics (Fogarty et al., 2001). Also, the questions were worded so that the respondents think about the statements rather than respond automatically. It also minimizes the effect of a response set towards either agreement or disagreement with whatever statement is made (Moser and Kalton, 1980). All items employed a Likert-style response format, with options ranging from 5 (Strongly agree), 4 (Agree), 3 (Neutral), 2 (Disagree), to 1 (Strongly disagree).

3. Results

The ability of the questionnaire to individually differentiate between motivation levels vis-à-vis the scales was measured using one-way analysis of variance (cf. Table 1). Here data is presented in mean (standard deviation) format. As seen from Table 1, low (L) motivation level students performed below those categorized with High (H) and Moderate (M) motivation levels. All these levels are significantly different when information of all items on every scale of the questionnaire are collated as seen in the last column (p<0.002). Self motivation levels govern outcome in the self-efficacy scale (p<0.002) as expected. Success in the SFCL as captured in the achievement goal (which is same for H, M and less for L) indicates that self-motivation plays a significant role in the outcome of SFCL. The M level students experience positive influence from the learning environment in comparison to H and L levels. This is expected since H level students are observed to be mostly self-focussed and learning environment stimulation had not had much impact on them, while L level students sitting mainly close together with like-minded L colleagues did not benefit much from the discussion, they had within themselves during the SFCL exercise. Similar reasoning can be reported under the scale of Active learning strategy. Though the numbers for H and L are encouraging they are lower than M type students. Note that the associated F ratios for these two scales aren't statistically significant in comparision to the remaining scales but are in reality close to our observation related to attitudinal difference that was also reported in (Weimar, 2013). Another belief is whether these two scales could be viewed as a single construct, since an active learning strategy is aided by the learning environment stimulation, and it is hard to detect perception to environment stimulation within a short span such as the one semester over which the SFCL is conducted. This is in fact reinforced by smaller values of the partial η^2 in Table 2 (viz. 0.14 and 0.17). However, we think that even transient emotional reaction to the learning environment stimulation scale may itself strengthen and deepen into firm attitudes if reinforced (Mandler, 1984). Hence, we will keep these two scales.

Table 1. One-way analysis of variance of high, medium and low-motivation level towards responses on the SFCL questionnaire.

Motivation	Self-efficacy	Learning environ. stim.	Achievement goal	Active learning strategy	Total Qns.
High	39.00 (0.00)	10.00 (0.00)	16.00 (0.00)	21.00 (0.00)	11.07 (0.00)
Moderate	27.50 (3.50)	12.00 (0.00)	16.00 (1.41)	23.00 (1.41)	10.76 (0.16)
Low	24.22 (4.11)	9.88 (2.08)	11.88 (1.69)	20.22 (2.99)	8.90 (0.82)
F	12.08 p<0.002	1.063 p=0.381	9.40 p=0.005	0.867 p=0.45	11.62 p=0.002

The results in table 2 discuss internal reliability of the questionnaire. Cronbach's alpha reliability coefficient for each scale, using an individual student as the unit of analysis, ranged between 0.24 and 0.85 and were estimated to be in general satisfactory. The discriminative validity or mean correlation with other scales refers to the extension to which each scale measured a dimension different from that measured by any other scale. In the SFCL the validity was close knit and ranged from 0.16 to 0.37, showing some overlapping with other scales. This feature was also captured as dicussed following Table 1 and only gets reinforced by the partial η^2scores, which ranged from 0.14 to 0.70 in Table 2. Two scales viz. achievement goal and self-efficacy enjoyed the lion's share of the proportion of this measure which was 0.65 and 0.70 respectively, in contrast to the remaining two scales.

Table 2. Internal consistency (Cronbach alpha coefficient) and discriminative validity (mean correlation with other scales).

Scale	Item no.	Mean	Std. dev.	Cronbach alpha		Partial analysis of var. (η^2)	Mean correlation with other scales
				Individual	Class mean		
Self-efficacy	Q1-Q12	27.00	2.63	0.85	0.80	0.70	0.24
Active learning strategy	Q13-Q20	20.76	2.68	0.46	0.42	0.14	0.24
Learning environ. stimulation	Q21-Q25	10.23	1.87	0.24	0.27	0.17	0.16
Achievement goal	Q26-Q30	13.15	2.44	0.62	0.59	0.65	0.37
Total Qns	Q1-Q30	71.15	9.56	0.81	0.80	0.70	---

The pattern matrix (see Table 3) derived from the factor analysis performed using principle axis factoring with oblique (oblimin) rotation provides 8 factors/ components. In Table 3, all components except the last one has at least one item from either self-efficacy and or achievement goal. The 8th component has item 18 and item 19 (see Appendix for details) worded similarly, but one in a positive spirit while the other is in a negative spirit to discount biasedness in response. Hence in general self-efficacy and achievement goal are significant over the other two scales. We emphasized the role of "self-motivation" in SFCL (Deshpande, 2017). However, based on the outcome of this exercise, we infer that one also needs the less emphasized learning environment stimulation to instigate "pro-activeness/active learning strategy" for reaping major benefits of SFCL. Hence one could induce it by providing contributing credit towards all the 10 steps or maybe first few steps, one for the middle step i.e. Exercise 6 and one at the end viz. Exercise 10. This will likely enhance the significance of active learning strategy through tenacity.

Table 3. Factor pattern matrix of items in SFCL questionnaire (n=13)

Pattern Matrix[a]								
	Component							
	1	2	3	4	5	6	7	8
SelfeffQ11	0.998							
SelfeffQ12	0.881							
SelfeffQ4	0.757							
LearnenvirstimQ22	-0.736							
SelfeffQ9	0.627							
SelfeffQ2		0.879						
ActLearnSratQ14		0.825						
ActLearnSratQ13		0.785						
LearnenvirstimQ25			0.938					
SelfeffQ8			0.756					
LearnenvirstimQ24			0.608					
LearnenvirstimQ21				-0.872				
SelfeffQ7				0.833				
AchivgoalQ27				0.644				
ActLearnSratQ15				0.524				
SelfeffQ5				0.484				
ActLearnSratQ17					0.876			
SelfeffQ3					0.711			
LearnenvirstimQ23						0.990		
ActLearnSratQ20						-0.702		
AchivgoalQ26						0.488		
SelfeffQ10						0.486		
SelfeffQ1						0.470		
AchivgoalQ28						0.447		
SelfeffQ6							-0.733	
AchivgoalQ29							-0.710	
ActLearnSratQ16							-0.641	
AchivgoalQ30							-0.439	
ActLearnSratQ18								0.975
ActLearnSratQ19								0.618

4. Conclusion

We test the efficacy of the 10 steps proposed in (Deshpande, 2017) by validating it via a questionnaire consisting of 30 questions that are divided into 4 scales. Positive outcomes to these 10 steps validated through the scale of computing confidence and achievement goal were however not strongly correlated with the scales of pro-activeness and learning environment stimulation. However, there is some evidence in support of including these two scales as an intrinsic part of self-motivational attitudes and group dynamics. One can ascertain from the analysis of Table 1, that low motivation students fared poorly in these two scales while high motivation students did not have much performance incentive due to gaining no credit for doing the 10 steps. Providing credit to the first few foundational yet simpler questions of (Deshpande, 2017) may enhance learning environment stimulation. This may then induce tenacity in low motivated students to build up pro-activeness. A similar outcome is anticipated for high motivated students. The results so obtained, in totality, also suggest one avenue of future investigation: We would like to know how “being clever” (Coughlan, 2016) influences success in implementation of these 10 steps. Though being clever helps, we believe that the success in implementation of these 10 steps does not and should not need any kind of cleverness. Our future aim would then be to also run these 10 steps for university students in countries ranked highly in the Organisation for Economic Co-operation and Development (OECD) index. We can then compare the implementation outcome vis-à-vis the UK. We anticipate similar outcomes amongst all participant OECD countries thus establishing invariance of SFCL towards predisposed cleverness.

5. References

Coughlan, S. (2016). *Which country really has the cleverest students*? BBC report, Available at: http://www.bbc.co.uk/news/business-37649892 [Accessed 29 December 2017].

Cretchley, P., Harman, C., Ellerton, N., and Fogarty, G. (1999). *Computation, exploration, visualisation: Reaction to MATLAB in First Year Mathematics*. 2nd Australian Symposium on Modern Undergraduate Mathematics, Proserpine, Qld, Australia.

Deshpande, A. (2017*). Financial computing literacy: 10 steps*. MSOR connections, 16(1), pp. 45-52.

Fogarty, G., Cretchley, P., Harman, C., Nerida E. and Konki, N. (2001). *Validation of a questionnaire to measure mathematics confidence, computer confidence, and attitudes towards the use of technology for learning mathematics.* Math Ed Res J, 13 (2), pp. 154-160.

Huang, W., Huang, W., Diefes-Dux, H. and Imbrie, P. K. (2006). *A preliminary validation of Attention, Relevance, Confidence and Satisfaction model-based Instructional Material Motivational Survey in a computer-based tutorial setting*. British Journal of Educational Technology, 37, pp. 243–259.

