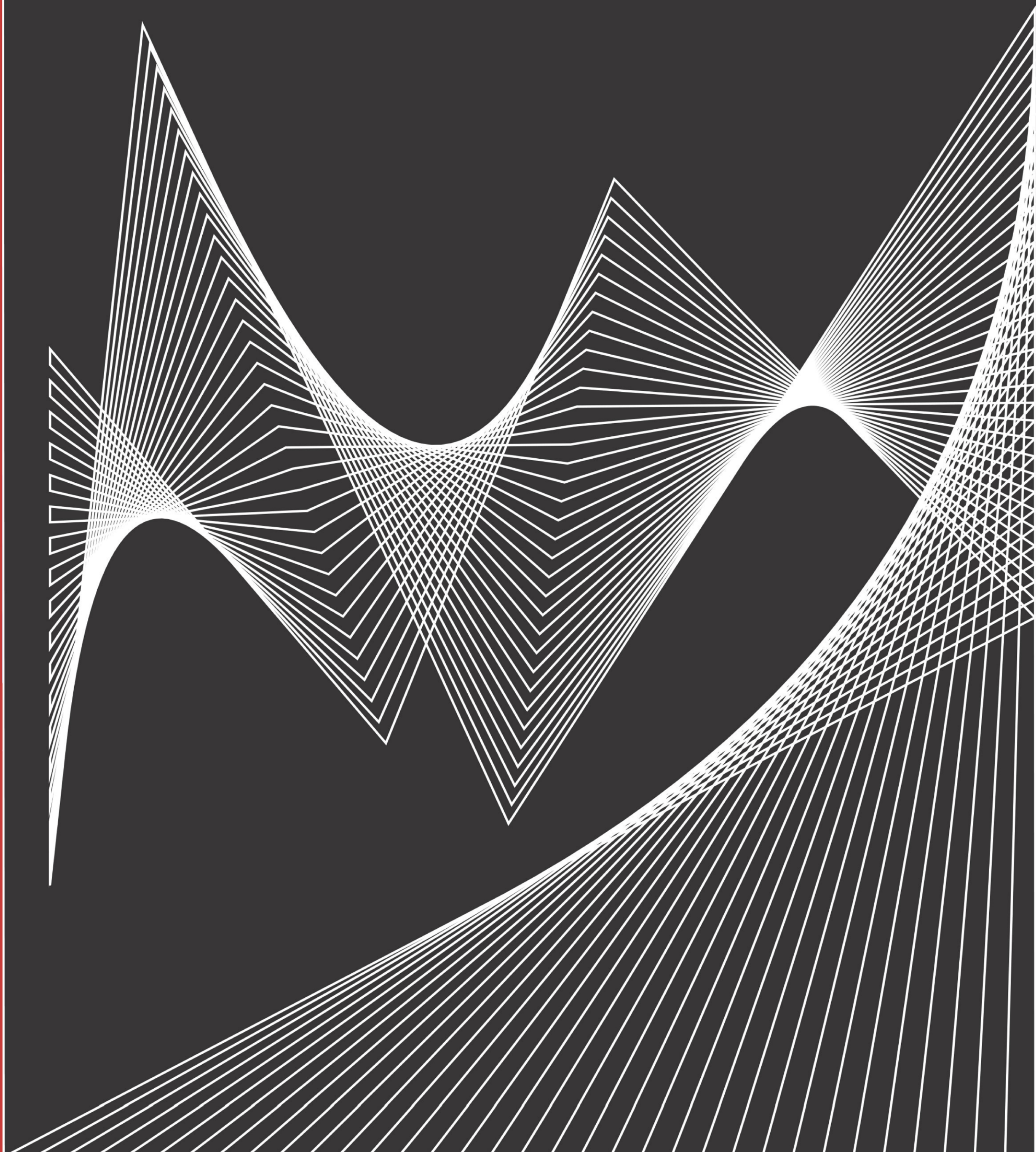
MSOR connections

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

Volume 15 **No.** 2

Special Issue: E-Assessment in Mathematical Sciences



Contents

Editorial – Peter Rowlett 3-4

E-Assessment in Mathematical Sciences (EAMS) Conference – Chris Graham and Christian Lawson-Perfect (EAMS Conference Organisers) 5

CASE STUDY: Using SOWISO to realize interactive mathematical documents for learning, practising, and assessing mathematics – André Heck 6-16

CASE STUDY: E-learning/e-assessment systems based on webMathematica for university mathematics education – Mitsuru Kawazoe and Kentaro Yoshitomi 17-24

CASE STUDY: Proctoring to improve teaching practice – Natasa Brouwer, André Heck and Guusje Smit 25-33

CASE STUDY: Using e-assessment to support flipped-style teaching – Karen Henderson 34-41

CASE STUDY: Numbas as an engagement tool for first-year Business Studies students – Tom Carroll, Deirdre Casey, Julie Crowley, Kieran F. Mulchrone and Áine Ní Shé 42-50

CASE STUDY: Creating statistics e-assessments using Dewis with embedded R code – Iain Weir, Rhys Gwynllyw and Karen Henderson 51-59

RESEARCH ARTICLE: STACK with state – Matti Harjula and Jarmo Malinen 60-69

RESEARCH ARTICLE: MathTOUCH: Mathematical Input Interface for E-Assessment Systems – Shizuka Shirai and Tetsuo Fukui 70-75

RESEARCH ARTICLE: A New Mathematics Input Interface with Flick Operation for Mobile Devices – Yasuyuki Nakamura and Takahiro Nakahara 76-82

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

Peter Rowlett, Mathematics, Sheffield Hallam University, Sheffield, U.K. Email: p.rowlett@shu.ac.uk

This special issue of *MSOR Connections* presents a selection of papers from the inaugural E-Assessment in Mathematical Sciences (EAMS) conference, held in September 2016 at Newcastle University. It is pleasing that there is enough activity in this area to support a two-day international conference on this topic. I am delighted that this issue offers international perspectives on the conference themes, including papers from Finland, Ireland, the Netherlands, Japan and the UK. Conference organisers Chris Graham and Christian Lawson-Perfect provide a brief overview of the conference.

Following the conference introduction, some case studies present approaches to conducting e-assessment in mathematical sciences.

SOWISO is used to create interactive online modules, combining e-assessment and instruction. Heck reports on the use of SOWISO for a course in basic mathematics for psychobiology students, which uses formative assessment and learning through worked examples, in which students reason by writing line by line equivalent expressions, with intelligent feedback. The paper includes an interesting analysis of the preparedness for and performance on the course of students with different entry qualifications.

Kawazoe and Yoshitomi outline the features of the system MATH ON WEB, used to deliver e-learning content and for e-assessment to encourage greater engagement with learning between classes, and describe its use with first year engineering students.

One process for conducting summative e-assessments is to use a computer room with restricted access rights to software and the web. However, there are circumstances where it makes good pedgagogic sense to allow students free(r) access to resources, or when practicalities mean that students cannot attend e-assessments on campus. Brouwer, Heck and Smit give three examples of use of online remote proctoring, using software that allows remote monitoring of students' use of computers.

Papers also focus on the use of e-assessment to facilitate teaching innovation. Henderson reports on the implementation of a flipped-style approach to a first year undergraduate calculus module, supported by formative e-assessment using Dewis. Carroll, Casey, Crowley, Mulchrone and Ní Shé describe the implementation of Numbas e-assessment to increase engagement by providing regular feedback for a large group of business studies students.

One theme of the conference was the presentation of innovative developments in what is possible with e-assessment.

Weir, Gwynllyw and Henderson report on an innovative development of the Dewis e-assessment system, which is capable of communicating with the R statistics package. This means greater sophistication in what can be assessed in statistics e-assessment, and an interface for developing Dewis e-assessments soley by writing R code.

One of the limitations of e-assessment, when compared to pen and paper assessment, is that it guides students and imposes a format on answers. The paper by Harjula, Malinen and Raslila describes an attempt to provide students with more choice when performing multi-step processes through a responsive interface. The paper is quite technical, but I would recommend even nontechnical readers to have a look at section 6 of the paper, which gives an example to show the potential of the system.

One difficulty using e-assessment is that of entering mathematics into computers or, especially, mobile devices. It is extremely cumbersome to type an expression such as '3*x^2-10*x-8' into a smartphone or tablet (which likely involves switching between multiple keyboards). Two possible solutions to this problem are presented. Shirai and Fukui have developed a HTML5 version of MathTOUCH, which accepts a colloquial text string and supports translating this into mathematical notation through a series of menus. Nakamura and Nakahara have developed a novel interface for inputting mathematical expressions using a nested keyboard with flick operation, a method similar to that used to input Japanese characters on a smartphone.

I enjoyed attending the conference, presenting at it and hearing the variety of work that was demonstrated, as well as editing this set of conference proceedings. I hope you will find this selection of papers offers an interesting taste of the nearly 30 presentations at that conference. As mentioned by Graham and Lawson-Perfect in their conference introduction, slides from the conference and videos of talks are available via eams.ncl.ac.uk.

I would like to thank the authors, peer reviewers and conference organisers for their hard work and for sticking firmly to a series of deadlines, which made editing this issue a smooth process.

E-Assessment in Mathematical Sciences (EAMS) Conference

Chris Graham, School of Mathematics & Statistics, Newcastle University, Newcastle, U.K. Email: christopher.graham@ncl.ac.uk
Christian Lawson-Perfect, School of Mathematics & Statistics, Newcastle University, Newcastle, U.K. Email: christian.perfect@ncl.ac.uk
EAMS Conference Organisers

The inaugural E-Assessment in Mathematical Sciences (EAMS) conference was held in September 2016 at Newcastle University. This two-day conference brought together researchers and practitioners in the field of mathematical e-assessment and was attended by over 70 delegates from all corners of the globe. Motivated by a desire to bring together projects and bubbles of development around the world, around 25 speakers gave a mixture of presentations and workshops.

The conference was opened with a keynote talk by Christian Lawson-Perfect (Newcastle University), who spoke about the student user experience with e-assessment. Other keynote speakers were Mike Gage (University of Rochester), founding developer of the WeBWorK system, Chris Sangwin (University of Edinburgh), who explored the automatic assessment of proof, and Sally Jordan (The Open University), who closed the conference with a talk encouraging delegates to reflect on their own experiences of e-assessment and a look ahead to the future of this field.

Speakers travelled from as far as Japan and South Africa to share their experiences, with insightful talks covering topics such as: flipped classrooms, distant learning, the transition to higher education, applications and development of e-assessment tools and mathematical input interfaces.

As well as the conference proceedings herein, slides and screencasts have been made available on the EAMS conference website via eams.ncl.ac.uk.

CASE STUDY

Using SOWISO to realize interactive mathematical documents for learning, practising, and assessing mathematics

André Heck, Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, The Netherlands. Email: a.j.p.heck@uva.nl

Abstract

Successes of computer aided learning in mathematics education have led to high expectations of teachers and learners. A teacher would like to be able to review what students actually do when learning online, how much progress they make, which difficulties arise during learning, and to which subjects attention must be paid in the next lessons. Learners expect that they can consult their instructional materials anywhere, anytime and device-independent, and that they receive immediate and intelligent feedback on their work. At the Faculty of Science of the University of Amsterdam we realize the envisioned interactivity and support in SOWISO. This is a cloud-based environment in which a course has the look-and-feel of an interactive module: texts, randomised examples and randomised exercises with automated feedback are integrated, and online formative and summative assessments are fully supported. We describe and evaluate a SOWISO-based course in basics mathematics for first-year psychobiology students, in which formative assessment and learning through worked examples are key elements of the instructional design.

Keywords: interactive mathematics documents, assessment driven learning, formative assessment, summative assessment, intelligent feedback.

1. Interactive mathematical documents

An interactive mathematical document can be understood as a collection of mathematical pages containing theoretical explanations, examples, exercises, and the like, delivered to the user through hypermedia. A fixed hierarchical structure is not required: users may take their own paths through the pages and they may study the same page, but get different examples and exercises. In the last decades, various systems offering interactive mathematical documents have been developed (Sangwin, 2013), ranging from homework and assessment systems like Maple T.A., STACK, WeBWorK, or Numbas, and notebooks and worksheets offered by systems like *Mathematica*, Maple, or Sage, to intelligent tutoring systems like ActiveMath. SOWISO differs from the aforementioned systems that it integrates instruction, practice and assessment in the form of an interactive online module.

The main characteristic of digital environments for assessment driven learning of mathematics is that each step in the learning path begins with a randomised question that a student is supposed (to try) to answer. If (s)he does not master it (completely), the environment often gives a hint or offers the option to show a worked example or to guide the student in a step-by-step approach. Hereafter the parameter-based implementation of the problem allows the loading of a new version of the problem for the student to demonstrate the newly acquired mastery. SOWISO extends these features with generous possibilities to give a student feedback during the process of completing a task, not only telling whether the given answer to a question was right or wrong, but also whether it meets requirements regarding the mathematical shape of a mathematical formula or the precision of a numeric answer, and what the mistake precisely was in case of an incorrect answer. Immediate feedback allows students to use the online mathematics document as a worksheet in which they progress line by line towards the final solution of a problem, while receiving feedback at each step. An example of a simplification task is shown in figure 1. The left-hand side of the equations are always given as a prompt and not entered by the student. The student reasons by equivalence, that is, writes line by line equivalent expressions, starting with the given expression and ending with the simplified result in the requested form. Automated, intelligent feedback at each step guides the work.

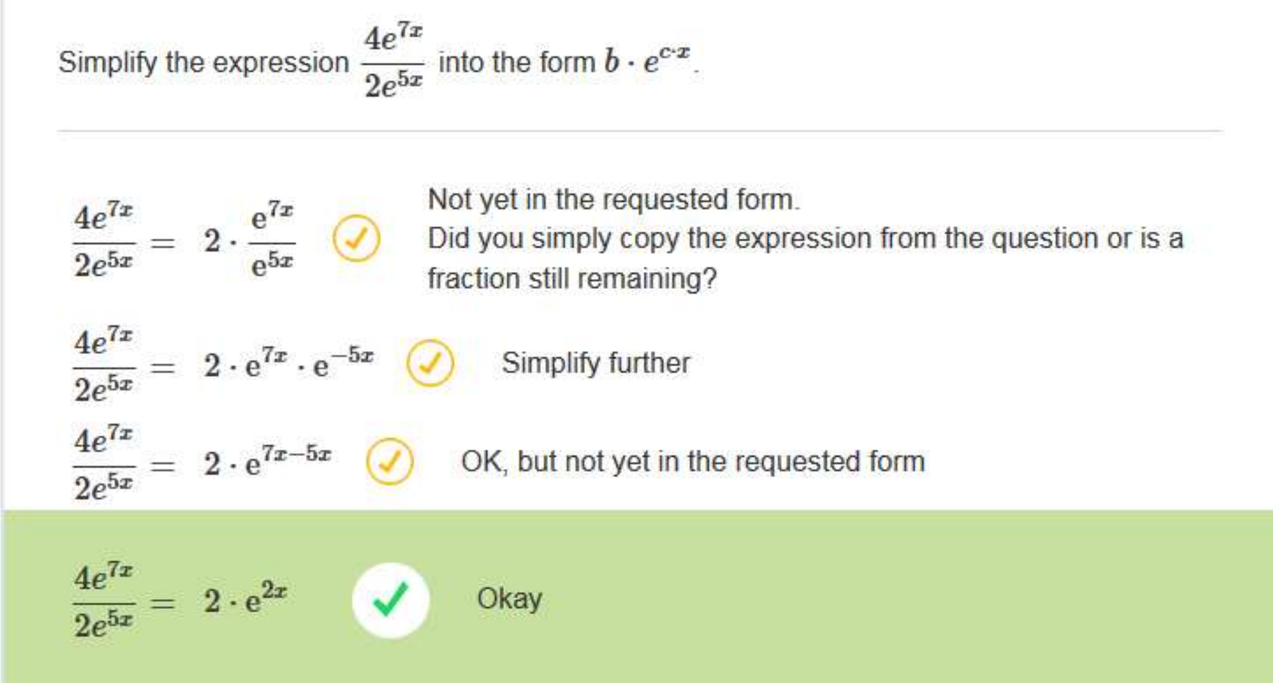


Figure 1. Use of the online document as a worksheet to solve a simplification task in small steps with intermediate feedback.

We speak of intelligent feedback when it is based on an analysis of the student's answer or action, and when it goes beyond a simple reaction as correct/incorrect. Intelligent feedback is a detailed reaction based on (1) expert knowledge about the mathematical content; (2) a model in which common mistakes and actions are classified; and (3) knowledge about learning and instruction so that the most suitable feedback can be chosen. Part of the feedback design of a prototypical integration example is listed in table 1.

Table 1. Feedback whilst calculating the antiderivative of $\frac{1}{5}x^9$.

Answer	*Feedback*	*Explanation*
$\frac{1}{50}x^{10}+c$	Well done	Correct answer
$\frac{1}{5}\left(\frac{1}{10}x^{10}+c\right)$	Correct, but simplify further.	Unusual correct answer that can be simplified
$\frac{1}{50}x^{10}$	Don't forget the constant of integration	No constant of integration
$\frac{1}{50}x^{10}+\frac{1}{5}$	This is a particular solution, but other values than 1/5 can also be a constant of integration.	No general solution
$\frac{9}{5}x^{8}$	You computed the derivative instead of the antiderivative	Reading error?