Kyng, T., Tickle, L. and Wood, L. (2013). *Perceptions of the software skills of graduates by Technology employers in the financial services industry*. International Journal of Mathematical Education in Science, 44 (8), pp. 1224-1238.

Mandler, G. (1998). Affect and learning: Causes and consequences of emotional interactions. In D. A. Grouws (Ed.), *Handbook on research on mathematics teaching and learning* New York: Macmillan. pp. 575–596.

Moser, C. A., and Kalton, G. (1980). *Survey methods in social investigations*. 2nd ed. London: Heinmann.

Ramsden, P. (2003). *Learning to teach in higher education.* 2nd edn. London: Routledge Falmer.

Tuan, H-L., Chin, C-C. and Shieh, S-H. (2005). *The development of a questionnaire to measure students' motivation towards science learning*. International Journal of Science Education, 27(6), pp. 639–654.

Weimar, P. (2013). *Better Group Work Experiences Begin with How the Groups Are Formed*. Available at: http://www.facultyfocus.com/articles/teaching-professor-blog/better-group-work-experiences-begin-with-how-the-groups-are-formed/ [Accessed 28 February 2017].

6. Appendix: SFCL survey questionnaire

Self-efficacy or computing confidence.

1. I have less trouble learning how to program using MATLAB than I do other things.

2. When I have difficulties using MATLAB, I know I can handle them.

3. I am not what I would call a computer programmer.

4. It takes me much longer to understand how to code in MATLAB than the average person.

5. I have never felt myself able to learn how to code in MATLAB.

6. I enjoy trying new ways to program in MATLAB.

7. I find MATLAB programming frightening.

8. I find many aspects of MATLAB programming interesting and challenging.

9. I don't understand how some people can seem to enjoy MATLAB programming.

10. I have never been excited about using computers.

11. I find doing MATLAB programming confusing.

12. I am nervous that I am not good enough with MATLAB programming to be able to use them in financial computing.

Active learning strategy or pro-activeness.

13. When learning new MATLAB programming concepts, I attempt to understand them.

14. When learning new MATLAB programming concepts, I connect them to my previous experiences.

15. When I do not understand a MATLAB programming concept, I find relevant resources that will help me.

16. When I do not understand a MATLAB programming concept, I would discuss it with my instructor or other students to clarify my understanding.

17. During the learning process, I attempt to make connections between the concepts that I learn.

18. When I make a programming mistake, I try to find out why.

19. When I meet MATLAB programming concept that I do not understand, I still try to learn them.

20. When new programming concepts that I have learned conflict with my previous understanding, I try to unerstand why.

<u>Learning environment stimulation</u>

21. I am willing to participate in this computing lab because the content is exciting.

22. I am willing to participate in this computing lab because the instructor uses a variety of teaching methods.

23. I am willing to participate in this computing lab because the instructor does not put a lot of pressure on me.

24. I am willing to participate in this computing lab because it is challenging.

25. I am willing to participate in this computing lab with a laptop because it allows me to consult my colleague(s) next to me.

<u>Achievement goal or student satisfaction</u>

26. I feel most fulfilled when I feel confident about the content in the computing lab.

27. During a computing lab, I feel most fulfilled when the teacher accepts my ideas.

28. During a computing lab, I feel most fulfilled when I am able to solve a difficult problem.

29. My MATLAB programming skills improved as I went completing the 10 steps.

30. I am now more confident than before of using Pseudo codes to do MATLAB programming.

CASE STUDY

Embeddedness of mathematics e-assessment and attitudes affecting adoption

Deirdre Casey, Department of Computing, Griffith College Cork, Cork, Ireland.
Email: Deirdre.Casey@griffith.ie
Julie Crowley, Mathematics Department, Cork Institute of Technology, Cork, Ireland.
Email: Julie.Crowley@cit.ie

Abstract

The Numbas e-assessment system was first introduced to Cork Institute of Technology (CIT) in 2014. It has been implemented in several modules and is used regularly by a number of Lecturers. The aim of this case study was to investigate how embedded Numbas is in the Mathematics department in CIT and to examine the perceptions of the lecturing staff which have led to their individual adoptions of the Numbas e-assessment tool. The results of a survey of lecturers within the department show that the use of Numbas is common and varied and that Lecturers' attitudes towards Numbas are positive. Several themes arose around ways to increase or improve the use of Numbas including the need for ongoing training and ideas for the broadening of use. Lessons learned could be applied to advancing the integration of Numbas in Cork Institute of Technology but may also be of use to others hoping to increase adoption of e-assessment in Mathematics within their institutions.

Keywords: Mathematics, e-Assessment, diffusion of innovations, implementation, computer aided assessment

1. Introduction

In 2014 Cork Institute of Technology (CIT) introduced e-assessment into some of its Mathematics modules. The initial motivation for using e-assessment was to manage lecturer workload in correcting continuous assessments thereby allowing more regular assessment and feedback. The Numbas e-assessment tool was chosen because it has a strong reputation, is user friendly and is compatible with the Virtual Learning Environment used on campus. The National Forum for the Enhancement of Teaching and Learning funded a joint project between University College Cork and Cork Institute of Technology in 2015 called 'Transitioning to e-Assessment in Mathematics Education (TEAME)'. The project focused on the implementation and evaluation of this Mathematics e-assessment tool and the dissemination of associated knowledge. Along with the expected benefits of facilitating more regular assessment and immediate feedback Numbas also increased student engagement and attendance in their tutorials at CIT (Carroll, et al., 2017). Numbas was rolled out to several large Mathematics modules and became integrated in the module delivery. The large scale and departmentally supported introduction of Numbas in CIT and its relative success could be used to help other institutions to encourage greater levels of adoption within their ranks.

The aim of this research is to evaluate the embeddedness of Numbas e-assessment in the department and to investigate the lecturer attitudes which have led to this level of adoption. This research focuses on the lecturer perspective rather that of the student and uses lecturer opinion to evaluate levels of embeddedness. Qualitative and quantitative data was collected through a survey in order to ascertain lecturer perceptions of the scale and width of Numbas use and their attitudes towards adopting Numbas.

The remainder of this paper is organised as follows. The next section discusses the literature in relation to diffusion of innovations and ways to evaluate the embeddedness or success of implementation. Next the methodology and methods used are considered. Then the results of the case study are presented. The final section discusses these results and their implications for Cork Institute of Technology and other institutions hoping to successfully implement Mathematics e-assessment.

2. Literature Review

The literature review focused on the spread of new ideas and how this knowledge can be applied to understanding the spread of Mathematics e-assessment.

2.1 Adoption and Diffusion of Innovations

Every new idea starts with an individual or small group and gradually diffuses outwards from that small starting point. *Adoption* is the process whereby an individual goes from first hearing about an innovation to finally adopting it. According to Rogers (1962) the adoption process can be viewed as five consecutive stages: awareness, interest, evaluation, trial and adoption. The spread of the adoption of new ideas is described by Rogers (*ibid.)* as *Diffusion of Innovation*. He described five user perceptions that effect the diffusion of any new innovation: relative advantage, compatibility, complexity, divisibility, communicability. These attitudes from Rogers work were refined and formed into an instrument to measure perceptions of adoption by (Moore & Benbasat, 1991) which measures eight perceptions:

- *Voluntariness* is the degree to which use of the innovation is perceived as being voluntary, or of free will;
- *Relative Advantage* is the degree to which an innovation is perceived as being better than its precursor;
- *Compatibility* is the degree to which an innovation is perceived as being consistent with the existing values, needs, and past experience of potential adopters;
- *Image* is the degree to which use of an innovation is perceived to enhance one's image or status in one's social system;
- *Ease of Use* is the degree to which the innovation is perceived as being easy to use;
- *Result Demonstrability* is the degree to which the results of using an innovation are perceived to be measurable, observable and easy to communicate to others;
- *Visibility* is the degree to which the innovation is perceived to be visible within the organisation/department;
- *Trialability* is the degree to which it is perceived that the innovation can be experimented with before adoption.

High scores in these eight perceptions predict adoption of the innovation by an individual.

2.2 Mathematics e-Assessment and its Diffusion

e-Assessment is assessment that uses or is enhanced by electronic technologies and it can offer many advantages over traditional pen and paper assessment. These advantages include immediate feedback, the possibility for students to make more than one attempt and supplying hints (Biggs & Tang, 2007; Jenkins, 2004). e-Assessment in Mathematics gained momentum in the 1990's with the development of WeBWorK at the University of Rochester and has continued to grow and diversify. In 2012 developers at Newcastle University released a new open source Mathematics e-assessment system called 'Numbas'. Numbas is now used successfully in several large institutions around the world (Loots, et al., n.d.; Foster, et al., 2012).