Because one does not know in advance which formula a student enters, a computer algebra system is used as back engine to recognize mathematical expressions with the same semantic meaning and to respond with appropriate feedback. It is important to realize that equivalence of learner's input with the answer preferred by the lecturer is not good enough to label the answer as correct: when asked to compute $\frac{1}{4}+\frac{1}{6}$, the answer $\frac{5}{12}$ is certainly correct and maybe $\frac{10}{24}$ is acceptable, but one can frown one's eyebrows with answers such as $\frac{1}{12}+\frac{1}{3}$, $\frac{1}{4}+\frac{1}{6}$, and $\frac{30}{72}$. Thus, it is important that automatic simplification of the back engine can be dealt with and that one can give appropriate feedback for different cases and at different stages of instruction.

Each randomised exercise can also be used as randomised example, so that a student can go through as many examples as needed. Figure 2 shows three instances of one and the same randomised example on powers.

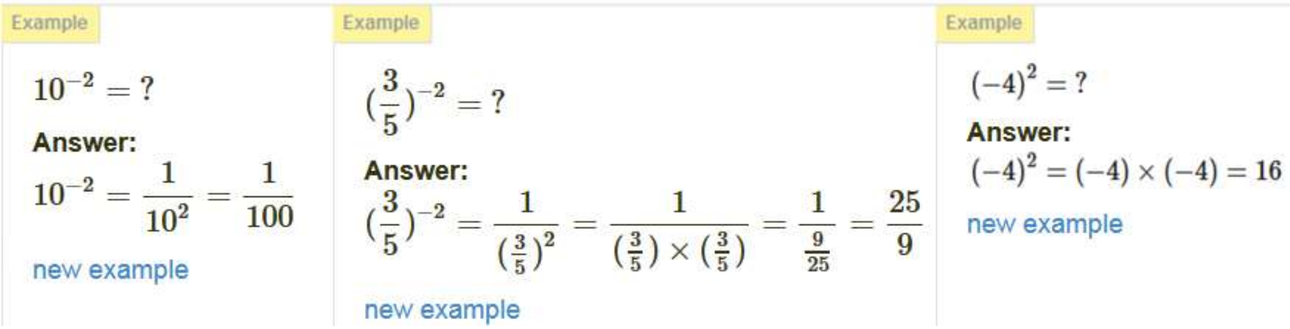


Figure 2. Repeated worked examples originating from one template exercise.

Learning from examples and example-problem pairs have been selected as the main mode of learning. Cognitive Load Theory (Kalyuga, 2011) has underpinned that use of worked examples is effective under certain conditions and is preferred by novices. A computer environment can provide the learner upon demand with as many good quality examples as needed and at the right level of detail to match the learner's needs.

SOWISO offers many tools to implement the instructional design principles associated with effective examples based learning and tutoring. It is a multimedia authoring environment to create interactive, mathematical, HTML5 compliant texts that can be viewed in a standard web browser on any device. A conversion into generated course notes such as PDF, Word or ePub format is supported. JavaScript can be used to add interactivity to a web page: examples are JSXGraph elements for dynamic mathematics (Gerhäuser, et al., 2011) and EjsS simulations of computer models (Saenz, et al., 2015). The dynamic graph of the logarithmic function $log_g(x)$ shown in figure 3 allows students to explore how the graph and the properties of the logarithmic function depend on the value of the base. A video clip from Mathcentre (Matthews, Croft and Lawson, 2013) extends the written explanation. Extra information can be folded behind paragraphs and opened by clicking the 'More' button. Students can rate theory pages and ask questions about theory pages or worked-out solutions of exercises on the users' forum of the class.

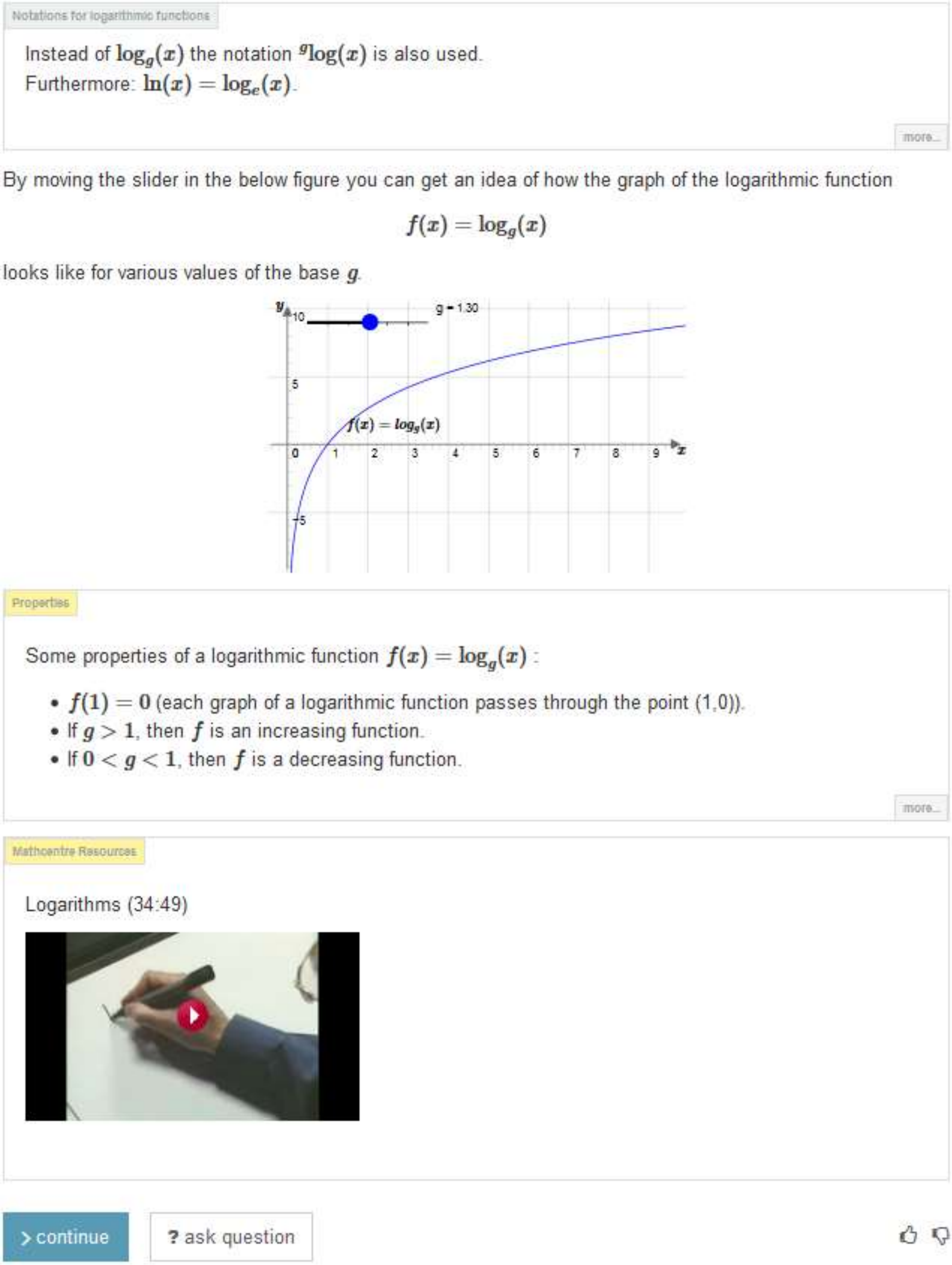


Figure 3. Course page, illustrating some online features of SOWISO.

2. Case study: Basic mathematics for psychobiology

In this paper, experiences with the ‘Basic Mathematics for Psychobiology’ course in the first year of the psychobiology bachelor programme at the University of Amsterdam are described. It concerns the course settings in the study year 2015-2016, course results, and dispositional learning analytics (Buckingham Shum and Deakin Crick, 2012), in which learning data and learner data are combined as a means to provide instructors actionable feedback.

2.1. Course settings

One hundred and fifty-six first-year students (38 male, 118 female) participated in this course of ten weeks. The course design was assessment driven and examples based instruction with intelligent feedback to acquire mathematical cognitive skills needed by psychobiologists. It was the second study year that Basic Mathematics was taught without being embedded in some neuroscience course. Course evaluation and research done in the previous year by Hoek (2015) showed that students had difficulties with providing correct and useful self-explanations while studying worked examples, and could not always recognize the important differences between an exercise and a worked example that they used before. This had resulted in adaptations of the online module. The course content was reduced, and video clips and more detailed explanations in worked examples were added. An example of an annotated solution to an exercise is shown in figure 4.

Example

Calculate the following integral via the substitution rule:

$$\int_4^9 \frac{1}{x+\sqrt{x}}\,dx$$

Answer:
We apply the substitution rule for integration with $u=\sqrt{x}$.
By differentiating u we find $du=\frac{1}{2\sqrt{x}}\,dx$, that is, $dx=2u\,du$.
The integration bounds change to $\sqrt{4}=2$ and $\sqrt{9}=3$.
So

$$\int_4^9 \frac{1}{x+\sqrt{x}}\,dx=\int_2^3 \frac{2u}{u^2+u}\,du$$

substitution rule with $u=\sqrt{x}$ and $du=\frac{1}{2\sqrt{x}}\,dx$, that is, $dx=2u\,du$

$$=\int_2^3 \frac{2}{u+1}\,du$$

simplification

$$=\Big[2\ln(u+1)\Big]_2^3$$

primitive of the integrand

$$=2\cdot\ln(4)-2\cdot\ln(3)$$

substitution of integration bounds

$$=\ln\left(\frac{16}{9}\right)$$

simplification

new example

Figure 4. Annotated solution to an exercise, with verbose explanations of each step in blue.

The online module of this course in SOWISO is available in Dutch and English at the website uva.sowiso.nl/auth/login (login name and password EAMS2016). The contents concern calculus and its applications in psychobiology: (1) Precalculus – calculating with numbers, calculating with letters, solving equations; (2) Chemical calculations; (3) Functions – all standard functions; (4) Differentiation and derivatives; (5) Differentials and integrals; (6) Complex numbers; (7) Models of growth – exponential growth, limited exponential growth, logistic growth; and (8) Ordinary differential equations – basic notions, separable 1st-order ODEs and linear 2nd-order ODEs with constant coefficients.

Treatment of each mathematical subject ended with a randomised digital assessment for which the pass mark was 7.5 (out of 10). Besides having a summative role, the main purpose of these tests was assessment for learning: students could objectively verify their learning progress in mastering a mathematics subject. These tests were short, but there were many. In total, students had to take forty-nine tests. Only in case they passed all tests, they received a course mark, regardless of positive results on the final examination. The final exam took place in a digital examination room (figure 5). The exam was taken in SOWISO. The questions were randomised having two versions of the same difficulty. Students solved the exercises by pencil and paper, entered their final answers in the SOWISO test, and submitted both their papers and online test. The digitalisation of the exam not only reduced the correction process, but also allowed students to compare their answers with the worked-out solutions and made the review process before the examination easier and more effective.

Figure 5. Picture of the digital examination room with psychobiology students at work.

2.2. Course results

Data of student activity were collected by logging students’ actions, their performance in the digital learning environment in formative assessments, and from evaluation reports of the students. Personal data of students such as mathematics background, initial level of mathematical competence, mathematics anxiety, test anxiety, and motivation and engagement, were collected in SOWISO using validated questionnaires (Hopko, et al., 2003; Spielberger, 1980; Martin, 2007).

By the mathematics background of the students we mean the mathematics examination programme that students took at upper pre-university level, namely, Mathematics A or B. Mathematics A is meant for students who prepare for academic studies in social or economic sciences, or other related studies. Its core subjects are statistics and probability and a little calculus. Emphasis is on applying mathematics. Mathematics B is the mathematics needed for exact sciences and technical studies, and its core component is calculus. In addition to the difference in core mathematical subjects the purpose and type of problems of the two curricula differ. The original intentions for Mathematics A, based upon realistic mathematics education, were (HEWET, 1980, p.19): “Mathematics A is intended for students who will have little further education in mathematics in their academic studies, but who must be able to use mathematics as an instrument to a certain extent. In particular, we have in mind those who have to prepare themselves for the fact that subjects outside the traditional sciences are more frequently being approached with the use of mathematics.” However, soon after its introduction in 1989, the initial objectives were not met anymore in the central Mathematics A final examinations and nowadays open ended problems generally tend to ask of the students only to do a calculation, draw a graph, read a graph, substitute a few values in a formula, and so on, and do not really assess modelling and problem solving competencies.

A significant percentage (30%) of the first-year psychobiology students had taken at secondary school the Mathematics A examination programme, which prepared them less for exact sciences than the Mathematics B examination programme taken by the rest of the first-year students. This became obvious when the exam results of the basic mathematics course were inspected. Only 19% of the students with Mathematics A background passed the exam, whereas 69% the students with Mathematics B background were successful; see the cross tabulation in table 2. In figure 6, three histograms of exam marks are plotted: the left, middle, and right diagram show the distribution of marks for all first-year students with known mathematics background (107), for first-year students with Mathematics A background (26) and for first-year students with Mathematics B background (81), respectively. On the left-hand side and in the middle there appear to be binormal distributions of marks, whereas it becomes a normal distribution for Mathematics B students only. These diagrams suggest that there was a split between the students based on their mathematics background, which can almost be described by combining the distributions for the two groups. Indeed, when the Mann-Whitney statistic was calculated to determine whether there was any statistically significant difference in exam marks regarding the mathematics background ($U = 1752$, $z = 5.10$, $p < 0.001$), a statistically significant difference between students with Mathematics A and Mathematics B was found, with large effect size (0.49). A cross tabulation (table 2) found that Mathematics B performed better than the Mathematics A students on the final exam. From the diagnostic entry test in the first week of the course we already knew that students with Mathematics B background entered the course with a statistically significant higher competence level than their peers with Mathematics A background. The mathematical knowledge and skills gap between these two groups of students did not get smaller during the course.

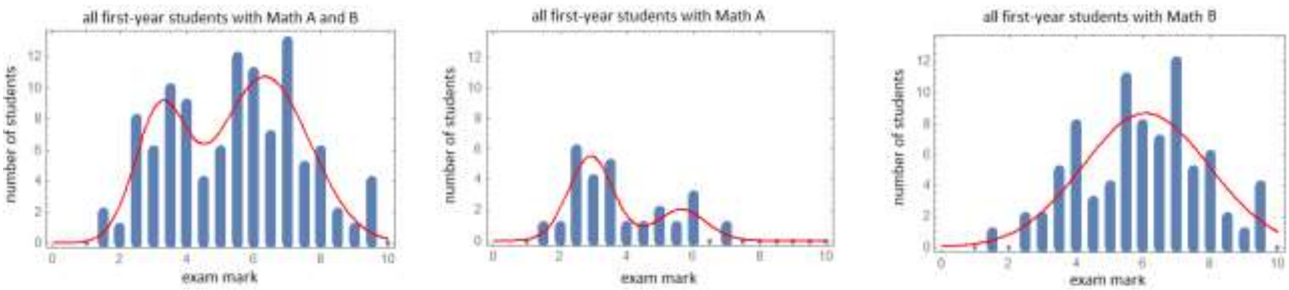


Figure 6. Distribution of exam marks: a binormal distribution for all first-year students with known mathematics background (left) and for first-year students with Mathematics A background (middle) and a normal distribution for first-year students with Mathematics B background (right).

Table 2. Cross tabulation of pass-fail exam results and mathematics background of the students.

	pass	*fail*	*total*
Mathematics A	5 (19%)	21 (81%)	26 (24%)
Mathematics B	56 (69%)	25 (31%)	81 (76%)

2.3. Mathematics anxiety and test anxiety

The level of mathematics anxiety amonst first-year students was measured via the Dutch translation of the Abbreviated Math Anxiety Scale (Hopko, et al., 2003). The mean AMAS score (M = 2.9, SE = 0.43) of 155 students was low on a scale from 1 (no anxiety at all) to 10 (panic) and in agreement with students' self-estimates. We calculated the Mann-Whitney statistic to determine whether there were any statistically significant differences in mathematics anxiety regarding mathematics background and gender. No significant differences were found at the significance level of 0.05 (U = 2491, z = 1.12, p = 0.263 for gender and U = 1806, z = 1.76, p = 0.079 for mathematics background). This is in agreement with result collected in previous study years in life sciences and in agreement with the data and results of Erurtan and Jansen (2015), who found in their studies on the relationship between children's emotional experiences of mathematics and their mathematics performances that only differences in test anxiety (boys had lower test anxiety than girls) but not in mathematics anxiety appeared amongst Dutch pupils in grades 3-8 (ages 7-15 years).

The level of test anxiety amonst first-year students was measured via the Dutch translation of the Test Anxiety Invenory (Spielberger, 1980). The mean TAI score (M = 3.7, SE = 0.96) of 153 students was low on a scale from 1 (no anxiety at all) to 10 (panic) and in agreement with students' self-estimates. We calculated the Mann-Whitney statistic to determine whether there were any statistically significant differences in test anxiety regarding mathematics background and gender. No significant difference was found at the significance level of 0.05 (U = 1915, z =0.98, p = 0.328) for mathematics background. But a statistically significant gender difference in TAI scores (U = 2887, z = 2.97, p = 0.003) was found, with moderate effect size (0.24). Male students had lower test anxiety than female students. This is in agreement with results collected in previous study years in life sciences and in agreement with the results of Erurtan and Jansen (2015).