Although e-assessment in Mathematics is far from a new innovation it's diffusion rates through the mathematics education community remain slow. The uptake of Computer Aided Assessment (CAA) has lagged behind the expectations of academics in the field (Warburton, 2009). This may be due to the fact that apart from the advantages that e-assessment may offer it also comes with difficulties and potential risks (Jenkins, 2004). A major barrier to the implementation of CAA is the time investment required in learning how to use the system and also in the development of questions (Jenkins, 2004). Staff require training on how to use the particular system along with general guidance on good question design (Sim, et al., 2004). The level of initial and ongoing staff training required relies, at least in part, on institutional support which seems to be very beneficial in terms of ensuring implementation of CAA (ibid.).

Judging or measuring the level of embeddedness of e-assessment has been considered by several authors. Metrics for successful implementation were discussed by Warburton (2009) but he doesn't describe an instrument to measure these metrics. However, McCann (2010) uses a questionnaire to attempt to measure these metrics and we have followed a similar approach. Here we measure 'Width' of practice by focusing on range of use and number of users (Warburton, 2009). A question was included on the survey to investigate the range of uses of the Numbas system in terms of the three types of assessment: Assessment of Learning (AoL), Assessment for Learning (AfL) and Assessment as Learning (AaL) (Zeng, et al., 2018).

3. Methodology

This case study was carried out through mixed methods research using a concurrent embedded design (Creswell, 2009). Most of the data collected was quantitative with embedded qualitative elements where the qualitative data is intended to provide a supporting role to the quantitative data. The two methods are used side by side to provide different perspectives on the same case. The survey used contained four parts: Demographics, width of use and embeddedness, perceptions of adoption, open-ended questions. The third section of the survey, dealing with perceptions of adoption, was adapted from the 25-item instrument developed and recommended by Moore and Benbasat (1991). The respondents were requested to rate their agreement to each statement on a seven-point scale, 7 = extremely agree and 1 = extremely disagree. The survey ended with two open ended questions for the purposes of collecting supporting qualitative data. The questions were as follows: (i) What do you think would increase or improve the use of Numbas within CIT? (ii) Do you have any other comments relating to the usefulness or adoption of Numbas?

4. Methods

The survey was sent to the current members of the Mathematics Department in CIT. Quantitative data analysis was carried out on the data from the first three sections of the survey. The mean and standard deviation were found separately for each item in the Perceptions of Adoption section. The data was then grouped by perception and an overall mean and standard deviation was found for each perception.

The responses to the open-ended questions were combined due to the similarity and overlap of responses. The text data was subjected to conventional content analysis where codes are derived directly from the data without reference to pre-conceived categories. Directive content analysis was also undertaken to attempt to code the data in relation to the eight perceptions described above but the data did not fit well into these categories and so the researchers felt that this would not give a valid representation of the data obtained.

5. Results

The survey received 9 responses from the 29 lecturers in the department giving a response rate of 31%. All respondents had some experience working with Numbas and no responses were received from non-users of the system.

5.1 Width of Practice/ Metrics for successful implementation

Figure 1 shows that respondents have used Numbas for all three types of assessment:

- Assessment of Learning (AoL): Testing prerequisite knowledge, Diagnostic testing;
- Assessment for Learning (AfL): Low stakes assessment and homework;
- Assessment as Learning (AaL): No stakes assessment in tutorials, Support tool.

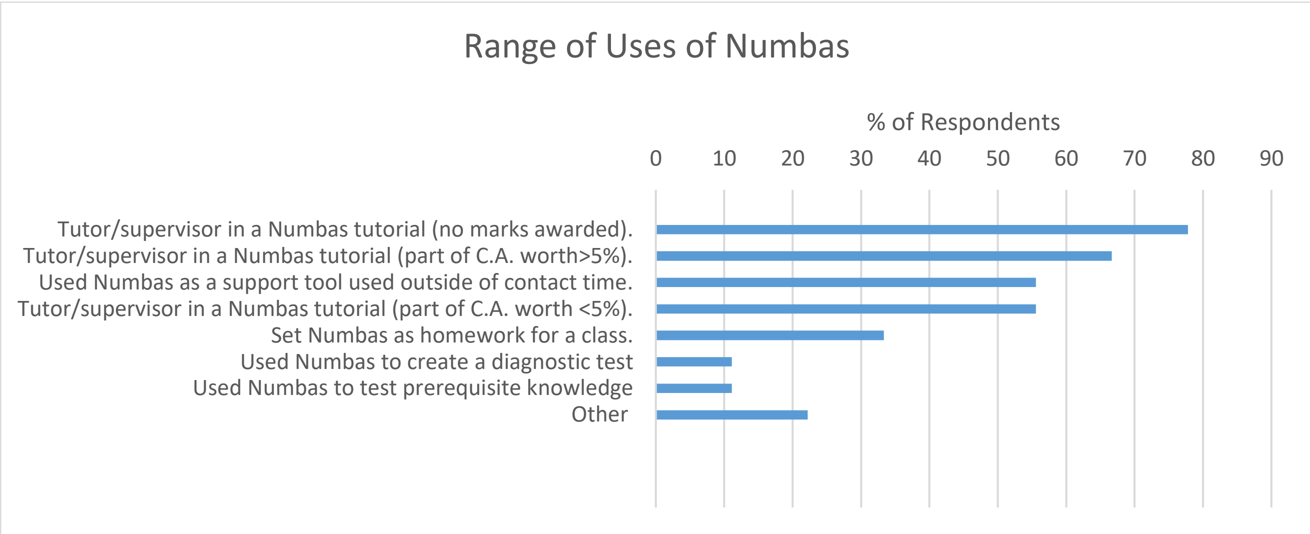


Figure 1: Range of Uses of Numbas among Respondents

5.2 Depth of use/knowledge

As shown in Figure 2 the respondents reported varied levels of engagement with the Numbas system. Over 50% of respondents have developed Numbas questions and so they are actively contributing to the question bank. Only one respondent had not done any of the listed activities using Numbas.

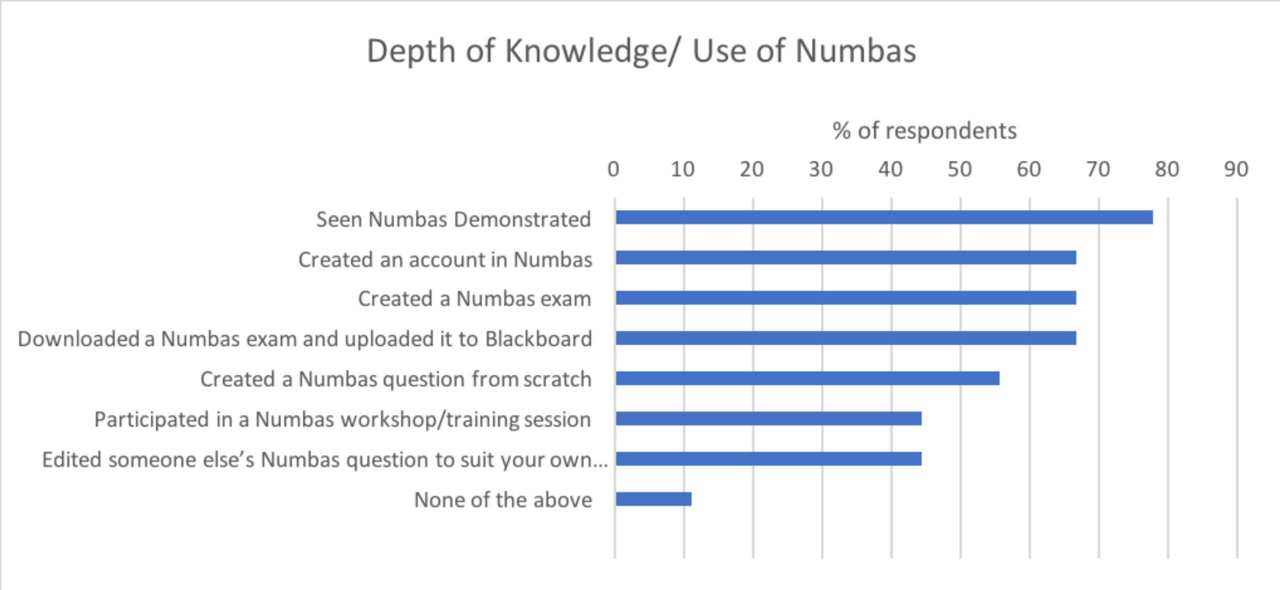


Figure 2: Depth of Use of Numbas Among Respondents

5.3 Attitudes/Adoption Attributes (of adopters)

The results of the Instrument to Measure the Perceptions of Adopting Numbas e-assessment are summarised in Table 1 where the respondents had rated their agreement to each statement on a seven-point scale, 7 = extremely agree and 1 = extremely disagree. The most positive perception was "Results Demonstrability" while the least positive was "Image". Trialability and Image were the only two perceptions that did not elicit agreement from the respondents.

Table 1: Overview of the Perceptions of Adopting Numbas: Where respondents were asked to rate their agreement with statements related to each attribute on a seven - point scale 7 = extremely agree and 1 = extremely disagree.