The AMAS scores were significantly but moderately related to the TAI scores (Spearman's rho r_s = 0.56, p (two-tailed) < 0.001, with bias corrected and accelerated bootstrap 95% confidence interval [0.40, 0.69]). This is not surprising because AMAS has items (in the mathematics evaluation anxiety component of the scale) that are comparable with items in TAI. Concerning the relationship between emotional experiences of mathematics and mathematical performance, only mathematics anxiety was statistically negatively related with exam scores at the significance level of 0.05, but it was a weak correlation (Spearman's rho r_s = -0.16, p (one-tailed) = 0.04, with bias corrected and accelerated bootstrap 95% confidence interval [-0.33, 0.02]).

2.4. Dispositional learning analytics

The 'Motivation and Engagement Wheel' framework (Martin, 2007) includes both thoughts and behaviours that play a role in learning and consequently in course performance. Both are subdivided into adaptive and maladaptive forms. Adaptive thoughts (the 'boosters') consist of self-belief (SB), value of school (V), and learning focus (LF), whereas adaptive behaviours consist of planning (PLN), task management (TM), and perseverance (P). Maladaptive thoughts (the ‘mufflers’) include (test) anxiety (A), failure avoidance (FA), and uncertain control (UC), and lastly, maladaptive behaviours (the ‘guzzlers’) include self-sabotage (SS) and disengagement (D). These behaviours and thoughts are assessed via the ‘Motivation and Engagement Scale – University / College’ (MES-UC) self-report instrument. One hundred and fifty-four (154) students filled out the MES-UC questionnaire at the start of the course. Cronbach’s alpha for the instrument was 0.76, close to the literature value of 0.78. No statistical significant differences between students with Mathematics A and B background were found on all scales of the MES-UC questionnaire, except for test anxiety: students with Mathematics A background scored higher on this scale. Gender differences were found: female students scored statistically significantly higher on boosters and guzzlers (in particular on planning, study management and test anxiety), but the effect size was small.

The effect of adaptive and maladaptive behaviours and thoughts on the activity level of students in the online module and on the course performance measured by the mark on the final exam was investigated. Activity level was measured by the number of theory pages that a student has viewed and the number of exercises that a student had tried to solve (by entering at least one response). Figure 7 represents the correlations between the MES-UC scales and the exam mark. It shows that self-belief, persistence and planning had the largest positive impact on the exam mark. As expected, maladaptive thoughts and behaviours (especially self-sabotage) had a negative impact.

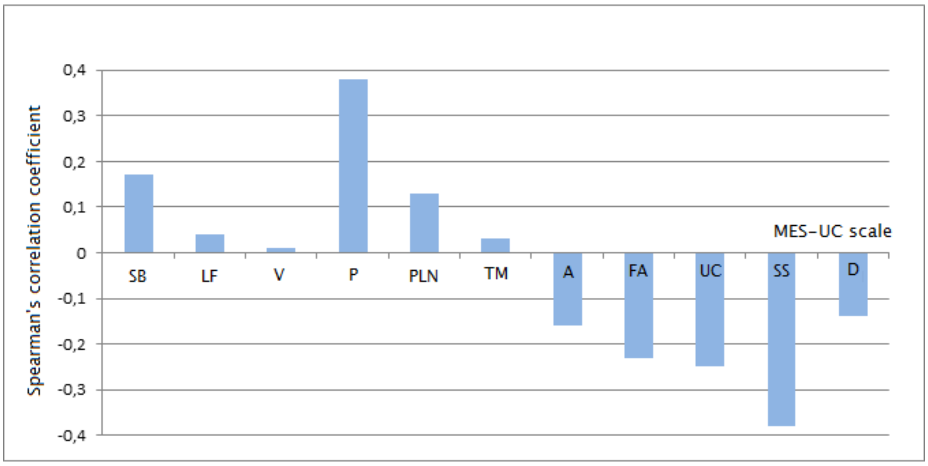


Figure 7. Correlation between MES-UC scales and the exam marks.

Figure 8 represents the correlations between the MES-UC scales and the activity level of a student. As expected, adaptive behaviours had a positive impact on the activity of students in the module. But adaptive thoughts apparently had less impact; self-belief may even have a negative impact. Anxiety and uncertain control had a positive impact on the activity level of students, whereas self-sabotage had amongst the maladaptive behaviours the strongest negative impact on the student’s activity level. These findings are in good agreement with results found in other studies, for example in mathematics and statistics courses for first-year economy students (Tempelaar, Rienties, and Nguyen, 2016). When the Jonckheere-Terpstra test (N = 121) was calculated to determine whether there was a statistically significant difference in exam scores and level of practise (with three categories of ‘no drop out’, ‘drop out as soon as new mathematics was taught’, and ‘no or irregularar participation’), a statistically significant difference ($J(2)$ = 334, z =5.47, p < 0.001) was found, with large effect size (0.50), between students who regularly practised and did not drop out, and those students who dropped out when it became difficult because of new mathematical subjects or who practised irregularly or not at all during the course. Students with Mathematics A background dropped out significantly more than students with Mathematics B Background, maybe because they were insufficiently prepared for the course. In summary, one could conclude that looking only at worked-out solutions does not lead to better marks, but that sufficient practising and stamina really helps to be successful.

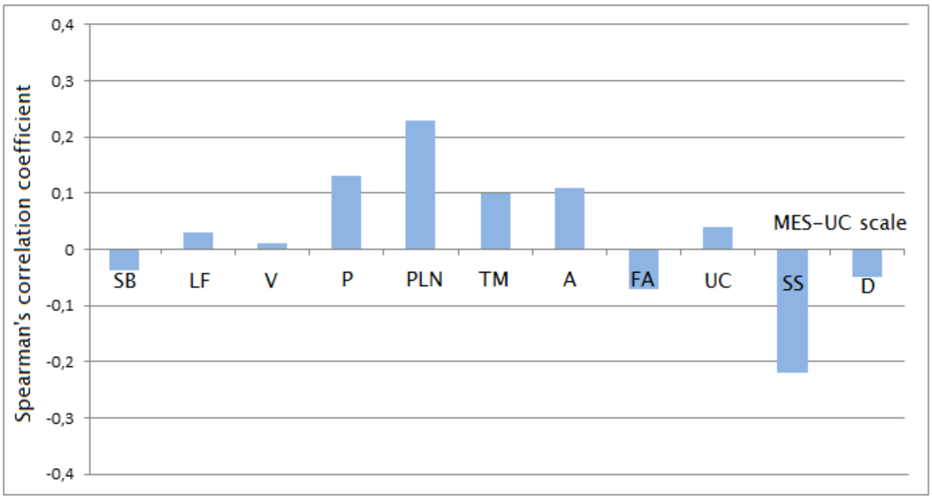


Figure 8. Correlation between MES-UC scales and the student activity level.

3. Conclusions

This case study and a similar case study about a mathematics course for computer science students (Heck and Brouwer, 2015) illustrate that the SOWISO environment offers an e-learning environment for mathematics in which students

- practise more than in a classical instructional setting;
- get automated feedback on their work without direct involvement of a teacher or tutor;
- use interactive material for self-study and practise mathematics;
- ask questions on the forum and communicate with peers;

and in which teachers

- create interactive online course notes (re-using instructional materials);
- monitor the students' level of activity and learning progress;
- carry out digital assessments (formative and summative).

More research is needed to find out if only procedural knowledge and skills have been mastered, or if problem solving skills have been developed, how long the acquired competencies retain, and whether also transfer of mathematical competencies has been reached. Anyway, this course has proved to many of the students that they are able to master mathematical knowledge and skills needed by psychobiologists. At least, when the students have a proper mathematics background: in this study we found that mathematical knowledge and skills gap between students with Mathematics A and B background was already significant at the start of the mathematics course and did not improve during the course. We also found that mathematics anxiety and test anxiety were low amongst the Dutch students, but female students reported a slightly higher level of test anxiety. The correlation between mathematics anxiety and mathematics performance was significant but weak in this study. No statistically significant relationship was found between test anxiety and exam performance. Dispositional learning analytics provided actionable feedback: in this study we learned for example that we should stimulate students (1) to actively use the interactive mathematical documents and do the exercises instead of only reading the text; and (2) not to rely on just one example, but to use more examples and read the theoretical explanations about them.

4. References

Buckingham Shum, S. and Deakin Crick, R., 2012. Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics. In: *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge*. New York: ACM Press. pp.92-101. Available at: http://oro.open.ac.uk/32823/1/SBS-RDC-LAK12-ORO.pdf [Accessed 9 October 2016].

Erturan, S. and Jansen, B., 2015. An investigation of boys' and girls' emotional experience of math, their math performance, and the relation between these variables. *European Journal of Psychology of Education, 30*(4), pp.421-435.

Gerhäuser, M., Valentin, B., Wassermann, A. and Wilfahrt, P., 2011. JSXGraph – dynamic mathematics running on (nearly) every device. *Electronic Journal of Mathematics and Technology*, 5(1).

Heck, A. and Brouwer, N., 2015. Digital assessment-driven examples-based mathematics for computer science students. In N. Amado & S. Carreira (eds.), *Proceedings of the 12th International Conference on Technology in Mathematics Teaching* (ICTMT12). University of Algarve, Portugal. pp.403-411. Available at http://hdl.handle.net/10400.1/6081 [Accessed 9 October 2016].

HEWET (1980). HEWET-Report (in Dutch) from the 'werkgroep van advies voor de herverkaveling eindexamenprogramma's Wiskunde I en Wiskunde II vwo [Working group for advice on the reallocation of the examination programmes Mathematics I and II VWO] (1980). The Hague: Staatsuitgeverij.

Hoek, R., 2015. *Setting the Right Example*. Master thesis, University of Amsterdam. Available at: http://www.scriptiesonline.uba.uva.nl/document/634733 [Accessed 9 October 2016].

Hopko, D.R., Mahadevan, R., Bare, R.L. and Hunt, M.K., 2003. The abbreviated math anxiety scale (AMAS). *Assessment*, *10*(2), pp.178-182.

Kalyuga, S., 2011. *Managing Cognitive Load in Adaptive Multimedia Learning*. Charlotte, NC: Information Age Publishing.

Martin, A.J., 2007. Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology,* 77(2), pp.413-440.

Matthews, J., Croft, T. and Lawson, D., 2013. Mathcentre 2013. *MSOR Connections*, 13(1), pp.34-40.

Saenz, R., Esquembre, F., Garcia, F.J., de la Torre, L., and Dormido, S., 2015. An architecture to use Easy Java-JavaScript Simulations in New Devices. *IFAC-PapersOnLine*, 48(29), pp.129-133.

Sangwin, C., 2013. *Computer Aided Assessment of Mathematics*. Oxford: Oxford University Press.

Spielberger, C., 1980. *The Test Anxiety Inventory*. Palo Alto, CA: Consulting Psychologist Press.

Tempelaar, D.T., Rienties, B., and Nguyen, Q., 2016. Achieving actionable learning analytics through dispositions. *IEEE Transactions on Learning Technologies* [Accepted].

CASE STUDY

E-learning/e-assessment systems based on webMathematica for university mathematics education

Mitsuru Kawazoe, Faculty of Liberal Arts and Sciences, Osaka Prefecture University, Osaka, Japan. Email: kawazoe@las.osakafu-u.ac.jp
Kentaro Yoshitomi, Faculty of Liberal Arts and Sciences, Osaka Prefecture University, Osaka, Japan. Email: yositomi@las.osakafu-u.ac.jp

Abstract

After-class learning is quite important to understand and master college mathematics. To promote students' after-class learning, we developed the web site "MATH ON WEB", which currently consists of two systems; Web-based Mathematics Learning System (WMLS) and Web-based Assessment System of Mathematics (WASM). Both systems have been developed with web*Mathematica*. WMLS is a system for assisting students' self-learning and WASM is a system for an online assessment. We implemented blended learning environments with these systems and more than 600 students at our university learn with the systems every year. In this article, we summarize the feature of the systems and report the result of our teaching practice for first-year engineering students with these systems.

Keywords: university mathematics education, e-assessment, e-learning, CAS.

1. Background and motivation

1.1. Role of e-learning/e-assessment systems for university mathematics education

After-class learning is quite important to understand and master college mathematics. Traditionally, after-class learning at university level has been promoted by giving students paper-based exercises. Such traditional after-class learning has some problems. If paper-based exercises are given as homework, teachers have to mark students' results. Otherwise, students want to know answers for the exercises. Each mathematics class at university in Japan includes 60-80 students. So marking students' homework every week is hard work for teachers. Should teachers give answers to students? Giving answers to students is also problematic. In college mathematics, there are many mathematics problems whose correct answers are not uniquely determined. In linear algebra, many problems have infinitely many correct answers. For such problems, students cannot check their answer even if teachers give answers to them. Another problem is that students cannot have any support during their work at home. Many students do their homework in the evening, night, or weekend. Therefore, when students have questions about how to solve problems at home, they can neither ask their teachers nor obtain real-time feedback.

E-learning is a useful tool to overcome the above issues and to build an effective after-class learning environment. We can use e-learning system to give and mark homework. Using e-learning system, we can develop a "homework system" without a heavy workload for teachers. Teachers cannot help students overnight, but e-learning systems can help students all day, even at midnight. And e-learning systems enable students to determine whether their answers are correct or not, even when there are infinitely many correct answers for the mathematical problem.

There is another merit of e-learning. To understand and master college mathematics, it is important to understand the meaning of mathematical concepts deeply. And to do so, it is important for students to explore various examples. E-learning is a good solution to develop effective interactive learning materials which promote students' understanding of mathematical concepts.

1.2. Why is CAS needed?

In college mathematics, there are many mathematical problems whose correct answer is not unique. For an example, let us consider the following problem: "Find a parametric representation of the plane in space given by the equation $x + y + z = 1$". Let $(a_1, a_2, a_3), (b_1, b_2, b_3), (c_1, c_2, c_3)$ be any three points, which satisfy the equation but are not on the same line. Then, we always get the correct answer for the problem as follows:

$$(x, y, z) = (a_1, a_2, a_3) + s(b_1 - a_1, b_2 - a_2, b_3 - a_3) + t(c_1 - a_1, c_2 - a_2, c_3 - a_3)$$

where s and t are any real numbers. Conversely, let us consider to determine whether the following student's answer is correct or not.

$$(x, y, z) = (p_1, p_2, p_3) + s(u_1, u_2, u_3) + t(v_1, v_2, v_3)$$

How can we do that? Since the correct answer is not unique, pattern-matching methods are not applicable. To check the above student's answer, it is needed to verify the following conditions:

- $(x, y, z) = (p_1, p_2, p_3)$ satisfies the equation $x + y + z = 1$;
- $(x, y, z) = (u_1, u_2, u_3)$ and $(x, y, z) = (v_1, v_2, v_3)$ satisfy the equation $x + y + z = 0$;
- The two vectors (u_1, u_2, u_3) and (v_1, v_2, v_3) are non-zero and non-parallel.

A computer algebra system (CAS) allows us to verify them easily. Finding a basis for the column space of the given matrix A is another example. For the second example, we need a more complicated algorithm to check students' answers.

Another reason is that CAS enables us to develop a simulation type of e-learning content. Using rich functions of CAS, we can develop effective interactive contents with which students can learn the meaning of mathematical concepts. We give an example of such content in section 2.2.

2. MATH ON WEB: webMathematica-based e-learning/e-assessment systems for university mathematics education

We developed e-learning/e-assessment systems for university mathematics education and offer them through the website "MATH ON WEB". Currently, the website consists of the two systems; Web-based Mathematics Learning System (WMLS) and Web-based Assessment System of Mathematics (WASM). WMLS is a system for assisting students' self-learning; It consists of the drill section and the simulation section (Kawazoe, Takahashi and Yoshitomi, 2013). WASM is a system for online assessment; It consists of online assessment materials associated to learning units (Kawazoe and Yoshitomi, 2016). Both systems have been developed with web*Mathematica* and cover all topics in the standard courses of calculus and linear algebra for the first-year university students.

2.1. WMLS: Drill section

The drill section of WMLS offers an online mathematics exercise environment. It has a large number of mathematics problems in calculus and linear algebra courses for the first-year university students. Actually, it has more than 200 learning units and each unit consists of five problems. Figure 1 shows an example of contents. For each unit, the five problems are presented in a fixed order and students have to solve them correctly in the order. Each ordered list of problems has been implemented carefully from the educational viewpoints. Only when a student solves a problem correctly, he/she can go to the next problem. After when all the five problems are correctly solved by a student, a mark indicating the "completion" of the learning unit appears in his/her learning summary page. There are three types of mark indicating the status of the task: *completed*, *suspended*, and *given-up*. The system shows various data to teachers. Teachers can see the learning status of students in their mathematics class.

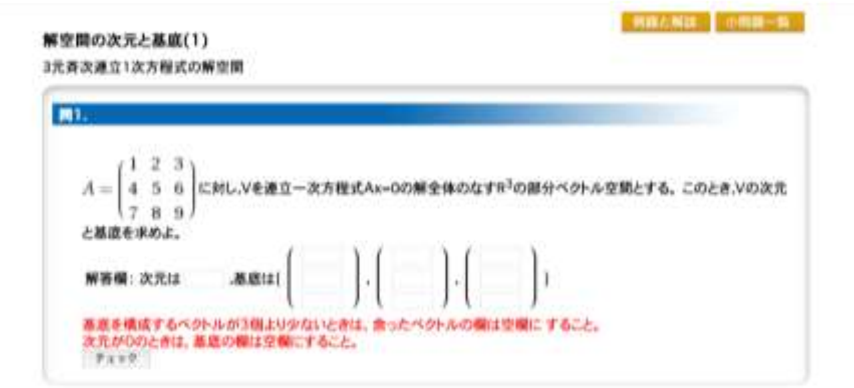


Figure 1. An example from the drill section of linear algebra (Kawazoe et al., 2013)

Translation: “For the matrix A, let V be the subspace of R^3 consisting of all solutions of $Ax = 0$. Find the dimension of and a basis for V.”

The problem data consists of a problem template (Text), a problem seed (a list of mathematical expressions in *Mathematica* form; sometimes given as a *Mathematica* program which generates mathematical expressions), an answer column (HTML form), an answer analyser (a *Mathematica* program which analyses students’ answers and identifies types of errors), a list of feedback messages (CSV), and a problem example with its model answer and a guide for how-to-solve (PDF). Each element in a list of feedback messages corresponds to the error type, and the system shows an appropriate feedback message to a student depending on the identified error type. Figures 2 and 3 show how the system works. Table 1 shows examples of feedback messages for the problem in figure 1. *Mathematica* programs analysing students’ answers are written as “Which statement” of Mathematica, and error analysis is done by first-match method. There is no time limit for solving problems. The correct answer is never shown. Students can retry the same problems repeatedly until they get correct answers, reading textbooks, notebooks, model answers and guides for how-to-solve, or asking friends and the teacher.

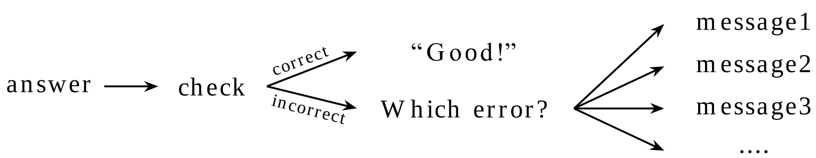


Figure 2. The flow of analyzing students’ answer (Kawazoe et al., 2013)

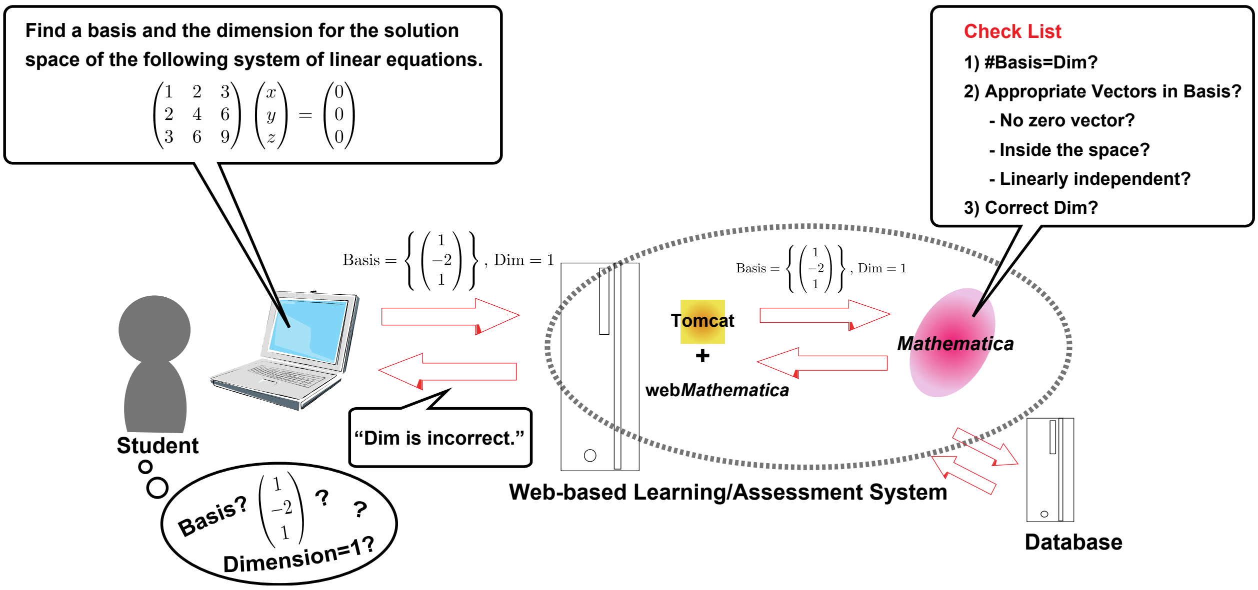


Figure 3. The basic structure of the drill section of WMLS

Table 1. Examples of feedback for the problem in figure 1

Student' answer	Feedback message
Basis= $\left\{\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\}$, Dim=2	The number of vectors in Basis is not equal to Dim. See the definition of basis and dimension.
Basis= $\left\{\begin{pmatrix}1\\-2\\1\end{pmatrix},\begin{pmatrix}0\\0\\0\end{pmatrix}\right\}$, Dim=2	Basis contains the zero vector. See the definition of basis.
Basis= $\left\{\begin{pmatrix}1\\-2\\1\end{pmatrix},\begin{pmatrix}2\\-4\\2\end{pmatrix}\right\}$,	Vectors in Basis are linear dependent. See the definition of basis.
Basis= $\left\{\begin{pmatrix}1\\-2\\1\end{pmatrix},\begin{pmatrix}2\\-1\\0\end{pmatrix}\right\}$,	Good!
Basis= $\left\{\begin{pmatrix}3\\0\\-1\end{pmatrix},\begin{pmatrix}0\\3\\-2\end{pmatrix}\right\}$,	Good!

2.2. WMLS: Simulation section

The simulation section of WMLS offers a simulation type of e-learning contents. A simulation type of content consists of interactive learning material to learn a mathematical concept and assessment material to assess students' understanding. Students learn mathematical concepts with a simulation type of content by varying parameter values and observing the results (figure 4). Assessment materials are the same as in the drill section.

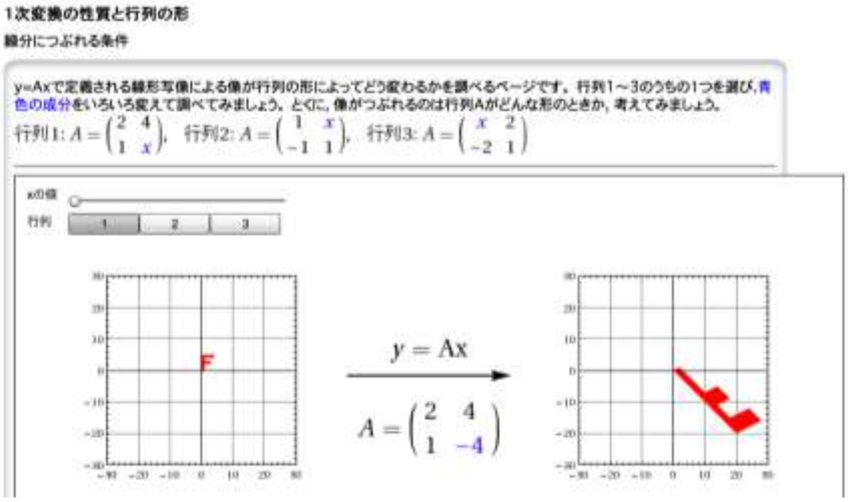


Figure 4. An example from the simulation section of linear algebra (Kawazoe et al., 2013)

Translation: "When does the linear transformation $y = Ax$ map the figure to a line segment?"

2.3. WASM: Assessment materials

WASM has two different modes: Assessment mode and Exercise mode. The exercise mode is almost the same as the drill section of WMLS. The main difference is that problems are presented to students in completely random order or randomly generated by the *Mathematica* program. Using the exercise mode, students can practise solving problems included in an assessment test before they try the assessment. Online assessment materials are associated to learning units and three levels of achievement: *basic*, *standard*, and *advanced.* An assessment test associated to a learning unit consists of multiple problem modules that are presented in random order. WASM has currently 140 problem modules. In the assessment mode, students need to solve each problem within the time limit. We show an example of assessment materials in figure 5.

Figure 5. An example of assessment materials in calculus

Translation: "Find an antiderivative of the following function."

In WASM, JavaScript-based popup keyboards based on jQuery Keypad (Wood, 2014) assist students to input their answers. In most CAS based e-learning/e-assessment systems, it is inconvenient that students have to input their answers in the input format of CAS. Actually, many of them have difficulty in inputting their answers into the answer fields. Since required symbols and functions differ with problems, the keyboard is provided in different layout depending on the problem (figure 6).

Figure 6. A popup keyboard for the problem in figure 5

In WASM, a tool for content developers has been improved in comparison with WMLS. A problem text, a problem seed, an answer column in HTML-form, a *Mathematica* program for analysing students' answers, a list of feedback messages and so on can all be edited online, while these data have to be uploaded as files separately by content developers in WMLS. In WMLS, handling and managing of variables in input forms of an answer column had been inconvenient to content developers. In WASM, all variables in input forms can easily be managed and handled online in a developer's environment.

2.4. Some remarks on the history of the systems

The project started from 2001 and the first version of MATH ON WEB started with simulation-type materials in 2002 (Kawazoe et al., 2003) and then drill-type materials were added in 2004. All materials are written in Japanese, but all of them are open to the public from the beginning of the project. In the academic year 2005, MATH ON WEB was officially released to all students at our university. Funded by the Japanese government, these materials have been developed as WMLS in 2009. Funded by the government again, WASM has been developed and added to MATH ON WEB in 2013.

3. Effectiveness of the blended learning class with our systems

3.1. A new blended learning class with WLMS and WASM

In the fall/winter semester of the academic year 2014/2015, we implemented a blended learning class with WMLS and WASM for the linear algebra course of the first-year engineering students. We had already implemented blended learning classes with WMLS (Kawazoe, Takahashi and Yoshitomi, 2013), but a blended learning class with both WMLS and WASM was a new approach. The class was designed as follows. Every week, students were recommended to use WMLS for after-class learning and their achievement was assessed with a mini-exam (paper-based test) in the next lesson. For a student who failed the mini-exam, a supplementary exam with WASM was imposed to him/her after class. We illustrate the design in figure 7.

3.2. Result of the questionnaire survey

At the end of the semester, we conducted a questionnaire survey about WASM in the class (63 students in total) and 56 students answered it. The questionnaire consisted of multiple choice questions; especially for measuring the usability and the usefulness of the system, we used a Likert scale. 94.6% of the 56 students answered that they used WASM during the semester. In the questionnaire, we asked the additional questions to those students who used WASM during the semester. When asked for the reason that they used WASM, 84.9% of them answered that they used it because supplementary exams were imposed, while four students answered that they used it because they wanted to check their understanding. On the modes they used, 49.1% of them answered that they only used the assessment mode, while 47.2% of them answered that they used both the assessment mode and the exercise mode. Only two students answered that they used only the exercise mode. On the usability of the system, they answered "very easy" (11.3%), "somewhat easy" (50.9%), "neutral" (18.9%), "somewhat difficult" (11.3%) and "very difficult"

(7.5%). On the usefulness for developing mathematical skills and self-evaluating of understanding, they answered “very useful” (18.9%), “somewhat useful” (64.2%), “neutral” (11.3%), “somewhat unuseful” (5.7%) and “very unuseful” (0%). We also asked them which supplementary work they prefer: a paper-based homework or a WASM exam. For that question, 58.5% of them answered that they prefer a WASM exam, while 26.4% of them answered that they prefer a paper-based homework. The remaining 15.1% of them answered that they prefer to use both works. The results indicate that this new blended learning environments using both WMLS and WASM is preferable to students.

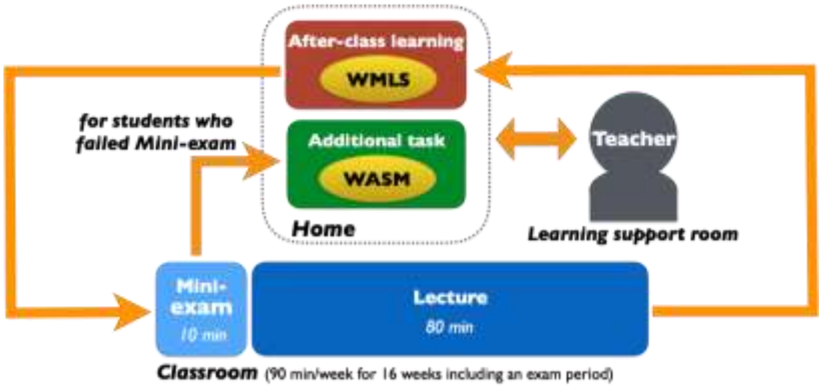


Figure 7. A new blended learning class with WMLS and WASM

4. Future works

4.1. Development of Moodle plugin of WMLS and WASM

WMLS and WASM are standalone all-in-one systems and isolated from any existing learning management system (e.g., Moodle). Most universities have learning management systems (LMS) and use them to enrich students’ learning environments and improve the educational qualities. If WMLS and WASM are seamlessly connected to LMS, then the mathematics teaching and learning environments with these systems become much more confortable to students and teachers. To develop a connection to LMS is one of the next goals of our research project. We plan to develop Moodle plugin of WMLS and WASM. The plugin is planned to be able to handle the actual WMLS/WASM problem data. The prototype version of the plugin was developed in 2016 (Nakahara, Yoshitomi and Kawazoe, 2016). We are planning to release the first version of the plugin in 2017.

4.2. Development of a contents sharing framework among heterogeneous systems

E-learning is undoubtedly a powerful tool for university mathematics education. Many systems have been developed and used at many universities in many countries. However, the development of learning content is a heavy workload for content developers. At least in Japan, content developers are usually mathematics teachers. Thus it is very hard for them to develop sufficient quantity of learning content. This situation seems to prevent e-learning becoming widespread. There are some attempts to share content, but they are only for users of the specific system: e.g., Maple T.A. Cloud by Maplesoft (2016), mathbank.jp (Taniguchi et al. 2016) for STACK. To overcome the above issue, we proposed to develop a contents sharing framework among the different mathematical e-learning/e-assessment systems (Kawazoe and Yoshitomi, 2013). As for the structure of content data in the e-learning/e-assessment systems for university mathematics education, the essential part of it is almost the same, we believe. Actually, at least between STACK and our systems, we can manually translate the question/problem data from each other (Yoshitomi, 2013). This fact implies the essential design of question/problem data is mutually compatible in heterogeneous systems. We started the research project called MeLQS (the Mathematics e-Learning Question Specification) to this aim. It is a collaborative research with other researchers who use Maple T.A. or STACK. The goal of the project is to develop a framework to share the MeLQS formatted data via a cloud service, which could be exported to as many as possible heterogeneous systems. In the near future, we hope, will come an era when teachers need only select the question design and use them in their own systems.

5. References

Kawazoe, M., Kimura, E., Okada, M., Okamoto, M., Seta, K. and Takahashi, T., 2003. Hypothesis-driven learning system for college mathematics. *JSiSE Research Report*, 18(4), pp.27-32.

Kawazoe, M., Takahashi, T. and Yoshitomi, K., 2013. Web-based system for after-class learning in college mathematics via computer algebra system. In M. Inprasitha ed. 2013. *Proceedings of the 6th East Asia Regional Conference on Mathematics Education*, Vol.2. Phuket, Thailand: Khon Kaen University, pp.476-485.

Kawazoe, M. and Yoshitomi, K., 2013. On the framework of database for e-learning contents of mathematics. *JSiSE Research Report*, 28(1), pp.23-28.

Kawazoe, M. and Yoshitomi, K., 2016. [Construction and operation of mathematical learning support web-site MATH ON WEB developed by webMathematica]. *Bulletin of JSSAC*, 22(1), pp.13-27, (in Japanese).

Maplesoft, 2016. Maple T.A. - Online Testing and Assessment System for STEM Courses - Maplesoft. Available at: http://www.maplesoft.com/products/mapleta/ [Accessed 14 October 2016].

Taniguchi, T., Nakamura, Y. and Nakahara, T., 2016. Mathbank.jp. Available via http://mathbank.jp [Accessed 14 October 2016].

Nakahara, T., Yoshitomi, K. and Kawazoe, M., 2016. Development of a Mathematica-based Moodle plugin for assessment in mathematics. *JSiSE Research Report*, 31(2), pp.15-16.

Osaka Prefecture University, 2016. MATH ON WEB Learning College Mathematics by webMathematica. [online] Available at: http://www.las.osakafu-u.ac.jp/lecture/math/MathOnWeb/ [Accessed 14 October 2016].

Wolfram Research, Inc., .2016. Mathematica (11.0). Champaign, IL.

Wood, K., 2014. jQuery Keypad (2.0.1). [computer program] Available at: http://keith-wood.name/keypad.html [Accessed 13 October 2016].