Attribute	**Mean**	**SD**	**Scale of Agreement**
Results Demonstrability	5.81	1.21	Agree
Voluntariness	5.5	2.26	Agree
Compatibility	4.92	1.66	Somewhat Agree
Ease of Use	4.85	0.88	Somewhat Agree
Visibility	4.69	1.31	Somewhat Agree
Relative advantage	4.64	1.21	Somewhat Agree
Trialability	4.37	2.17	Neutral
Image	2.11	1.15	Disagree

5.4 Qualitative Data

All respondents gave a response to at least one of the two open ended questions: (i) What do you think would increase or improve the use of Numbas within CIT? (ii) Do you have any other comments relating to the usefulness or adoption of Numbas? Five concepts presented themselves from thematic analysis:

- The need for training;
- Expansion of use;
- Fairness and transparency to students;
- Need for resources;
- Teaching and Learning concerns or considerations.

Of the five concepts the need for training was the most prevalent. The desire for more training was alluded to by 5 of the 9 respondents. Within the five key concepts there were thirteen categories. These are summarised in Figure 3.

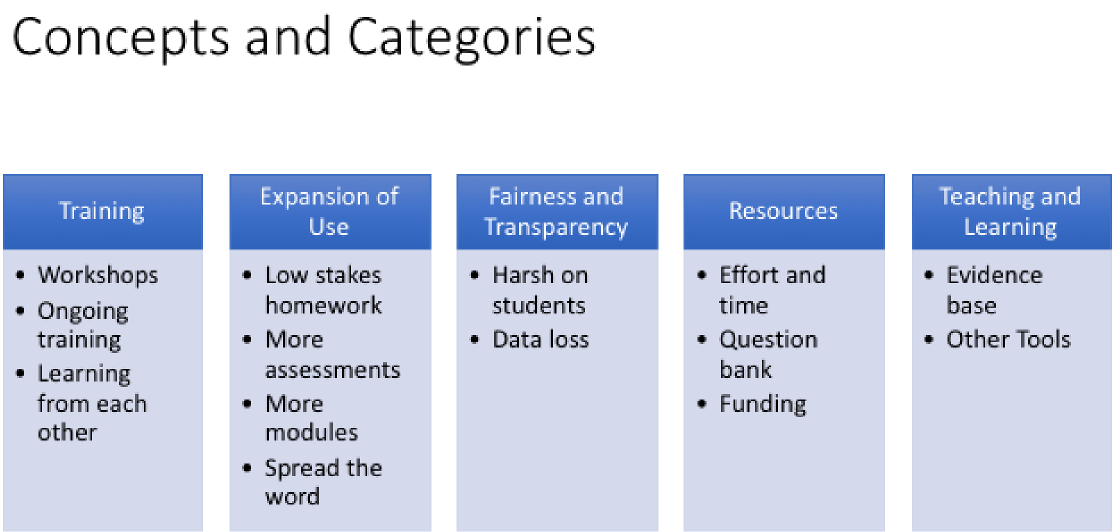


Figure 3: Five Main Concepts and thirteen categories emerging from the qualitative data

Table 2 provides a detailed breakdown of the concepts and categories along with illustrative quotes from the data for each category.

Table 2: Concepts, Captions and Supporting Quotes

Concepts	% of 23 items	Categories	Sample Items
Training	26.09% (6 items)	Workshops	*"Training on setting up and writing questions in Numbas"*
		Ongoing Training	*"a workshop focusing on the creation of more advanced questions"*
		Learning from each other	*"an experienced user and a new user [could] develop content for a course that they are both teaching together"*
Expansion of use	29.09% (6 items)	Low Stakes Homework	*"could potentially be used for homework too"*
		More Assessment	*"more assessments via Numbas"*
		More Modules	*"more accessible to other Mathematics Modules rather than basic first year modules"*
		Spread the Word	*"introducing it to a larger number of staff"*
Resources	21.74% (5 items)	Time and Effort	*"Time to develop more questions – writing new questions is time consuming"*
		Question Bank	*"Bigger question banks"*
		Funding	*"Funding"*
Fairness and transparency	13.04% (3 items)	Harsh on Students	*"It doesn't take into account for people that get correct answer but could be off by a decimal place"*
		Data Loss	*"being able to see the students' responses to the questions"*
Teaching and learning	8.7% (2 items)	Evidence Base	*"Evidence of students using Numbas having improved learning"*
		Other Tools	*"I think it is a good tool, but just one of many different tools you can use to teach"*
Other	4.35% (1 item)		*"I haven't used Numbas yet: still developing tools"*

6. Discussion

The results show a broad width of practice using Numbas within the department both in terms of range of use and number of users. Together these results indicate that Numbas has become reasonably well embedded in the department. All the respondents have had some level of engagement with Numbas. They all fall somewhere between the evaluation, trial and adoption stages of Rogers stages of adoption of an innovation (Rogers, 1962) tending mostly towards the adoption phase. No conclusions can be reached from our results for perceptions or attitudes of non-adopters or reasons for not adopting. Among responders the range of uses of Numbas were large and varied. It is used for Assessment of Learning (AoL), Assessment for Learning (AfL) and Assessment as Learning (AaL) (Zeng, et al., 2018) within the department. Respondents showed various levels of engagement with the Numbas system ranging from downloading pre-existing exams to creating new questions from scratch.

Overall results demonstrability was the most positive adoption attribute displayed by the respondents. This indicates that the degree to which the benefits of Numbas can be measured, observed and easily communicated to others is the main reason for the successful embeddedness of Numbas within the department. Voluntariness was also very positive but there were some more diverse opinions on this with some lecturers feeling that there was a requirement on them to use Numbas. The most interesting and unexpected result was that the lecturers disagreed that using Numbas would enhance their image or standing within the department. Image, at 2.11, was by far the lowest scoring attribute of Numbas. It also had the second lowest standard deviation at 1.15 indicating that there was a high level of agreement between participants This means that they adopted the system in spite of this attitude being in contradiction of the findings of previous work. It is possible that this is in some way related to Irish culture as Ireland has a very low Power Difference Index, meaning that Irish people believe that power and status should be distributed equally (Hofstede, 1983). And so, it is perhaps not surprising that this group would find it difficult to see that using any one system would be a status boost as status is not easily seen or given.

The lecturers highlighted the need for training, both initial training to increase awareness of Numbas and ongoing or more advanced training to improve the skills of the adopters. In relation to training there was also comment about collaboration between new users of Numbas and the more experienced adopters as a form of ongoing training. There was mention that the use of Numbas should be broadened to more lecturers and courses and that purposes of use should also be extended e.g. to low stakes or no stakes homework. Lecturers commented on the time required to implement Numbas and cited the need for larger CIT specific question banks and additional funding as possible support. There were some concerns regarding the need to be fair to students in the marking of questions and also there was concern about the current inability to see all data relating to the question that an individual student was presented with an also concerns regarding loss of data between Numbas and the virtual learning environment. The final theme that was mentioned related to concerns about the pedagogical foundations of using Numbas and a need to see evidence of the teaching and learning benefits particular to Numbas that might differentiate it from other tools or methods.

7. Conclusions and Future Work

Numbas e-assessment has been used in Cork Institute of Technology for several years and the aim of this case study was to investigate whether Numbas is embedded within the department, how lecturers feel about Numbas and what lessons can be learned regarding successful implementation of Mathematics e-assessment. Diffusion of an innovation relies on many individuals deciding to adopt the innovation and this depends on eight main attitudes of the individual towards the innovation: Voluntariness, Relative Advantage, Compatibility, Image, Ease of Use, Results Demonstrability, Visibility and Trialability. The case study was designed around an online questionnaire collecting mainly quantitative data with an element of embedded qualitative data. Quantitative and qualitative analysis was conducted on the data. There is broad use of Numbas within the department for various types of assessment and in various contexts, both inside and outside contact time. As expected the lecturers expressed positive perceptions towards most of the attributes with the notable exception of Image which runs contrary to the work of previous researchers. Themes emerging from the qualitative data included the need for ongoing staff training, concerns about fairness and transparency to students and ideas for the expansion of Numbas use. Future work will involve follow up interviews with some of the survey participants and, possibly more importantly, with some of the non-adopters who did not complete the survey.

8. Appendix

Survey Instrument adapted from Moore and Benbasat (1991) with the breakdown of the results for each item.