Yoshitomi, K., 2013. On a formulation of “Mathematics e-Learning Contents Specification” and its applications to some systems. In Japanese Society for Information and Systems in Education, *Proceedings of the 39th Annual Conference of Japanese Society for Information and Systems in Education,* pp. 167-168. Available at: http://www.jsise.org/taikai/2014/program/index.html [Accessed 13 October 2016].

CASE STUDY

Proctoring to improve teaching practice

Natasa Brouwer, University of Amsterdam, Faculty of Science, Amsterdam, Netherlands. Email: natasa.brouwer@uva.nl
André Heck, University of Amsterdam, Faculty of Science, Amsterdam, Netherlands. Email: A.J.P.Heck@uva.nl
Guusje Smit, University of Amsterdam, Faculty of Science, Amsterdam, Netherlands. Email: G.L.H.Smit@uva.nl

Abstract

Universities organise digital summative assessments in special assessment computer rooms and they conduct strong restrictions on the resources. This policy assures safety and security of assessments but obstructs authentic problem solving where open resources are needed. A digital assessment room is also not a proper solution when students for some reason cannot attend the assessment on campus. We present three scenarios from the science and mathematics teaching practice at the University of Amsterdam in which we have used proctoring to create the flexibility of online exams that was needed. Online remote proctoring of computer screens on the campus and/or students at home enabled (1) more authentic exams by allowing open resources, (2) use of own laptops (BYOD) in a regular lecture room, (3) off campus online exams as a part of an online programme, for example, at home anywhere in the world. We have shown that proctoring is not just a different way to control students using computers, but that it enabled improvement of the quality of learning processes in which summative assessments are an integral part.

Keywords: e-assessment, online remote proctoring, summative assessment, higher education.

1. Introduction

Assessment is an indispensable part of any education. It is widely accepted that assessment in a course should be aligned with intended learning outcomes and with the learning process (Biggs, 1996) and that assessment should be adapted when the learning process changes. E-assessment, also called Computer-Assisted Assessment (CAA) enables frequent testing of knowledge and understanding of students. This also raises fundamental questions about the role of the assessment in the education process as a whole (Conole and Warburton, 2005). E-Assessment for learning or online formative assessment is used in courses to support students in self-regulated learning and constructing understanding. In this case, the pedagogical role of assessment shifts from a teacher-centred instructive role to a student-centred supportive role and in the technology perspective shifts assessment from a rigid time and space setting toward a fully flexible one. Heck and Brouwer (2015) studied performance of students using online examples-based mathematics formative assessments through which students regulated their own learning process. The authors found a large impact on academic performance of students in the Numerical Recipes Project course in which ninety percent of the students passed the exam after examples-based and assessment-driven teaching and learning. In higher education usually at the end of the course or at several pre-defined moments within a course a summative test is taken to assess whether students have reached the desired learning outcomes. When digital formative assessment is used in a course it is natural to also take the exam online in a similar design as during the course. When students use computers for problem solving, computer skills need to be assessed as well. In the course Business and Enterprise (Cornock, 2016) the assessment was entirely through coursework tasks.

In any summative assessment security is very important. Online examinations at the universities are taken in specially designed and secured digital examination rooms. Apampa, Will and Argles (2009) defined security goals that are specific for e-assessments. They showed that besides confidentiality, integrity, and availability, the electronic presence security has to be seen as one of the security goals. Specific cheating problems in e-assessment and counteractions have been listed by Rowe (2004). One of the cheating problems is unauthorised help during the e-assessment. In digital examination rooms, students cannot communicate via the computers. All internet connections through which students could reach unauthorised help outside the room or social media are blocked. The websites that are allowed or the computer software need to be whitelisted. In courses in which students work on open problems and use computer software of their own choice, whitelisting of resources is almost impossible.

For many students, online courses create new possibilities for personal development or for a career switch. Students can combine work and study, and they can live far away from the university campus. Coming to the campus to take an exam can be extremely time-consuming and expensive. But taking off campus online exams worries faculties because of the risk of cheating. Fask, Englander and Wang (2014) designed an experiment to assess the difference in student performance between students taking a traditional exam in an examination room and those taking an unproctored exam online. They found no significant difference between online and class exam scores. They found evidence that disadvantages of online assessment offset opportunities for unproctored students to cheat.

2. Three cases of proctoring

In this paper we describe three different educational settings in which a digital examination room was not a suitable place to take a digital summative assessment. Online remote proctoring was used to assure security of these assessments.

We have used two commercial applications, namely ProctorExam Pro and ProctorExam Light (www.proctorexam.com). ProctorExam Pro is an application where a student is observed by a live remote proctor via the webcam of the laptop and a smartphone put behind the student. The computer screen of the student was observed, too (figure 1). All three sources were video recorded, which made it possible to double check suspected fraud marked by the proctor.

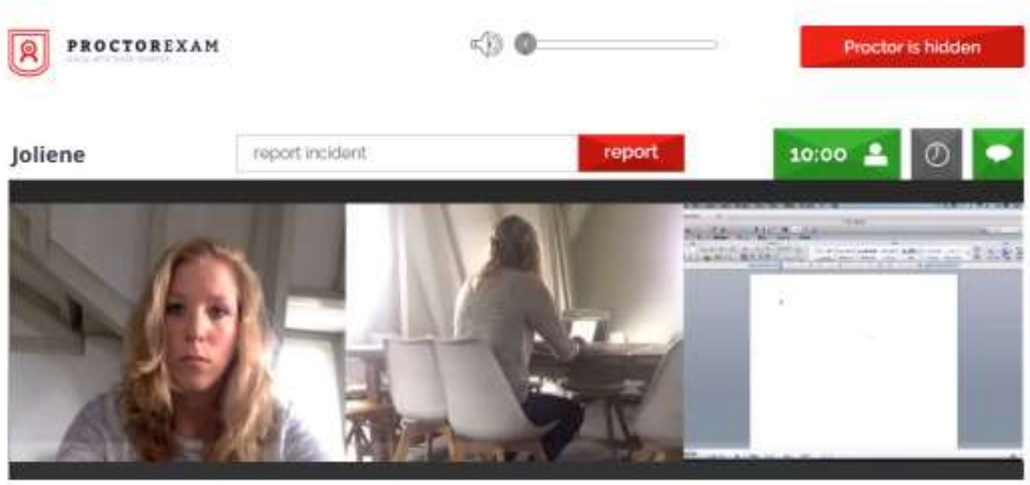


Figure 1. Proctor’s view of a student taking an exam using ProctorExam Pro (image from instructional material). *Left*: laptop webcam, *Centre*: smartphone camera image, *Right*: computer screen recording.

ProctorExam Light was used to only observe the computer screen of the students. After the examination, the screen recording was checked at higher speed in order to detect any unauthorised actions.

Back office we monitored on a dashboard (figure 2) when students connected to the proctor online application and registered if the connection was lost during the examination.

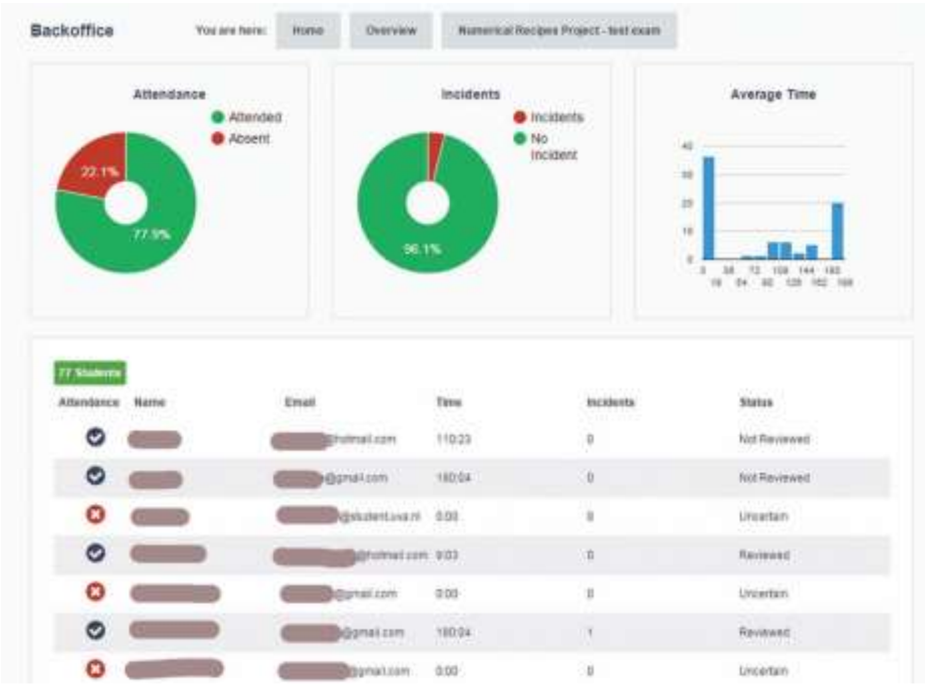


Figure 2. ProctorExam dashboard: report of an exam

Suspected fraud incidents were marked in the proctoring report and the videos of all students' assessments could be watched herein as well (figure 2). An incident was not always a case of cheating: it could also be an innocent deviation in the behaviour of a student or a technical issue from a small disturbance in the internet connection. The lecturer had to evaluate whether an incident was fraud or not. In our cases, no incident indicated cheating.

The students who took a digital assessment with online remote proctoring followed a three step model (figure 3). First, they had to sign a privacy agreement about sharing personal data. Next they had to install a ProctorExam browser add-on on their computers and pass a technical test to assure the required technical standard of their equipment (laptop, webcam, audio, smart phone and internet connection) met the security conditions for the examination. The technical test was taken several weeks before the examination to give students enough time to fix the equipment if necessary. A student could not take part in the examination until the technical requirements were satisfied. In step 2 (figure 3) the online exam was mimicked. A dummy assessment was used to experience the online examination as realistically as possible. This diminished unnecessary stress during the examination. Students who had already taken an off campus online exam could skip step 2. According to Dutch law regarding protection of personal data students could refuse online remote proctoring for personal reasons. Thus we have always arranged a possibility to take the exam on the campus. Some students chose for a pencil and paper exam and others preferred an extra invigilator to be present and observe their computer screen from behind during the whole exam.

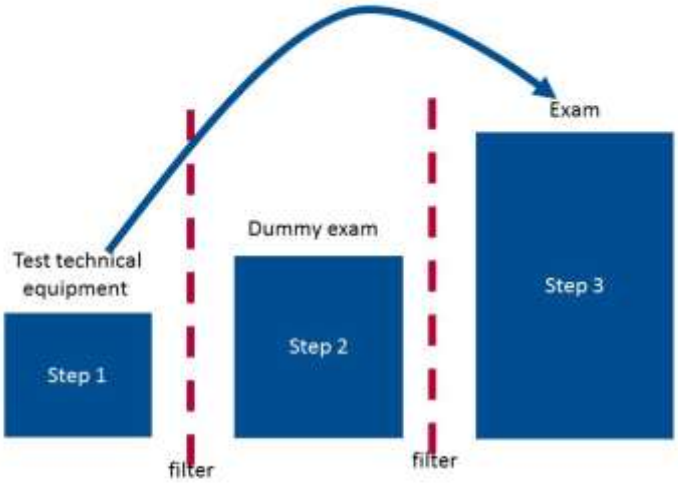


Figure 3. Three step model of digital assessment with online remote proctoring

2.1. Case 1: digital exam with open resources

Title of the course: Operating Systems.
Student population: 137 first-year students in the Computer Science bachelor.
Course design features: problem solving, authentic problems, open resources. Students could use any resources they wanted, including search engines such as Google.

The lecturer designed an open resources exam with problem solving questions. This kind of exam could not be organised in digital examination rooms on the university campus for the following reasons:

- whitelisting of all possible resources that students could use was impossible;
- even if a whitelist were possible, allowing wide access to resources while preventing contact with each other or people outside the class via the internet was impossible;
- observing individual screens by invigilators in the room was not feasible. A privacy protecting foil (fraud prevention) prevented reading the computer screen from an angle. This also prevented the invigilators to notice any unauthorised actions like using chat boxes or social media;
- software for online remote proctoring was not available in digital examination rooms.

As a workaround the lecturer organised the digital exam in a regular computer room at the faculty and he used online remote proctoring of computer screens during the whole exam (figure 4).

The lecturer monitored the digital exam in the lecture room on his laptop using ProctorExam back office tool. This way he could see which students logged in to the digital assessment and if all of these students were connected to the online remote proctoring app. No additional invigilators or other support staff was needed during this exam.

This situation had also several drawbacks. The exam had to be scheduled in several sequential groups because the room was too small for the whole group. On the computer screens there was no privacy foil. The students had to be seated further away from each other. To prevent cheating in the room each student got a different set of questions, which costed the lecturer more time to prepare the assessment.

Figure 4. Digital exam using ProctorExam Light. The lecturer is answering a question of a student.

2.2. Case 2: digital exam using Bring Your Own Device (BYOD)

Title of the course: Numerical Recipes Project.
Student population: 77 second-year students in the Computer Science bachelor.
Course design features: learning based on worked-out examples and formative assessments. Students worked on their own laptops (BYOD) in SOWISO, a cloud-based environment for learning, practising, and assessing mathematics (www.sowiso.com). Most of the face-to-face activities were tutorial sessions in parallel groups. Each week students had to pass several assessments. They could take them anywhere, any time within the deadline and as often as they wanted. Every time a different set of questions appeared. Passing the assessments was obligatory to get the final mark for the course but no credits were given. More details about the course design can be found in Heck and Brouwer (2015).

The lecturer designed two summative assessments for a final mark in SOWISO on BYOD, just as students were used to working during the course. This could not be organised in the digital examination rooms of the university for three reasons:

- it was too expensive because the size of the digital examination room did not match the size of the student population taking the exams;
- it was obligatory to use the available computers there;
- there was no WiFi and no sockets for BYOD power supply.

The lecturer organised the summative assessments in the regular lecture rooms where students could use their laptops (figure 5). Online remote proctoring was used and all laptop screens were captured in order to prevent fraud. The students started the assessment at the same moment in several parallel lecture rooms. The students could use pencil and paper to work out the assignments before they filled in their final answers in the SOWISO assessment. The digital exam was automatically marked. The lecturer checked by hand only the erroneous responses to assign partial points if the answers were partly correctly worked out on paper. In practice, the availability of the digital final answers more or less halved the correction work by the teacher and offered students the possibility to compare their answers with the worked-out solutions.

Figure 5. Students taking an e-assessment in SOWISO on their laptops. The lecturer observes the students from the last row.

2.3 Case 3: online pre-master programme

Programme: Pre-master Information Studies, five online courses given three times a year.
Target group of students: students with knowledge deficiencies who wished to start a masters' programme Information Studies. An intake procedure assured that a level of pre-education of applicants was appropriate to start the pre-master programme.
Number of students: about 100 students per year
Programme design features: fully online courses on Blackboard (university virtual learning environment). The course design of four courses was test-based learning using adaptive release of lesson material in Blackboard (figure 6).

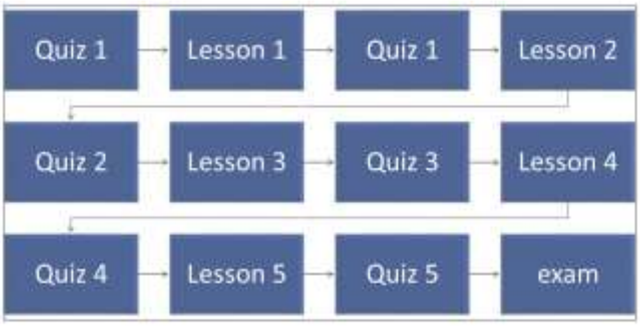


Figure 6. Test-based learning course design of Pre-master Information Studies courses

Students had to pass each quiz with a score of at least 80% before the next lesson would open. A student could repeat a quiz as often as (s)he wanted. Every time a different set of questions was displayed in the quiz. In one course (Academic Skills) the adaptive release was less strictly applied. In each online course there was a moderator who supported students in their learning.

The final exam was taken online in Blackboard. The students could take the exam at home using online remote proctoring (ProctorExam Pro) or come to the university campus. The final exam of the Pre-master Information Studies courses was not organised in digital examination rooms of the university for two reasons:

- the students were located all over the world. It would be very expensive to come to the campus only to take the exam. Some of the students would first have to arrange an entrance visa before they could come to take the exam;
- also for most Dutch students it was more convenient to take the exam at home. Thus the group of students who preferred to take the exam on campus each time was very small (between 3 and 8 students). One invigilator could assure security of the exam in a small regular computer room on campus.

3. Experiences with online remote proctoring

The experiences of the lecturers with online remote proctoring were positive. They could create flexibility for their summative assessments that would not have been possible otherwise. All lecturers were concerned about the reliability of online remote proctoring to assure security. They found it very important to have full insight in the data of the proctoring. The lecturers would have appreciated to get a more explicit and clearer report about the observed incidents so that they could analyse it more efficiently. In our case, they first had to watch all video fragments to find out what exactly the proctor had marked as incidents before they could decide if it was necessary to start any procedures about cheating.

Five to ten percent of the students per course refused proctoring for personal or privacy reasons and took the exam in a campus computer room having sufficient invigilators (1 invigilator per maximum 5 students). A large majority of students who took an exam with online remote proctoring was very positive about the flexibility this had offered them. Several students who took the assessment at home explained why by statements such as: “I prefer to take the exam at home because at home I can concentrate better”. Several students who took their exam on campus answered the question “Why did you take the exam on campus and not at home?” by “I prefer to take the exam on campus because there is less disturbance than at home and I can concentrate better.”

In practice, no summative assessment can be 100% secure from fraud or cheating. We used online remote proctoring only in case a university digital examination room was not a feasible option for a specific course or group of students. In table 1 (see appendix), a tool is presented to estimate benefit/costs ratio that could help instructors or programmes to make a choice for proctoring online in relation to the assessment setting and the flexibility that can be gained with it.

4. Conclusions

Remote online proctoring makes it possible that summative e-assessments are more flexible in place and time and that open resources can be used. This way a summative assessment can be better constructively aligned with the learning process of the course. Instead of a list of allowed aids for the exam, students are asked to handle according to the academic aptitude. Thus online remote proctoring is not just a different way to control students and to prevent cheating, but it makes it possible to improve the quality of the learning process as a whole in which summative assessment is an integral part.

5. Acknowledgements

We acknowledge the following education innovation projects:
Online Switch (University of Amsterdam, 2013-2014,
http://starfish.innovatievooronderwijs.nl/project/21/),

Surveillance on distance, (SURF, Innovatieregeling Digitaal toetsen voor Onderwijs op maat, 2015-2016, https://www.surf.nl/innovatieprojecten/onderwijs-op-maat/digitaal-toetsen/innovatieregeling-digitaal-toetsen-voor-onderwijs-op-maat-2015-2016/index.html),
Digital Assessment Matched (Faculty of Science, University of Amsterdam 2015-2017, http://starfish.innovatievooronderwijs.nl/project/634/).

6. References

Apampa, K.M., Wills, G. and Argles, D., 2009. Towards security goals in summative e-assessment security. *Internet Technology and Secured Transactions*, *ICITST 2009. International Conference for*, *London*, pp.1-5., doi: 10.1109/ICITST.2009.5402505. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5402505&isnumber=5402499 [Accessed 5 October 2016].

Biggs, J., 1996. Enhancing teaching through constructive alignment*, Higher Education,* 32, pp.1-18.

Conole, G. and Warburton, B., 2005. A review of computer-assisted assessment, *Alt-J*, 13(1) pp.17-31. doi: 10.1080/0968776042000339772.

Cornock, C., 2016. Teaching mathematics to Business and Enterprise students in a module based around Excel. *MSOR Connections*, 15(1), pp.22-27. Available at: https://journals.gre.ac.uk/index.php/msor/article/view/315/pdf [Accessed 13 October 2016].

Fask, A., Englander, F. and Wang, Z., 2014. Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment. *Journal of Academic Ethics*, 12(2), pp.101-112. doi: 10.1007/s10805-014-9207-1.

Heck, A. and Brouwer, N., 2015. Digital assessment-driven examples-based mathematics for computer science students. In N. Amado and S. Carreira (Eds.), *Proceedings of the 12th International Conference on Technology in Mathematics Teaching*, pp.403-411. Faro, Portugal: Universidade do Algarve. Available at: http://hdl.handle.net/10400.1/6081 [Accessed 13 October 2016].

Rowe, N.C., 2004. Cheating in online student assessment: Beyond plagiarism. *On-Line Journal of Distance Learning Administration, Summer 2004*. Available at: http://hdl.handle.net/10945/36015 [Accessed 5 October 2016].

7. Appendix

Table 1. Institution benefit/cost ratio

Benefit/cost ratio: high ***, medium **, low *

Number of students	A resit with a small group of students	An exam with access to the internet	An exam on BYOD	Students are abroad / unable to come to the university location
1 or 2 students	Exam location outside the university (e.g. at home) *ProctorExam Pro* ***	Exam location outside the university (e.g. at home) *ProctorExam Pro* ***	Exam location outside the university (e.g. at home) *ProctorExam Pro* ***	Exam location outside the university (e.g. at student's home) *ProctorExam Pro* ***
	No location costs. No local computer costs.	No location costs. No local computer costs.	No location costs. No local computers costs	No location costs. No local computer costs.
< 50 students	In a lecture room *ProctorExam Light* **	In a university lecture room *ProctorExam Light* **	In a university lecture room *ProctorExam Light* **	Exam location outside the university (e.g. at student's home) *ProctorExam Pro* ***
	Low location costs room invigilators computers in the room	Low location costs, room invigilators computers in the room	Low location costs room invigilators extra computers (safety / privacy)	No location costs. No local computer costs.
50-100 students	n.a.	In a university lecture room *ProctorExam Light* **	In a university lecture room *ProctorExam Light* **	Exam location outside the university (e.g. at student's home) *ProctorExam Pro* **
		Location costs room invigilators computers in the room	Location costs room invigilators extra computers (safety / privacy)	No location costs. No local computer costs. Higer costs per student when group is larger.
> 100 students	n.a.	In a university lecture room *ProctorExam Light* *	In a university lecture room *ProctorExam Light* *	Exam location outside the university (e.g. at student's home) *ProctorExam Pro* **
		Location costs room invigilators big computer rooms or multiple computer rooms	Location costs room invigilators extra computers (safety / privacy)	No location costs. No local computer costs. Higer costs per student when group is larger.

CASE STUDY

Using e-assessment to support flipped-style teaching

Karen Henderson, Department of Engineering Design and Mathematics, University of the West of England, Bristol, U.K. Email: karen.henderson@uwe.ac.uk

Abstract

We show how weekly formative e-assessments are used to support flipped-style teaching of a module delivered to all first year Mathematics students at the University of the West of England, Bristol (UWE). The flip lecture approach places students at the centre of the learning process. For the module described here, a highly scaffolded approach was employed. A workbook containing gapped lecture notes was created as well as a handbook containing exercise sheets and extra reading material. Each week students were expected to independently: watch screencasts and fill in the relevant gaps in their workbooks; take a formative e-assessment; try some basic questions from the exercise sheet and optionally do some extra reading and/or work through a Maple file. During the following two hour class, TurningPoint questions and group activities were used to encourage active learning. Student feedback of this new teaching approach has been very positive.

Keywords: flipped-style teaching, e-assessment, gap notes, active learning.

1. Introduction

Flipped-style teaching or the flipped classroom has seen a surge in interest recently (Brame, 2013; Maciejewski, 2016). This style of teaching, pioneered by Mazur (1997), is a change to the traditional lecture model used in universities for hundreds of years. In the traditional model, the lecturer is in charge of the class and largely dictates the material and pace at which this is delivered. Typically students are then required to work through more challenging material on their own before attending tutorials/problem classes for support. The idea behind the flipped classroom is that students' initial exposure to material takes place in their own time, so students work through material independently at their own pace before the formal class. Class time may then be used for active learning, where students are able to deepen their understanding of the material, through for example problem-solving, peer instruction and discussion. In this case study we describe the process and results of flipping part of a level 1 (first year undergraduate) calculus module, focussing on how we have used e-assessment to support this.

2. Background

The module considered in this case study is called Calculus and Numerical Methods (CNM). This is a level 1 module delivered to all first year students on the BSc (Hons) Mathematics and BSc (Hons) Mathematics & Statistics courses at the University of the West of England, Bristol (UWE). CNM is taught over the whole academic year; the first semester is delivered in a traditional way whereas the second semester has been delivered using a flipped approach since January 2015. The material covered in the second semester includes the topics of differential equations, numerical methods, series and report writing.

The module is assessed through a written examination (worth 75% of the module mark) and coursework which comprises four e-assessments (7.5%) and a group case study (17.5%). E-Assessment has been used on this module for many years and our assessment strategy has evolved from being merely summative to also include formative assessments that give high quality feedback from which students actively learn (Gill and Greenhow, 2008). Using online tests to support learning has become standard practice in many institutions (Sangwin, 2013). We use Dewis (2012) to deliver the e-assessments on this module.

Dewis is a fully algorithmic open-source web-based e-assessment system which was designed and developed at UWE. It was primarily designed for the assessment of mathematics and statistics and supports a range of inputs, such as numeric entry, algebraic entry, matrix entry, multiple choice and drop-down selection. An example of several e-assessment questions used for CNM is illustrated in figure 1 together with the full feedback received for one of the questions. Using an algorithmic approach enables the separate solution, marking and feedback algorithms to respond dynamically to a student's input and as such can perform intelligent marking. In addition, the Dewis system is data-lossless; that is, all data relating to every assessment attempt is recorded on the server. This enables the academic to efficiently track how a student or cohort of students has performed on a particular e-assessment (Walker, Gwynllyw and Henderson, 2015). Recent developments include using embedded R code to facilitate the assessment of students' ability to perform in-depth statistical analyses (Gwynllyw, Weir & Henderson, 2016) and using Dewis to automatically mark computer code (Gwynllyw, 2016). Implemented for the first time in 2007 the system is now well-established and in 2015/16 within UWE and partner institutions, Dewis was used for formative and summative tests to support over 3,500 students involving more than 50,000 assessment attempts.

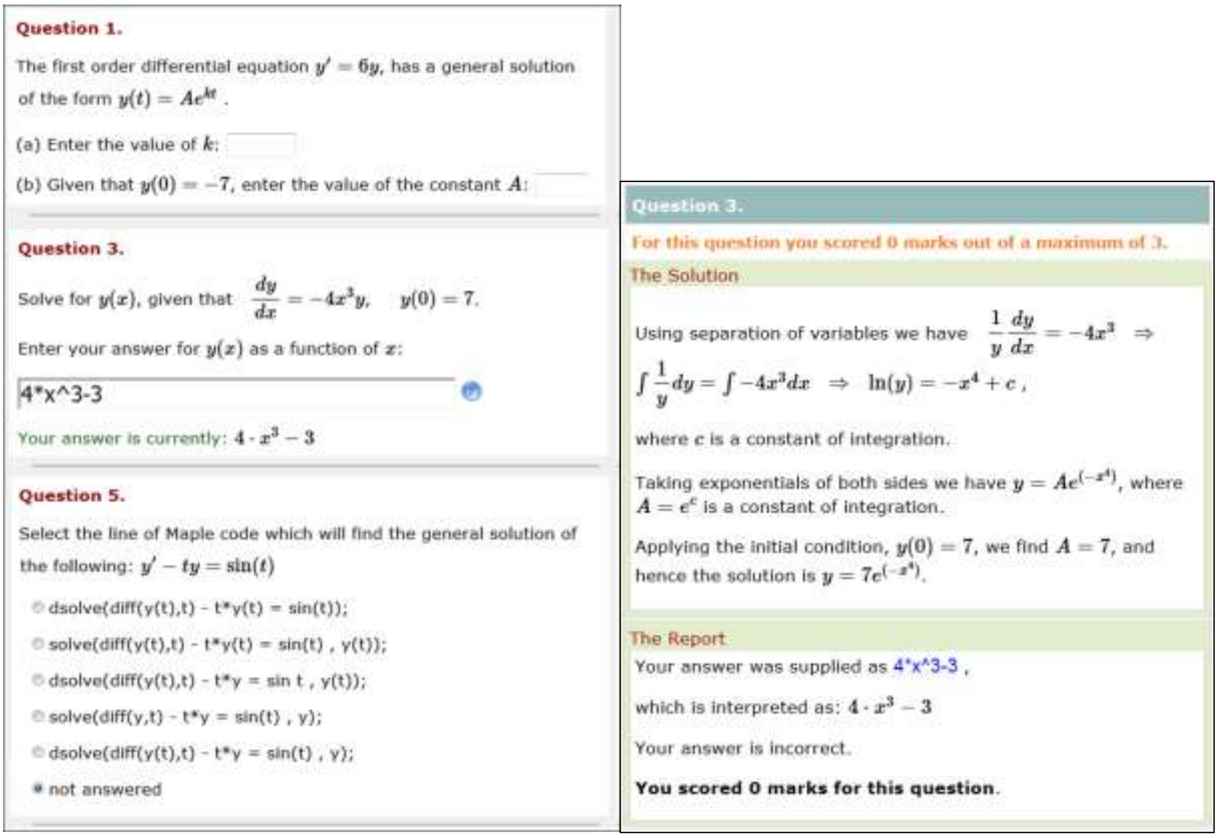


Figure 1. Example Dewis questions, together with feedback and marking bespoke to the random parameters used for one of these questions.

3. Methodology

3.1. Motivation

There were several motivations for deciding to adopt the flipped approach on this module. Despite always having good student feedback and results I was concerned as to how much my students were learning and in particular whether they could still remember techniques and methods when they came to their final year. I also wanted students to have a deeper learning experience and to take control of their learning. In order to investigate how the flip lecture could best be incorporated into my teaching, I attended an HEA STEM workshop on `Lectures without lecturing' in February 2013 and participated in classes run by fellow practitioners. Mathematics has been slower than other subjects to embrace this approach and I felt for it to work effectively, especially at level 1, students needed scaffolding. Building on my experience of using technology (Hooper, Henderson and Gwynllyw, 2014) I created materials to use pre-class and in-class and these are described in the following two sections.

3.2. Pre-class material

Prior to the start of the second semester, all students were issued with a workbook of gapped notes. This contained background material, key mathematical theorems and examples, all of which contained gaps in selected places. Each week, prior to the scheduled class, students were expected to:

- watch a series of screencasts and fill in the relevant gaps in their workbooks;
- take a formative Dewis e-assessment test;
- try some basic questions from the exercise sheet;
- optionally do some extra reading and/or work through a Maple file.

Typically there were four screencasts to watch each week lasting on average 10 minutes each. A total of 35 screencasts were produced with a tablet PC using Camtasia Studio software. These were made available through SCORM packages on the University's VLE (Blackboard) and students were deemed to have completed that task if they watched 95% of the screencasts, with their progress monitored through the VLE. Additionally, at the end of each screencast students were asked to provide feedback as to whether they "thought that the video was (a) Good (b) OK (c) Poor: Please re-record for next year". This gave me timely feedback on the quality and relevance of the screencasts. There was no limit on the number of attempts allowed at each weekly formative Dewis e-assessment which typically contained five questions. Only three attempts were allowed for each summative e-assessment, which comprised a selection of questions already seen in these weekly tests, so students had an additional incentive to attempt the practice tests prior to the summative e-assessments becoming live.

3.3. In-class activity

Each week students were timetabled for a two hour class in a flat teaching room, with a one hour optional support session available every fortnight. The two hour class started with listing the learning outcomes for the week followed by a suite of TurningPoint (TP) questions, which typically took an hour to complete (Hooper, Henderson and Gwynllyw, 2014). An example of a typical TP question is shown in figure 2.

1. Consider the differential equation

$$y'' - 9y' + 18y = -12e^{6x}$$

Determine how many of the following statements are true:

- The auxiliary equation is $m^2 - 9m + 18 = 0$.
- The complementary function is $y_{CF} = Axe^{6x} + Be^{3x}$
- There exists a constant p such that $y = pe^{6x}$ satisfies (1).
- The particular integral is $y = 4xe^{6x}$
- The general solution is $y = (A - 4x)e^{6x} + Be^{3x}$

A 1
B 2
C 3
D 4
E 5
F I don't know

Figure 2. An example of a typical TurningPoint question used in the class (left), together with the voting options for this question (right).

Depending on the success of the cohort with a particular question I would re-poll (if the vote was significantly split), go through the answer in detail (if several did not know the answer – this was always one of the response options, or answer correctly) or move onto the next question (if everyone answered correctly). Question sheets were handed out at the start of each class, so that students didn't have to waste time writing out the question and could move onto the next question if they finished early. Students who attended the class but who had not done the pre-class work were able to view a PDF of the completed workbook from the University's VLE. This became available at the start of the class.

4. Results

4.1. Monitoring engagement

Three measures were used to monitor engagement with the module. These were whether the student had that week: attended the class; watched the screencasts; attempted the practice e-assessment. The semester 2 attendance for both years that the flipped teaching has been employed stayed fairly constant throughout the first ten weeks and was at a similar level to that experienced in semester 1 (approximately 75%) in which a more traditional lecture/tutorial delivery was employed. The remaining two weeks of semester 2 was used for revision purposes. Table 1 gives an indication of how many students did the pre-class work in 2015-16. The second column gives the number of students who watched at least three-quarters of the screencasts for that particular week at some stage prior to the final exam. In the next two columns we show details of the number of students and total number of attempts at each weekly practice e-assessment prior to the class. These numbers are disappointingly low, particularly for the practice tests from week four onwards. We note that practice tests 1-3 contributed to summative e-assessment 3 and practice tests 4-7 contributed to summative e-assessment 4, whilst the questions in practice tests 9 and 10 did not contribute to a summative e-assessment. Columns five and six show the number of students and total number of attempts made at the weekly practice tests prior to the written exam in May. We can see that, particularly for weeks 1-7, although not all students tried the tests, those that did made multiple attempts at each test.

Table 1. Details of the number of students (out of a total of 36) who did the pre-class work in 2015-16.