Attribute	Mean	SD
Voluntariness (agree)	***5.5***	***2.26***
My institute does not require me to use Numbas	4.89	2.71
Although it might be helpful, using Numbas is certainly not compulsory in my job.	6.11	1.62
Relative advantage (somewhat agree)	***4.64***	***1.21***
Using Numbas improves the quality of my work	4.78	1.39
Using Numbas enables me to accomplish tasks more quickly	5	1.18
Using Numbas makes it easier to do my job	5.22	0.97
Using Numbas enhances my effectiveness on the job	4.33	0.87
Using Numbas gives me greater control over my work	3.89	1.36
Compatibility (somewhat agree)	**4.92**	**1.66**
Using Numbas is compatible with all aspects of my work	4.56	1.74
Using Numbas fits into my work style	5.11	1.69
I think that using Numbas fits well with the way I like to work	5.11	1.69
Image (disagree)	***2.11***	***1.15***
People in my department who use Numbas have more prestige than those who do not	4.09	1.15

People in my department who use Numbas have a high profile	2.44	0.88
Using Numbas is a status symbol in my department	1.56	0.53
Ease of Use (somewhat agree)	***4.85***	***0.88***
I believe it is easy to get Numbas to do what I want it to do	4.44	0.73
Overall, I believe Numbas is (would be) easy to use	5.13	0.83
Learning to operate Numbas is easy for me	5	1
Results Demonstrability (agree)	***5.81***	***1.21***
I would have no difficulty telling others the results of using Numbas	6.33	1.32
The results of using Numbas are apparent to me	4.78	0.83
I would NOT have difficulty explaining why using Numbas may or may not be beneficial	6.33	0.71
Visibility (somewhat agree)	***4.69***	***1.31***
In my department, one sees many people using Numbas	4.33	1.12
Numbas IS very visible in my department	5.22	1.39
Trialability (neutral)	***4.37***	***2.17***
I am able to experiment with Numbas as necessary	4.89	1.76
Before deciding whether to use Numbas I was able to try it out properly	4.56	2.29
I was permitted to use Numbas on a trial basis long enough to see what it could do.	3.67	2.45

9. References

Biggs, J. & Tang, C., 2007. *Teaching for Quality Learning at University.* 3rd Edition ed. s.l.:Open University Press.

Carroll, T. et al., 2017. Numbas as an engagement tool for first-year Business Studies students. *MSOR Connections,* 15(2), pp. 42-50.

Creswell, J. W., 2009. *Research Design: Qualitative, Quantitatvie and Mixed Methods Approaches.* 3rd Edition ed. s.l.:SAGE Publications.

Foster, B., Perfect, C. & Youd, A., 2012. A completely client-side approach to e-assessment and e-learning of mathematics and statistics. *International Journal of e-Assessment,* 2(2).

Hofstede, G., 1983. The cultural relativity of organizational practices and theories. *Journal of International Business Studies,* 14(2), pp. 75-89.

Jenkins, M., 2004. Unfulfilled Promise: formative assessment using computer-aided assessment. *Learning and Teaching in Higher Education.*

Loots, I., Fakir, R. & Roux, A., n.d. *An evaluation of the challenges and benefits of e-assessment implementation in developing countries: a case study from the University of Pretoria.* Cape Town, s.n., pp. 418 -420.

McCann, A. L., 2010. Factors affecting the adoption of an e-assessment system. *Assessment & Evaluation in Higher Education,* 35(7).

Moore, G. C. & Benbasat, I., 1991. Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research,* September, 2(3), pp. 192-222.

Rogers, E., 1962. *Diffusion of Innovations.* New York: Free Press.

Sim, G., Holifield, P. & Brown, M., 2004. Implementation of computer assisted assessment: lessons from the literature. *Research in Learning Technology,* 12(3), pp. 215 - 229.

Warburton, B., 2009. Quick win or slow burn: modelling UK HE CAA uptake. *Assessment & Evaluation in Higher Education,* 34(3), pp. 257 - 272.

Zeng, W., Huang, F., Yu, L. & Chen, S., 2018. Towards a learning-oriented assessment to improve students' learning-a critical review of the literature. *Educational Assessment, Evaluation and Accountability,* August, 30(3), pp. 211 - 250.

CASE STUDY

Developing a math e-learning question specification to facilitate sharing questions between different systems

Mitsuru Kawazoe, Faculty of Liberal Arts and Sciences, Osaka Prefecture University, Sakai, Japan. Email: kawazoe@las.osakafu-u.ac.jp
Kentaro Yoshitomi, Faculty of Liberal Arts and Sciences, Osaka Prefecture University, Sakai, Japan. Email: yositomi@las.osakafu-u.ac.jp
Yasuyuki Nakamura, Graduate School of Informatics, Nagoya University, Nagoya, Japan. Email: nakamura@nagoya-u.jp
Tetsuo Fukui, Mukogawa Women's University, Nishinomiya, Japan. Email: fukui@mukogawa-u.ac.jp
Shizuka Shirai, Osaka University, Toyonaka, Japan. Email: shirai@ime.cmc.osaka-u.ac.jp
Takahiro Nakahara, Sangensha LLC., Chitose, Japan. Email: nakahara@3strings.co.jp
Katsuya Kato, Cybernet Systems Co., Ltd., Tokyo, Japan. Email: katsuyak@cybernet.co.jp
Tetsuya Taniguchi, Nihon University, Tokyo, Japan. Email: taniguchi.tetsuya@nihon-u.ac.jp

Abstract

In recent years, e-assessment has become increasingly popular in mathematics education. However, there are several different systems, and hence, the contents need to be developed independently in each system. Sharing contents between different systems is important for the diffusion of math e-learning/assessment systems. This study focuses on sharing computational questions, which are the main contents in most systems. The structure of such questions seems essentially compatible between many systems. Based on this observation, a specification, namely mathematics e-learning question specification (MeLQS) is proposed and described as a common base for developing contents in computer-algebra-system-based mathematics e-learning/assessment systems. Furthermore, the development of authoring tools for MeLQS is reported.

Keywords: mathematics e-learning system, contents sharing.

1. Introduction

E-assessment has become increasingly popular in mathematics education. Several different and independently developed math e-learning/assessment systems have been reported—STACK (Sangwin, 2013; Stack.ed.ac.uk, 2018), Möbius Assessment (DigitalEd.com, 2018) (formerly Maple T.A.), Numbas (Numbas.org.uk, 2018), MATH ON WEB (Osakafu-u.ac.jp, 2018), WeBWorK (The Mathematical Association of America, 2018), and many others. In these math e-learning/assessment systems, questions need to be implemented independently in each system due to non-compatibility between the systems. Because the development of content is difficult, it is important to reduce the burden on content developers. There have been a few attempts to share e-learning contents. The Abacus project (2018) and ItemBank (Nakamura et al., 2014) aim to share questions used in STACK, and Möbius Cloud (DigitalEd.com, 2018) aims to share questions used in Möbius Assessment. These attempts also contribute to developing communities of content developers at different institutions. However, these are closed to specific systems.

Sharing content between different systems makes it possible to not only reduce the burden on contents developers but also connect the communities of content developers for different systems and enhance the diffusion of math e-learning/assessment systems. Therefore, computational questions in math e-learning/assessment systems developed with computer algebra systems

(CASs), such as solving systems of linear equations and solving differential equations, are focused on, because they are implemented in a similar way in each system.

One approach is to develop a tool for importing questions from one system to another. Kinnear and Sangwin (2016) developed a 'MapleTA question importer' which imports Maple T.A. questions to STACK. However, its functionality is restricted to 'import questions placing corresponding fields in the correct places pending editing by an intelligent human' (ibid.). Developing an import tool is difficult, even if it is a semiautomatic one, because translating codes between different systems is complex and difficult. In addition, with this approach, several tools for all pairs of systems must be developed.

Another approach is to make a common 'question specification' of math e-learning questions and share the data of questions, containing the idea of questions, marking principles, and algorithms needed for implementation, in a common format that does not depend on any specific e-learning/assessment system or any CAS language. With this approach, content developers do not need to learn codes of questions in another system in order to implement them in the target system. They only need to implement a question according to its question specification. Hence, knowledge of multiple CAS languages and multiple systems is not required.

This work shows that an approach that involves making a common question specification for math e-learning is possible and proposes such a question specification. The proposed specification is called the 'mathematics e-learning question specification' (MeLQS).

We think that it is useful to design MeLQS along with a process of developing questions in math e-learning/assessment systems. Each question is produced from an idea of a mathematics teacher. First, a mathematics teacher who would like to create a question for an online test acquires an idea based on his/her knowledge of mathematics teaching. The idea includes the following: the types of questions that need to be presented to students; methods to verify the correctness of students' answers; the types of errors often found in students' answers; and the types of feedback that need to be provided to students. Then, the teacher's idea is formulated as a set of algorithms on marking, identifying errors, and so on. Finally, the question is implemented in a target system by coding algorithms. Thus, a process of developing questions can be divided into the following three phases: 1) describing a teacher's idea, 2) algorithmizing the teacher's idea, and 3) implementing algorithms. Based on these three phases, we propose MeLQS as a pair of two specifications, concept design and implementation specification. The teacher's idea in the first phase is described in a concept design, and its algorithmization in the second phase is described in an implementation specification.

The outline of this work is as follows. Section 2 focuses on the three systems used in Japanese universities, and it is observed that these three systems have a common data structure for questions. In Section 3, the MeLQS is proposed based on the observations in Section 2. Finally, conclusions are provided in Section 4.