Week no.	No. students who watched 75% of weekly screen-casts	No. of students attempting test before class	No. of attempts at test before class	Total no. of students attempting test	Total no. of attempts at test	Comments
1	29	14	19	25	78	Practice tests 1-3 contributed to assessment test 3
2	31	11	19	22	90	
3	27	15	23	26	69	
4	24	1	1	15	45	Practice tests 4-7 contributed to assessment test 4
5	25	0	0	17	47	
6	21	8	10	17	47	
7	19	6	8	17	43	
8	19	Report writing and group work - no Dewis test				
9	23	0	0	4	8	Practice tests 9-10 not used in a summative assessment
10	17	4	5	7	9	

4.2. Student performance

In table 2 we display the exam marks and exam pass rates for the last four years. We can see that there has been a marked increase in performance since we started flipping in 2014-15. However, this was not the only change brought in at that time. We also increased the e-assessment coverage across the whole syllabus (semester 1 and semester 2), introduced fortnightly Dewis practice tests in semester 1 and actively monitored engagement with all practice tests.

Table 2. Comparison of exam marks and pass rates over the last four academic years.

Exam and year	Number of attempts	Pass rate	Average Mark
2012-13	37	75.6%	55.5
2013-14	65	66.1%	45.8
2014-15	64	92.0%	72.6
2015-16	36	88.9%	71.0

We found a strong correlation between engagement and exam performance. Further, the three students who failed the exam in 2015/16 did not attempt any of the practice e-assessments, engaging through watching the screencasts and attending only.

4.3. Student feedback

Student feedback has been overwhelmingly positive to the flipped-style approach. At the end of the first year (April 2015) I ran an in-class questionnaire to gauge student feedback. Answers to selected questions have been collated in table 3.

Table 3. Outcomes of some of the questions used in an in-class questionnaire (April 2015).

Qn 1: *On average I spent the following amount of time on the pre-class work*				
30 mins	1 hour	1.5 hours	≥ 2 hours	Did not do
5.5	12.5	14	2	1
Qn 2: *Doing the weekly Dewis tests was helpful to my learning*				
Strongly agree	Agree	Neutral	Disagree	Strongly disagree
14	15	6	0	0
Qn 3: *I liked the new style of teaching*				
Strongly agree	Agree	Neutral	Disagree	Strongly disagree
20	10	4	1	0

It was encouraging to see that the majority found the weekly e-assessments helpful to their learning. A similar response to Qn 3 in table 3 was received from the second group of students via the university formal online module evaluation in May 2016. Specific comments from students were collected via module evaluations and some are shown below:

"I thought that the Flipped Learning technique … was very effective, it meant that we could get the majority of the knowledge in our own time and at our own speed (we could pause the videos if we wanted), then in the lecture we could go over any problems that we may have had with harder questions."

"… the flipped class teaching worked very well, as often the easy, basic stuff is taught in lectures and then the harder stuff is left to questions out of class. Whereas instead the videos online before the lecture were all encompassing of the information we needed and then in lectures everything was made clearer by group questions and further demonstrations."

"The flipped learning approach helped me to learn the semester two work well. The weekly tests were good as they helped with learning content and the online coursework tests."

"the pre work was amazing, I got so much more out of the lectures because of it."

"The flipped classroom approach was something I wasn't expecting to get on with, but instead was a far better style of teaching than I had anticipated and would like this to be continued next year."

"The flipped approach worked really well for me as it allowed me to get a basic understanding of the work first for myself, and it actually made me do the work before the lecture so that I knew what was going on."

All comments received were positive of the flipped-style teaching.

5. Discussion

The flipped-style teaching worked much better than I had hoped and students responded very positively to it. Having a highly-scaffolded approach worked well, in that students were very clear what they needed to do each week and had a range of different learning activities to work from. It also meant that students who were not able to attend the class, for whatever reason were still able to keep up with the work to a certain extent. However the drawback was that it was quite time-consuming to set up which could potentially be a barrier to adoption on other modules.

During the class itself, not everyone in the room voted using the TP clickers. This did not appear to be due to technological problems; instead it seemed that some students preferred not to partake, despite responses being anonymous. The drawback to this was that I may not always have had an accurate picture of the class' understanding of different topics. It was slightly disappointing that more students didn't attempt the Dewis practice tests prior to class. Anecdotally students said to me that they didn't want me to see if they achieved a low mark, so preferred to try them after the class when they were sure of the material. To address this for future years I may amend the wording of these practice tests from "engagement will be monitored" to "non-engagement will be monitored".

From January 2016 the class will be delivered in a Technology-Enhanced Active Learning (TEAL) space (MIT iCampus, 2016). This space contains collaborative working pods which each seat up to six students and include a PC. Students within each pod can work independently on their PC and the lecturer can choose to project the pod's or the podium's screen to the whole class if desired. Using this TEAL space may encourage better small group discussion and peer instruction in class. In addition, this different learning environment will enable students to use relevant software, e.g. Maple during class.

6. References

Brame, C., 2013. Flipping the classroom. Vanderbilt University Center for Teaching. Available at: http://cft.vanderbilt.edu/guides-sub-pages/flipping-the-classroom/ [Accessed 23 December 2016].

Dewis Development Team, 2012. Dewis welcome page. Available at: http://dewis.uwe.ac.uk [Accessed 14 October 2016].

Gill, M. and Greenhow, M., 2008. How effective is feedback in computer-aided assessments? *Learning, Media and Technology*, 33, pp.207-220. Available at: http://dx.doi.org/10.1080/17439880802324145 [Accessed 23 December 2016].

Gwynllyw, R., Weir, I. and Henderson, K., 2016. Using Dewis and R for multi-staged statistics e-assessments. *Teaching Mathematics and Its Applications*, 35(1). pp.14-26. Available at: http://teamat.oxfordjournals.org/content/35/1/14 [Accessed 23 December 2016].

Gwynllyw, R., 2016. Automatic marking of C code applied to mathematics problems. In preparation.

Hooper, A., Henderson, K. and Gwynllyw, R., 2014. Using technology in the module delivery of mathematics. Presented at the CETL-MSOR Conference, Cardiff. Available via http://www.sigma-network.ac.uk/wp-content/uploads/2014/09/K_Henderson_A_Hooper_R_Gwynllyw_-session5.pdf

Maciejewski, W., 2016. Flipping the Calculus Classroom: an Evaluative Study. *Teaching Mathematics and its Applications*. Available at: http://teamat.oxfordjournals.org/content/35/4/187 [Accessed 23 December 2016].

Mazur. E., 1997. Peer Instruction: A user's manual. *Series in Educational Innovation*. Prentice Hall.

MIT iCampus, 2016. TEAL – Technology Enabled Active Learning. Available at: https://icampus.mit.edu/projects/teal/ [Accessed 23 December 2016].

Sangwin, C., 2013. *Computer Aided Assessment of Mathematics*. Oxford: Oxford University Press.

Walker, P., Gwynllyw, R. and Henderson, K., 2015. Diagnosing student errors in e-assessment questions. *Teaching Mathematics and its Applications*, 34(3). pp.160-170. Available at: http://teamat.oxfordjournals.org/content/34/3/160 [Accessed 23 December 2016].

CASE STUDY

Numbas as an engagement tool for first-year Business Studies students

Tom Carroll, Department of Mathematics, University College Cork, Cork, Ireland.
Email: t.carroll@ucc.ie
Deirdre Casey, Department of Mathematics, Cork Institute of Technology, Cork, Ireland.
Email: deirdre.casey.cit@gmail.com
Julie Crowley, Department of Mathematics, Cork Institute of Technology, Cork, Ireland.
Email: julie.crowley@cit.ie
Kieran F. Mulchrone, Department of Applied Mathematics, University College Cork, Cork, Ireland.
Email: k.mulchrone@ucc.ie
Áine Ní Shé, Department of Mathematics, Cork Institute of Technology, Cork, Ireland.
Email: aine.nishe@cit.ie

Abstract

In this paper we report on the implementation of e-assessment in mathematics with a large cohort of Business Studies students in their first year at Cork Institute of Technology, Ireland. The assessment tool used was Numbas, a freely available e-assessment tool for mathematics developed at Newcastle University. The main motivation in introducing Numbas for this course was to increase attendance and engagement at tutorials but also to make regular assessments with feedback a practical possibility for large groups. In this paper we discuss the effect the introduction of Numbas had on student engagement, in particular on student participation, attendance, and on the student experience.

Keywords: Numbas, student engagement, online assessment, computer aided assessment, formative assessment.

1. Introduction

Historically, attendance and engagement in mathematics modules were reported as low among first year Business Studies students at Cork Institute of Technology (CIT). Moreover, students' feelings about mathematics were often negative. Attendance at tutorials in particular was poor, especially among those students most likely to need the help offered by tutorials. In CIT, the average teaching load for a lecturer is 18 to 20 hours per week with lecturers often teaching five or more modules concurrently. With this heavy workload it is difficult to have regular assessment with timely and constructive feedback. As a result of these time and workload restrictions, this cohort of students traditionally had only one exam during the semester together with the final exam. Other written homework sheets were assigned for the students to work through for tutorials but these were not incentivised with marks so many students did not always attempt them. Hence, students lacked regular formative feedback on how they were doing and what they needed to improve on.

We wished to investigate if using Computer Aided Assessment (CAA) would help alleviate some of these problems. In particular, we focused on increasing attendance and engagement at tutorials. We also wanted to understand if using CAA would improve the students' experience of mathematics. Here we report on the use of Numbas as a tutorial and assessment tool for the first year Business Studies students over two semesters. We settled on the use of Numbas because it has a strong reputation, is user friendly and is compatible with the Virtual Learning Environment used at the CIT campus.

The remainder of this paper is organised as follows. First we discuss the background underpinning our research, then we go on to describe the methodology we used to implement e-assessment and collect data, then we describe the results of our research and finally we draw our conclusions.

2. Background

2.1. Engagement

While student engagement is central to Higher Education there is not, however, general consensus in the literature on a definition of the term. For example, Trower and Trowler (2011) define student engagement as follows:

"Student engagement is concerned with the interaction between the time, effort and other relevant resources invested by both students and their institutions intended to optimise the student experience and enhance the learning outcomes and development of students and the performance and reputation of the institution"

In this paper, we follow Warwick (2008) and Linnenbrink and Pintrich (2003) where student engagement is divided into three components: behavioural engagement, cognitive engagement and motivational engagement. They describe behavioural engagement as

"the observable behaviour we see as teachers in the classroom. This relates to the efforts students are putting into mathematical tasks and how students relate to each other and to the teacher in terms of their willingness to seek help, attendance at the classes etc.".

We focus on this definition and use attendance, participation and student enjoyment as indicators of engagement with the modules.

2.2. Assessment and Feedback

Regular assessment and quick feedback improves learning (Black & Wiliam, 1998). Students tend to be very judicious in where they focus their efforts and can be 'selectively negligent' when there is no assessment associated with a topic (Gibbs & Simpson, 2004-5). Assessment and feedback, although widely accepted as increasing engagement, are difficult and time consuming in practice. The idea of automating (or partially automating) these processes seems attractive as students derive the benefits without the substantial increase in workload involved. It would seem that e-assessment makes it possible for the practitioner to synthesise best practice in encouraging engagement. There is a long history of e-assessment in mathematics back to the 1980s when WeBWorK, the online homework system, was developed by Michael Gage and Arnold Pizer at the University of Rochester. Feedback is also well documented as increasing engagement (Cairini, et al., 2006). Bearing this in mind and in an effort to increase engagement for students of first year Business studies modules we introduced the mathematics e-assessment tool Numbas.

2.3. Numbas

Numbas is a freely available e-assessment tool for mathematics developed at Newcastle University. It generates random variations of uploaded questions and interacts with Learning Management Systems such as Blackboard and Moodle. It allows students to input mathematical formulae easily and creates a similar but different question for each student. It is an excellent formative assessment tool giving students instant feedback with features such as 'try another question like this one', 'show steps' and an 'advice' section which gives the solution to the question, as shown in figure 1. One can also add images, video, graphs and embed GeoGebra in a question. Since Numbas is an open source tool with a global community of users, expertise and experience can be shared nationally and internationally. It is actively maintained at Newcastle and is easy for lecturers and students to work with. The Numbas system has a proven record of accomplishment and a strong reputation (Foster, et al., 2012 and Perfect, 2015). It is currently being used in Cork Institute of Technology and University College Cork as well as in Newcastle University, University of Leicester, Kingston University, London (Denholm-Price & Soan, 2014) as well as at universities in Norway and South Africa.

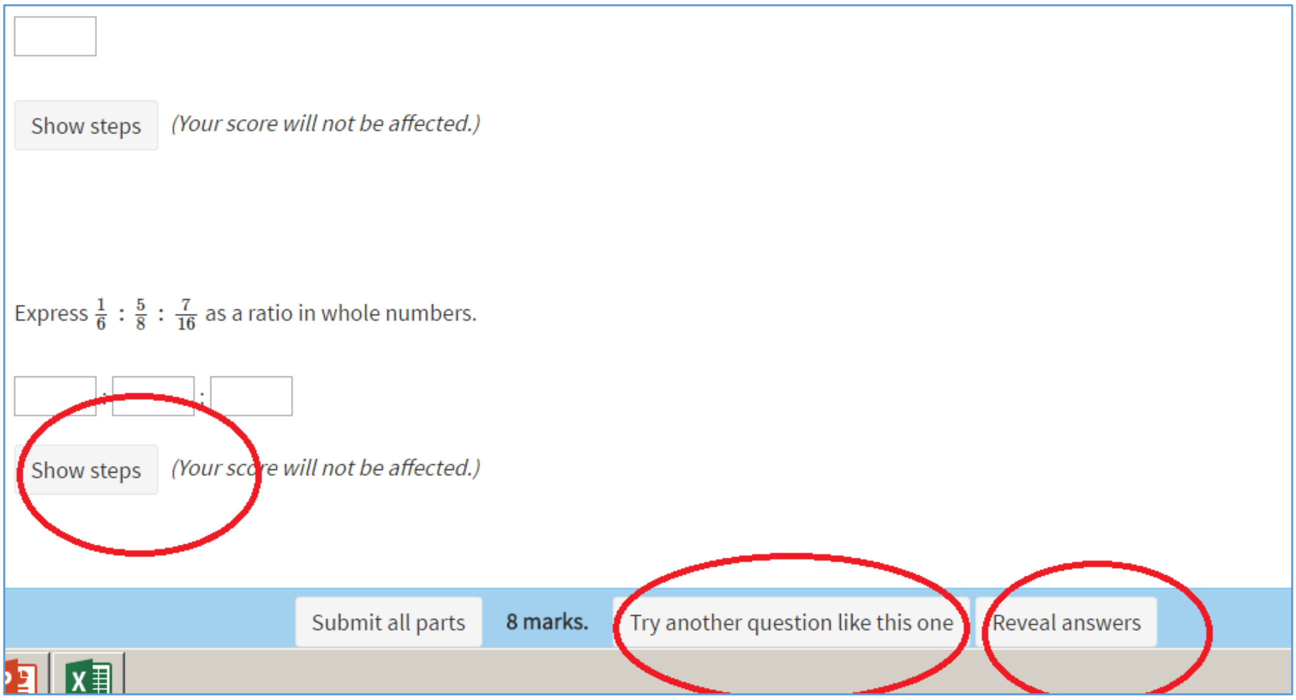


Figure 1. Numbas as a formative assessment tool

3. Methodology

3.1. Implementation

Prior to the introduction of Numbas, students had three hours of lectures each week and a one-hour pen and paper tutorial every second week. Assessment consisted of a mid-semester exam (20%) and an end of semester summative exam (80%). Following the introduction of Numbas, the schedule of lectures and tutorials stayed almost the same but the format of the tutorials changed. Under the new system students still had 3 lecture hours per week and they now had a tutorial every week, but the format of the tutorial varied depending on the week (see table 1).

The final mark for the module was now made up of three Numbas assessments (20%), a statistical software package exam (20%) and an end of semester summative exam (60%).

The Numbas tutorials were very structured. The relevant questions for a given tutorial were available for students to practise for at least one week before the tutorial. The students were given an opportunity to work through the questions in the tutorial for 40 minutes with the tutor available to give help and guidance. The practice ‘tests’ were designed to make the most of the medium of formative assessment. Students got instant feedback on whether their answer was correct or not, had unlimited attempts and had access to hints, answers and fully worked solutions as well as tutor support. Each student worked at their own pace but students could help each other with the method of how to do a question. The last 15 minutes of the tutorial were devoted to the Numbas assessment. The assessment questions were a subset of the questions that the students had been working on for the previous week and in the tutorial. Assessments were set up so that students could only submit one attempt at an answer and there was a fixed time limit.

Table 1. Tutorial schedule

Week	Type of Tutorial	Description
1	Numbas	Getting started with Numbas and a practice Numbas test
2	Statistical software	Computer Lab
3	Pen and Paper	Traditional pen and paper tutorial, working through students' questions on assigned homework
4	Numbas	Tutorial and a 15 minute assessment
5	Statistical software	Computer Lab
6	Pen and Paper	Traditional pen and paper tutorial, working through student's questions on assigned homework
7	Pen and Paper	Traditional pen and paper tutorial, working through students' questions on assigned homework
8	Numbas	Tutorial and a 15 minute assessment
9	Statistical software	Computer Lab
10	Numbas	Tutorial and a 15 minute assessment
11	Statistical software	Computer Lab
12	Statistical software	Statistical Software exam

In the academic year 2015/16, the modules in question had 459 registered students and were delivered by a teaching team of ten lecturers/tutors. As CIT does not currently have a large electronic exam hall and occupancy of each computer lab is limited to 22 students, the group needed to be split into 25 different tutorial groups, which occurred at 21 different times. This complexity posed some logistical challenges and required careful preparation and planning. A number of copies of printable versions of the assessment were also available as a back up.