2. Observation

This section focuses on a data structure of questions in the following three systems used in Japanese universities: STACK, Möbius Assessment, and MATH ON WEB.

2.1. Question data in STACK

STACK is an online assessment system for mathematics and STEM that works as one of the question types for Moodle and ILIAS; it can also be integrated more widely via LTI. Maxima (Maxima.sourceforge.net, 2018) is used for the manipulation of symbolic and numerical expressions in STACK. The question data of STACK consist of question variables (potentially random variables that can be used to generate a question), question text (a statement of the question students actually see), general feedback (a general feedback text, usually a worked solution, that is shown to all students after they attempt the question), inputs (areas where students input answers and some input types—algebraic input, numerical input, matrix, and so on are allowed) to be placed in a question text, and potential response trees that are the algorithms that establish the mathematical properties of the students' answers and generate feedback with scores.

2.2. Question data in Möbius Assessment

Möbius Assessment is an online testing and assessment system designed especially for courses involving mathematics, and it can be incorporated into virtually any learning management system, such as Moodle. It consists of a wide variety of question types to assess the mathematical ability of students. The system makes it possible to generate mathematical questions algorithmically as well as evaluate the students' responses flexibly with a grading code programmable by Maple, a computer algebra system. When a Maple-graded question is used, students' responses are graded with Maple (Maplesoft.com, 2018). The dataset basically consists of five primary items: question text (a mathematical problem statement), response area (a field for students to enter their own answers), grading code (a field to write Maple code to grade the student response), algorithm (a field to write algorithmic code to generate formulae, matrices, and plots), and feedback (a field to write a comment for the question itself). Compared with the potential response tree (PRT) of STACK, Möbius Assessment does not have an identical capability to the PRT but can have multiple parts, which could be subquestions or messages (feedback), in a single question with adaptive sections (separators to divide questions). Each part is selectively displayable only with the students' responses: correct or incorrect answers. When a question is generated online, its algorithm code is executed once. Then, the algorithmic elements, such as its answer and mathematical expressions, are referred from the other items, such as question text and grading code.

2.3. Question data in MATH ON WEB

MATH ON WEB is a web site that offers two mathematics e-learning/assessment systems: WMLS and WASM (Kawazoe and Yoshitomi, 2017). These two systems have been developed with web*Mathematica* (Wolfram Research, Inc., 2018). Although it has different types of contents, questions in the drill section of WMLS and questions in WASM have almost the same structures as the ones in STACK and Möbius Assessment. The question data comprises the following: a problem template (a question text with places where mathematical expressions are inserted); problem parameters (mathematical expressions given as a list or as a *Mathematica* program generating mathematical expressions) to be inserted in a question text; an answer column (HTML form with variables setting); an answer analyser (a marking algorithm given as a *Mathematica* program) that marks students' answers and also identifies the type of error if students fail to solve the question; a list of feedback messages corresponding to the output of the marking algorithm; and a problem example with its model answer and a guide for how to solve (PDF). Marking algorithms are written as the which statement of *Mathematica*, and marking is performed by a first-match method.

2.4. A common data structure of STACK, Möbius Assessment, and MATH ON WEB

From the observation of the above three systems, a common structure of question data can be observed, which consists of the following data.

- Question text with places where mathematical expressions are inserted
- Mathematical expressions inserted into the question text (given as a list or as a CAS program generating them)

- Answer box (given as an HTML form, including variable setting)
- Marking algorithm (given as a CAS program)
- Fixed feedback message shown to all students
- List of feedback messages for adaptive feedback according to the result of marking

Because Möbius Assessment does not have a function to give a different feedback message depending on the result of marking, the last data were regarded as optional. Each of the data is named differently in each system. The results of the observation are summarized in Table 1.

Table 4. Correspondence between data of each system

Item	STACK	Möbius Assessment	MATH ON WEB
Question text	Question text	Question text	Problem template
Mathematical expressions	Question variables	Algorithmic variables	Problem parameters
Answer box	Inputs	Response area	Answer column
Marking algorithm	Potential response trees	Grading code (available only in Maple-graded question)	Answer analyser
Fixed feedback message	General feedback	Feedback	Model answer of a problem example, and a guide for how to solve
Adaptive feedback messages	Feedback defined in potential response trees	Available but limited inside the sequence of question text with adaptive sections by correct or incorrect answers	A list of feedback messages with respect to the result of answer analyser

3. MeLQS

With the observation presented in the previous section, a common question specification (MeLQS) is proposed that can be used at the design stage of questions for different CAS-based math e-learning systems. In previous work (Nakamura et al., 2018), only a brief idea of MeLQS was presented, and a prototype of an authoring tool for concept design that is a part of MeLQS was shown. Here, the details of MeLQS and authoring tools are described. They have been updated from the ones reported previously.

3.1. Overview of MeLQS

MeLQS is designed as a pair of two specification: concept design and implementation specification. Concept design is a specification describing a teacher's idea. Its objective is to share the idea of questions from the viewpoint of mathematics teachers. Hence, it is described with the mathematics teachers' language, not with programming language. Implementation specification is a specification describing the algorithmization of the teacher's idea. The objective of implementation specification is to provide sufficient information needed in implementing questions to programmers or engineers who are engaged in authoring questions for online tests. In the implementation specification, algorithms for implementing questions are provided, but their descriptions do not depend on a specific math e-learning system. These specifications are described in the following sections.

3.2. Concept Design

The concept design aligns with the phase of describing a teacher's idea. The concept design consists of data corresponding to the items in Table 1 together with some supplemental data (subject, course, learning unit, title, the aim of the question, answer example, and remark). The question text, mathematical expressions, and answer box in Table 1 are compiled into 'Question' in the concept design. A 'question' is described as a sample question with examples of mathematical expressions and an answer box. The marking algorithm and adaptive feedback messages are compiled into 'Check List'. A check list is the most important data in the concept design, and it describes how to assess students' answers based on the teachers' knowledge of students' errors. The fixed feedback message in Table 1 is optional data which can be attached to 'Answer Example' in the concept design. Sample data of concept design are shown below.

- *Subject/course/learning unit*: mathematics(college)/linear algebra/Euclidean vector space
- *Title*: Length of vectors
- *The aim of the question*: To assess students' understanding of the definition of length of vectors in the numerical vector space
- *Question*:

Find the length $\|\boldsymbol{a}\|$ of the vector $\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 3 \end{pmatrix}$ with respect to the dot product. [Answer] $\|\boldsymbol{a}\| =$ _______

- *Answer example*: $\|\boldsymbol{a}\| = \sqrt{15}$
- *Check list*:

No	Condition	Feedback	Score (%)	Remark
1	The input answer is the same as the correct answer.	Correct!	100	
2	The input answer is a negative number.	The length cannot be negative.	0	
3	The input answer is an imaginary number.	Your answer is an imaginary number. The length cannot be an imaginary number.	0	
4	The input answer is a sum of squares of the coordinates. (Taking a square root is forgotten.)	Recall the definition of the length.	0	

- *Remark*: no comment.

An authoring tool has been developed for concept design (Figure 1). With the authoring tool, a teacher can input data step-by-step. Mathematical equations can be described in TeX format. The next version of the authoring tool will include a math input interface MathTOUCH (Shirai and Fukui, 2017), which would help users in inputting mathematical equations.

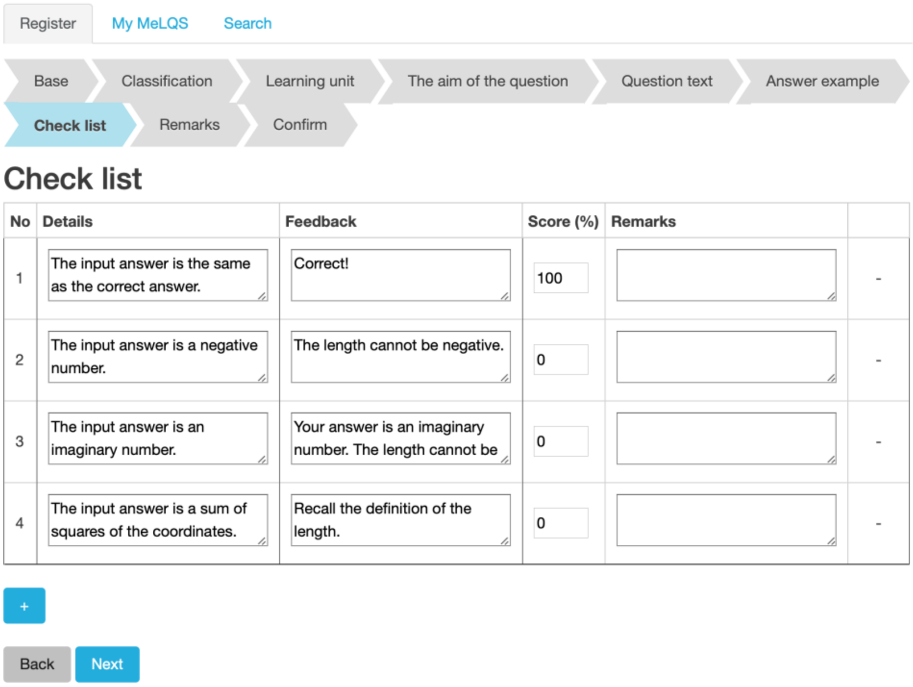


Figure 1. Screenshot of the MeLQS authoring tool

3.3. Implementation Specification

The development of implementation specification is ongoing work. Implementation specification aligns with the phase of algorithmizing the teacher's idea. The main contents of the implementation specification are the following three algorithms.

- [Algorithm 1] Algorithm for generating mathematical expressions embedded in a question text (it can be replaced by a random choice from a list of selected mathematical expressions)
- [Algorithm 2] Algorithm for generating an answer example
- [Algorithm 3] Algorithm for marking students' answers (and identifying students' errors)

Implementation specification consists of the above algorithms, a pointer to a concept design, and a question text with explicit insert positions of mathematical expressions generated by Algorithm 1.

Implementation specification is created from concept design data, but it does not necessarily have to be created by the same person who created the corresponding concept design. Anyone who can understand the concept design and can algorithmize it may create an implementation specification.

An authoring tool for implementation specification is still in progress. The user interface of the tool will be developed with a block programming interface like Blockly (Google, 2018), and the data will be stored as XML. A sample of a marking algorithm described with Blockly is shown in Figure 2. In Figure 2, 'input_answer', 'mark', 'feedback', and 'correct_answer' are variables for a student's answer, score, feedback message, and correct answer, respectively.

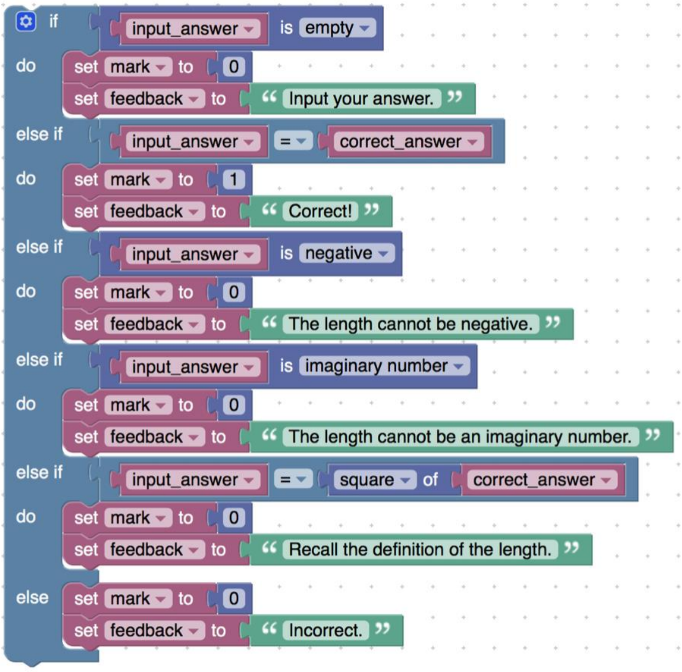


Figure 2. Sample of a marking algorithm described in implementation specification

To make descriptions of algorithms simple, each algorithm needs to be described with blocks representing mathematical functions or mathematical procedures. A sufficient number of blocks of mathematical functions/procedures will be given as preset blocks, but custom blocks could be created if they are needed.

3.4. How to implement questions based on implementation specification

Finally, we mention the implementation of questions based on the implementation specification, which is ongoing work. Questions can be implemented in each system by writing codes based on the data described in the implementation specification. Implementation is performed by a person who has the skill to write codes for a target system. Implementation is intended to be performed manually at this moment. Developing an automatic export tool from the implementation specification, although it might be partially automatic, is a future task. To develop such a tool, libraries for each target system must be developed, consisting of codes of the target system representing mathematical blocks used in algorithms in the implementation specification.

4. Conclusion

In this work, with the observation of the common data structure of three math e-learning systems, MeLQS, which can be used in different CAS-based math e-learning systems, was proposed. MeLQS consists of two types of specification: concept design and implementation specification. An authoring tool for concept design was developed. An authoring tool for implementation specification is still in progress, but the idea of using a block programming interface in describing algorithms is outlined. Implementation based on the implementation specification is supposed to be performed manually at this moment. An automatic export from the implementation specification is a future task.

5. Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP16H03067.

6. References

Abacus.aalto.fi, 2018. *Abacus project*. Available at: https://abacus.aalto.fi. [Accessed 9 November 2018].

DigitalEd.com, 2018. *MÖBIUS ASSESSMENT – ONLINE ASSESSMENT SYSTEM FOR STEM COURSES.* Available at: https://www.digitaled.com/products/assessment/ [Accessed 20 November 2018].

Google.com, 2018. *Blockly*. Available at: https://developers.google.com/blockly/ [Accessed 12 November 2018].

Kawazoe, M. and Yoshitomi, K., 2017. E-learning/e-assessment systems based on webMathematica for university mathematics education. *MSOR Connections*, 15(2), pp.17–24.

Kinnear, G. and Sangwin, C., 2016. MapleTA question importer. Available at: https://github.com/maths/moodle-qformat_mapleta [Accessed 9 November 2018].

Maplesoft.com, 2018. Maple for Education & Research: The Essential Tool for Mathematics. Available at: https://www.maplesoft.com/MapleEducation/ [Accessed 28 November 2018].

Maxima.sourceforge.net, 2018. Maxima, a computer algebra system. Available at: http://maxima.sourceforge.net [Accessed 28 November 2018].

Nakamura, Y., Taniguchi, T. and Nakahara, T., 2014. Item bank system for the mathematics e-learning system STACK. *Electronic Journal of Mathematics & Technology*, 8(5), pp.355–362.

Nakamura, Y., Yoshitomi, K., Kawazoe, M., Fukui, T., Shirai, S., Nakahara, T., Kato, K. and Taniguchi, T., 2018. Effective use of math e-learning with questions specification. In: J. Silverman and V. Hoyos, eds. *Distance Learning, E-Learning and Blended Learning in Mathematics Education*, ICME-13 Monographs. Springer.

Numbas.org.uk., 2018. Numbas: Really versatile maths e-assessment. Available at: https://www.numbas.org.uk [Accessed 12 November 2018].

Osakafu-u.ac.jp, 2018. *MATH ON WEB: Learning college mathematics by webMathematica*. Available at: http://www.las.osakafu-u.ac.jp/lecture/math/MathOnWeb/ [Accessed 12 November 2018].

Sangwin, C., 2013. *Computer aided assessment of mathematics*. Oxford: Oxford University Press.

Shirai, S. and Fukui, T., 2017. MathTOUCH: Mathematical input interface for e-assessment systems. *MSOR Connections*, 15(2), pp.70–75.

Stack.ed.ac.uk, (2018). STACK. Available at: http://www.stack.ed.ac.uk [Accessed 28 November 2018].

The Mathematical Association of America, 2018. *WeBWorK*. Available at: http://webwork.maa.org [Accessed 12 November 2018].

Wolfram Research, Inc., 2018. Mathematica, Version 11.3. Champaign, IL.

OPINION

Mathematics support centres from a sociocultural point of view

Kirsten Pfeiffer, School of Mathematics, Statistics and Applied Mathematics, NUI Galway, Ireland.
Email: kirsten.pfeiffer@nuigalway.ie

Abstract

Mathematics support centres provide opportunities for students to work in groups or explore mathematics with some guidance, which is often difficult to achieve in large lectures or tutorials. In this article I discuss the role of mathematics support centres from a socio-cultural point of view. In this view learning takes place as participation in a community of practice. Providing access to such a community is seen as crucial for the transition from school to university, however it is particularly challenging to provide authentic opportunities for mathematical practice in a university environment. I argue that mathematics support centres have potential to provide such opportunities for students and are therefore significant for the progress of newcomers in the practice.

Keywords: Maths Support Centre, Community of Mathematical Practice, Learning as Participation in a Community of Practice

1. Introduction

Maths support centres support learning of mathematics at further and higher education institutions to complement lectures, tutorials, etc. In this role many maths support centres aim to provide friendly, informal environments where students of all mathematical levels of mathematical attainment can work, study and support each other, take control of their own learning and build confidence in their own mathematical ability.

Figure 5. Students studying in the Maths Support Centre at NUI, Galway

In this article I take a sociocultural view and argue that maths support centres particularly provide opportunities for students to participate in *communities of mathematical practice* and therefore play an important role for the development of mathematical identities.

The next section explains what is meant by community of mathematical practice, and how learning appears in such a community.

2. Learning from a sociocultural perspective

In the field of education research the view of *learning* has changed in the last 40 years: from mostly influenced by constructivist views concerned with the development of individual learners (e.g. Piaget or Bruner), to more sociocultural approaches to learning based on the work of Vygotsky (1986). A key feature of this view of human development is that learning develops out of social interaction. Vygotsky argues that a child's development cannot be understood by a study of the individual in isolation. The external social world in which that individual life has developed must also be examined. Vygotsky describes learning as being embedded within social events and occurring as a child interacts with people, objects, and events in the environment.

Influenced by Vygotsky's theories, Lave and Wenger established their *theory of legitimate peripheral participation*. This theory argues that knowledge is distributed amongst a community of practice, and can only be understood with the 'interpretive support' provided by *participation* in the community of practice itself.

I will now explain what Lave and Wenger mean by *community of practice* and how they interpret *learning* within this theory.

Wenger characterizes a community of practice by its three dimensions:

- Mutual engagement (engaged diversity, doing things together, relationships, social complexity, community, maintenance);
- Joint enterprise (negotiated enterprise, mutual accountability, interpretations, rhythms, local response);
- Shared repertoire (stories, styles, artefacts, tools, actions, historical events, discourses, concepts).

A community of practice is not static, it is changing continuously: over time, forms of participation and identities change as newcomers themselves become ‘old-timers’ with respect to the next set of newcomers.

In this view, *learning* is about becoming, about participating in practices. It occurs through participation in activities and contributes to a growing identity within or across communities of practice. Identity is built around social engagement and is constantly being renegotiated as individuals move through different forms of participation.

Lave and Wenger's examples of learning in communities of practice are mostly located in 'real world' work situations, for example a group of claim processors in an insurance company. Practices here are beside other common activities defined by the practical work the members share. Those communities of practice differ from communities at universities as these are mostly defined by common or shared knowledge of its members and some researchers refer to them as *communities of knowledge*. Communities at universities are complex, and interaction and tensions between different communities can cause difficulties for members to establish their own professional identities.

The role of maths support centres is to support learning of mathematics at universities, between a variety of communities of knowledge, and to invigorate students' development of mathematical identity and mathematical confidence. How does learning or development of identity appear in such a community?

3. Learning in Communities of Mathematical Practice

Kirsti Hemmi considers a maths department as community of mathematical practice. This community consists of "*all people exercising and learning mathematics at the department of mathematics as members of a community of practice of mathematics. There are mathematicians, doctoral students, teaching assistants and students. It is a dynamic practice and the joint enterprise for all participants is the learning of mathematics in a broad sense*".

Examples for each of the three dimensions in a community of mathematical practice may be

- Mutual engagement: studying, teaching, learning, communicating maths;
- Joint enterprise: learning and developing the practice;
- Shared repertoire: courses, words, symbols, concepts, proofs, artefacts, or computers.

In a mathematical practice *learning* or *becoming knowledgeable* includes for example learning to talk about mathematics, using its language and symbols, learning to prove, or validating mathematics.

Mathematics and its applications as well as statistics are taught and used in diverse ways at varied communities in a university: for example students of Engineering or Commerce may have different attitudes towards these subjects than a student of Mathematics. While the latter may aim to become a mathematics researcher, the others aim to become sufficient applicants of mathematical methods. I use the term 'Communities of Mathematical Practice' to include a mathematics department as described by Hemmi as well as the intersection of mathematical practice with other university communities.

To promote learning of mathematics and development of mathematical identities in a broad sense, teachers of mathematics need to encourage students to participate in these communities of mathematical practice. In particular they need to provide opportunities for students to participate in these communities. Considering the abstract nature of the topic and the relatively large size of many mathematics classes at universities, this seems to be quite a challenge.

Wenger claims that in a Community of Practice newcomers need to have "*broad access to arenas of mature practice"*. "Mature practice'" at communities of mathematics practice include activities like mathematics research, communication with colleagues, participating at conferences, reading mathematical texts, or using mathematical methods to approach problems appearing in other fields such as industry or medicine.

However, typically students get to see mathematicians in lectures, where they might tell them about their work, and of course they see them teaching, but the students don't get to participate in other aspects of the mathematicians' work. Yvette Solomon describes a discussion of first-year undergraduate students' personal epistemologies of mathematics and mathematics learning. In her interviews with 12 first year university students the students described themselves as "*outside of the mathematics community. Their relationship with the lecturers required them to engage only with mathematics as already created rather than with the disciplinary process of creation and validation of knowledge, and their experience of missing explanations and exclusivity placed them on the periphery of the community*". Drawing on Wenger's theories, Solomon emphasises the importance of the interface between an individual's beliefs about the discipline and their self-positioning within the community, be it school classroom or university lecture theatre. Within this theoretical framework, she considers students' epistemologies as central to an analysis of the nature of undergraduate learning and the shifts that are required in moving from pre-university to university mathematics.

According to De Corte et al. the overall picture of students' beliefs of mathematics include:

- Mathematics is associated with certainty.
- Doing mathematics corresponds to following rules prescribed by the teacher.
- Knowing mathematics means being able to recall and use the correct rule.
- Mathematics becomes true when it is approved by the authority of the teacher.
- Mathematics problems have one and only one right answer.
- Mathematics is a solitary activity, done by individuals in isolation.

The following paragraph explains why maths support centres are seen as very valuable resources to provide opportunities for students to participate in communities of mathematical practice and to develop appropriate mathematical identities and epistemologies.

4. The role of Maths Support Centres in Communities of Mathematical Practice

It is no doubt a challenge to provide opportunities for incoming students to participate in communities of mathematical practice and to negotiate their beliefs of mathematics. However, I argue that maths support centres do provide such opportunities. Students are encouraged to explore mathematics, somewhat in a similar way as research mathematicians do their work.

For example in these centres students

- Talk, read and validate mathematics,
- Discuss different approaches,
- Use mathematical symbols and artefacts,
- Teach and learn from each other, and
- Share experience with 'oldtimers', i.e. PhD students or lecturers.

As the manager of the maths support centre at the National University of Ireland (NUI), Galway, I have observed the development of several students from first year at university to graduation and beyond who were very regular visitors of our centre. Some of them told me that the centre and people from the community played a major role in their development and decision for professional progression after graduating. It was the everyday exploring of mathematics within the community which inspired them to apply for PhD programmes in mathematics. This phenomenon is an important aspect of our support centre and the theory described above may provide a framework to describe and explain it.

While the evidence for this phenomenon is anecdotal at this stage, I will describe my subjective observations. The MSOR conference in Glasgow allowed me to discuss these ideas with colleagues to get an insight if they share these views and experiences and consider further investigations worthwhile.

The maths support centre at NUI Galway is on the ground floor of the building where all mathematics and statistics lecturers of the 'School of Maths' have their offices. The maths support centre is located between the School's office, a meeting room and the staff room and therefore even locally plays a central part in our community.

Over the years we observe students who come in during their first year at university, often not confident about their mathematical competence, looking for procedures or 'recipes' to approach their homework questions. Some of these students become regular visitors and often already after a few months significant changes of behaviour and attitudes towards approaching mathematics can be noted. Very quickly these students start working together and it is not unusual for tutors to hear "We may get back to you later, but would like to think about this ourselves first". Often finding a solution to a mathematical problem is not sufficient for them anymore, they start considering alternative methods. So changes to the above described beliefs about mathematics can often be identified after only a few months of visiting the support centre. There are several white boards in the support centre. Some of the students like to use these to explore their own ideas or to work in small groups. Several students who started studying mathematics with poor confidence, became almost daily visitors and applied for a PhD in mathematics after finishing their degrees. These students claim that they 'basically lived in' the maths support centre. Over the years they became more and more part of the community, took part at workshops and volunteered to help out in their final year. They got to know all our research students very well.

Of course, not all of our visitors aim to become mathematics researchers, and many students mostly use the maths support centre to get help with the requirements to pass their mathematics or statistics modules, however each year a significant number of students use the centre as an ideal environment to explore mathematics and develop problem solving skills.

In my view providing opportunities for participation in mathematical communities of practice and for development of mathematical identities are crucial aspects of maths support centres which may be underestimated. Here I have described my own subjective observations of a phenomenon. In case colleagues from maths support centres in other institutions share these views, a joint project to investigate further might be worth consideration.

5. References

De Corte, E. Op't Eynde, P. Depaepe, F and Verschaffen, L., 2010. The reflexive relation between students' mathematics related beliefs and the mathematics classroom culture. In L. Bendixen & F. Feucht (Eds.), *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice* (pp. 292-327). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511691904.010

Hemmi, K. 2006. *Approaching Proof in a Community of Mathematical Practice*. Doctoral Thesis, Stockholm University.

Lave, J. and Wenger, E., 1991. *Situated Learning: Legitimate peripheral participation.* Vol. 521423740. Cambridge University Press, Cambridge.

Solomon, Y., 2006. Deficit or difference? The role of students' epistemologies of mathematics in their interactions with proof. *Educational Studies in Mathematics* 61(373). https://doi.org/10.1007/s10649-006-6927-1. Kluwer Academic Publishers.

Wenger, E., 1998. *Communities of Practice: learning, meaning and identity.* Cambridge University Press, Cambridge.

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 $\frac{-2x^3-x^2-3x+2}{-x-3}$	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.