Once created, the Numbas tests were uploaded to the virtual learning environment Blackboard. Students in CIT are familiar with using Blackboard for other learning and assessment purposes and so adapt very easily to doing their mathematics assessments through this system. This integration with Blackboard allows their results to be automatically tracked and recorded. The adaptive release feature available on Blackboard was used to control the times that students could access the Numbas assessment.

3.2. Data Collection

An online survey was emailed to all 459 first year Business students at the end of Semester 1 of the 2015/2016 academic year. In total, 83 responses were received. We asked students to respond to statements on a 5-point Likert scale with the five options: strongly disagree, disagree, neutral, agree and strongly agree. The survey also included some open-ended questions. The most apparent limitation to this survey study is the response rate of 18%. However, despite this low response rate, valuable insight was gleaned in that the students that did respond gave a full and detailed response. In addition to the student survey, 10 lecturers/tutors involved in the modules were asked to complete a survey anonymously. Seven responses to the lecturer survey were received. The authors applied both qualitative and quantitative methods to the data collected. Attendance records were available for six tutorial groups in Semester 1 and four tutorial groups in Semester 2.

4. Analysis and Results

In relation to engagement, the themes emerging from our research are increased student participation, increased attendance and improved student experience in terms of enjoyment. As expected, instant feedback was a popular feature for students. The main barrier we expected was usability but this was not reported as an issue.

4.1. Student Participation

We found that students like to the get the 'green tick' to say that they have answered a question correctly. This seemed to be a key motivation and students were *"more likely to try again than in traditional pen and paper tutorials",* according to one of the lecturers. *S*tudents interacted more with the material being taught and lecturers/tutors felt that the students were more engaged in tutorials then they had been in previous years. The students asked more questions and took more control of their own learning. In the lecturer/tutor survey, 6 out of 7 said they agreed or strongly agreed with the statement *"Numbas has changed the manner in which students engage with Mathematics in college."* One lecturer commented that students "*do more revision*" since the introduction of the Numbas system. Other comments by lecturers included that Numbas *"gets the students working on material"* and *"encourages student participation."*

4.2. Attendance

We found that attendance at the Numbas tutorials was higher than at traditional tutorials. Figure 2 shows the average attendance at tutorials (both Numbas and non-Numbas tutorials) in Semesters 1 and 2 of academic year 2015/16. In both semesters attendance at Numbas tutorials was about 20% higher than at the tutorials where Numbas was not used.

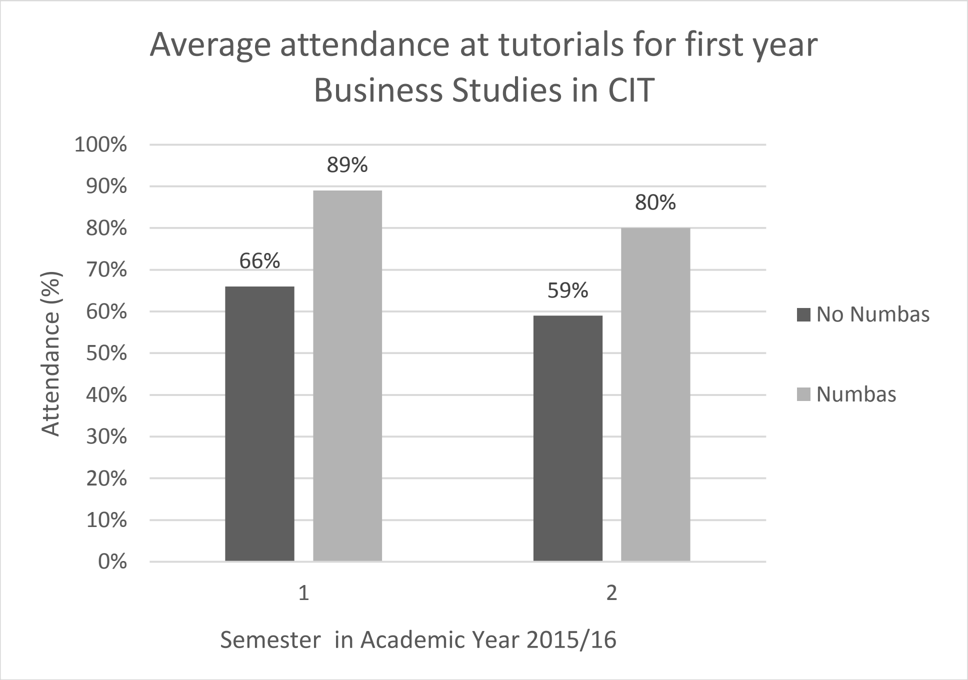


Figure 2. Average attendance at mathematics tutorials for students of first year Business Studies.

Of those that replied to the lecturer/tutor survey, 6 out of 7 said that they agreed or strongly agreed with the statement "The use of Numbas has increased attendance at tutorials." One lecturer commented, *"When (pen and paper) tutorials are the only thing scheduled, less students turn up and they are less inclined to work even if they do turn up".*

4.3. Student Enjoyment

Numbas tests gave rise to a more positive feeling about mathematics and the students found it enjoyable to use. In the survey, students were asked, *"Do you feel that Numbas Assessments have allowed you to enjoy maths more in college?"* When the answers were coded, the result was that 64% answered positively, 31% answered negatively with 5% giving a neutral answer. Some negative comments included:

"No, I hate maths full stop."

"No. under too much time pressure when completing assessments."

"Not at all. I found it quite draining"

The majority of students, however had positive responses:

"It is definitely more enjoyable than normal maths."

"A bit yes I'll never enjoy maths but Numbas really helps"

"Yes it's a change to listening to a lecturer all day and gives you the opportunity to work on maths."

"Yes, definitely. It is something I really didn't mind practicing at home in my own time."

"Yes, it was a new way of learning maths then before and it's much easier."

"Yes, you are more engaged with assessments than versus a class and it is more enjoyable."

"Yes. The interactive section of Numbas helped me to enjoy maths more in college. I looked forward to practicing my Numbas at home in preparation for Numbas assessments."

4.4. Feedback

Figure 3 shows student responses to the statement *"Feedback given by the Numbas program is useful to me"*. One student commented, *"You get constructive feedback on your work"* while another said *"I found it the more interesting and helpful way to study certain topics and it's a good way to challenge yourself in your own time"*. Backing this up, when lecturers were asked to rate the statement *"Feedback given by Numbas is useful for the students"*, they unanimously agreed with the statement.

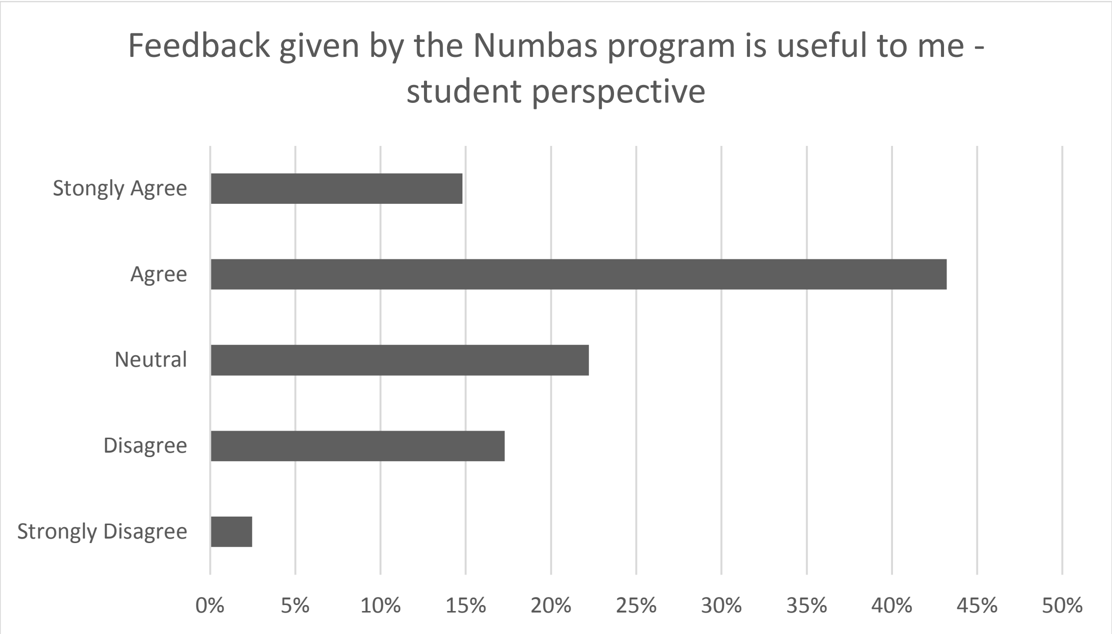


Figure 3. "Feedback given by the Numbas program is useful to me" - student perspective.

4.5. Usability

User experience (UX) refers to the quality of the user's interaction with and perceptions of a system. We expected the inputting of mathematical expressions to be a key issue particularly with a cohort of students not familiar with inputting mathematics into a computer. Sangwin (2013) refers to notation and syntax as "*the most significant barrier to CAA use*". However, inputting mathematics did not seem to be a problem for students when they were using the Numbas system. Students were asked to rate their agreement or otherwise with the statement "*The Numbas system is straightforward for me, as a student, to use"*. As shown in figure 4, the majority of students found Numbas straightforward to use.

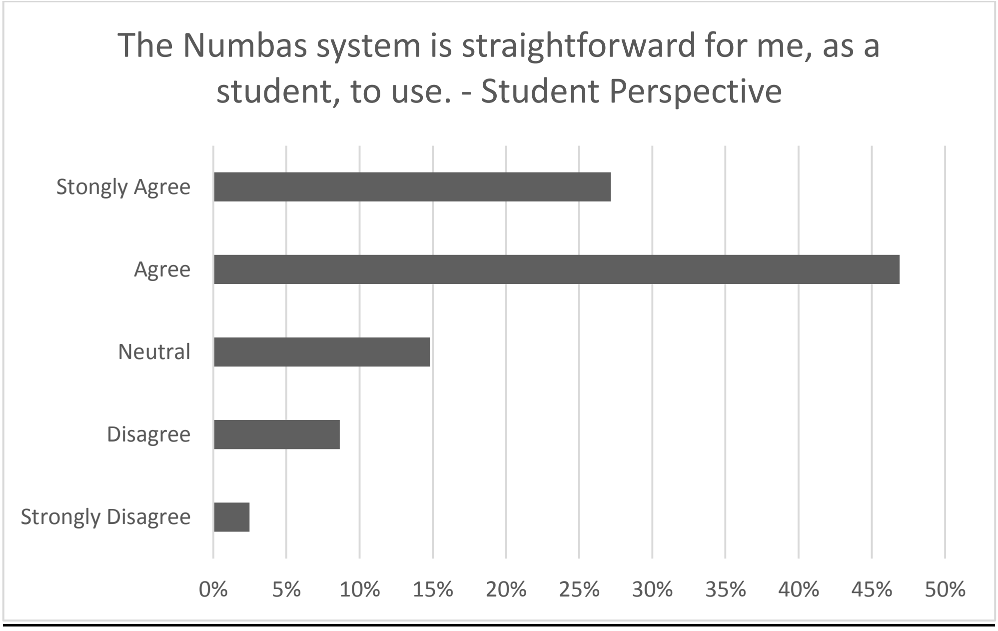


Figure 4. "The Numbas system is straightforward for me, as a student, to use." - student perspective.

Even though initially there was some apprehension among the teaching team to the introduction of a CAA to the modules, after using Numbas in their classes the lecturers/tutors all either agreed or strongly agreed with the statement "*The Numbas system is straightforward for me, as a lecturer or tutor, to use."* One lecturer commented that *"it is a very good system"* and other said *"There were minor details that I needed to work out myself but overall very user friendly."*

5. Conclusion

In order to address low attendance at tutorials and low engagement at all classes we introduced the online mathematics e-assessment system Numbas as a tutorial and assessment tool. This improved the student experience and increased engagement by increasing attendance, student participation and student enjoyment. It made regular assessment with timely feedback a practical possibility for a large group of students. Usability did not prove to be the barrier that we expected it to be. Through the feedback it gave to students, Numbas proved to be a very successful formative assessment tool. We conclude with a comment from one of our teaching team, which summarizes much of this paper

"I think Numbas is a good thing, useful for getting students engaged."

Acknowledgements

The authors would like to acknowledge the financial contribution from the National Forum for the Enhancement of Teaching and Learning in Higher Education through the Teaching and Learning Enhancement Fund 2014. The authors would also like to acknowledge the helpful comments of the reviewer.

References

Black, P. & Wiliam, D., 1998. Assessment and Classroom Learning. Assessment In Education: Principles, Policy & Practice, 5(1).

Cairini, R., Kuh, G. & Klein, S., 2006. Student engagement and student learning: Testing the linkages. Research in Higher Education, Volume 47, pp. 1-33.

Denholm-Price, J. & Soan, P., 2014. Mathematics eAssessment using Numbas: experiences at Kingston with a partially 'flipped' classroom. Presentation at the STEM National Conference.

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 vol. 2, iss. 2.

Gibbs, G. & Simpson, C., 2004-5. Conditions under which assessment supports students' learning. Learning and Teaching in Higher Education, 1(1), pp. 3-31.

Linnenbrink, E. A. & Pintrich, P. R., 2003. The Role of Self-Efficacy Beliefs Instudent Engagement and Learning Intheclassroom. Reading &Writing Quarterly 19.2, pp. 119-137.

Perfect, C., 2015. A demonstration of Numbas, an e-assessment system for mathematical disciplines. CAA Conference .

Sangwin, C., 2013. Computer Aided Assessment of Mathematics. Oxford: OUP.

Trowler, V. & Trowler, P., 2011. Student engagement toolkit for leaders. Leadership foundation for higher education and Higher Education Research and Evaluation.

Warwick, J., 2008. Mathematical self-efficacy and student engagement in the mathematics classroom. MSOR Connections 8.3 , pp. 31-37.

CASE STUDY

Creating statistics e-assessments using Dewis with embedded R code

Iain Weir, Department of Engineering Design and Mathematics, University of the West of England, Bristol, U.K. Email: iain.weir@uwe.ac.uk
Rhys Gwynllyw, Department of Engineering Design and Mathematics, University of the West of England, Bristol, U.K. Email: rhys.gwynllyw@uwe.ac.uk
Karen Henderson, Department of Engineering Design and Mathematics, University of the West of England, Bristol, U.K. Email: karen.henderson@uwe.ac.uk

Abstract

We report on the creation of statistics e-assessments using the Dewis system with embedded R code. Dewis is a fully algorithmic open-source e-assessment system designed and developed at the University of the West of England, Bristol (UWE). Dewis' ability to communicate with the R programming language greatly facilitates the task of generating bespoke data and its subsequent analysis. This approach has allowed us to successfully test students' ability to perform involved statistical analyses on individual data sets and led to the creation of a suite of open access online e-learning modules on the UK national **stats**tutor website. Development of a Dewis-R interface allows the creation of sophisticated e-assessments solely by writing an R script file. The goal is to create a community of Dewis-R practitioners who will be able to author and share relevant, authentic and engaging statistics e-assessments that enrich the learning experience of students.

Keywords: statistics, SPSS, R, e-assessment, Dewis.

1. Background

Great strides have been made in the use of computer aided assessment for mathematics over recent years and many mathematics departments regularly use formative and summative e-assessments for their students (Sangwin, 2013). Such e-assessments can supply challenging practice in mathematical skills that, given a suitable platform and connection, students can access anytime or anywhere and obtain instantaneous and timely adaptive feedback. However, there has been far less progress on using e-assessment to test the ability to perform statistical analyses despite the fact that statistics is a component in many university courses.

Dewis (2012) is a fully algorithmic open-source e-assessment system designed and developed at the University of the West of England, Bristol (UWE). Originally designed to support and assess the learning of mathematics, it is currently used in the fields of business, computer science, nursing, engineering and mathematics. Implemented for the first time in 2007 the system is now well-established and in 2015/16 within UWE and partner institutions, Dewis was used for formative and summative tests to support over 3,500 students involving more than 50,000 assessment attempts.

A recent innovation by Weir, Gwynllyw and Henderson (2015) was to successfully embed R statistics code within Dewis to create e-assessments that could take advantage of the statistical functions available within R (2014). This approach led to the creation of complex statistics e-assessments that replicate full statistical analyses and report writing. These e-assessments have been used on level 2 (second year undergraduate) research skills modules, delivered to 850+ Business School students at UWE, since 2014. Students receive their own unique data set, generated by Dewis, which they are able to download directly into Excel, perform the necessary operation in SPSS before re-logging back into Dewis to submit their answers (Gwynllyw, Weir and Henderson, 2016).

The success of this approach led to the creation of an open access statistics resource which is freely available from the **stats**tutor site (*Dewis on **stats**tutor*, 2015). This resource has been made available under a Creative Commons licence by the authors of this article and reviewed by Dr Nadarajah Ramesh, University of Greenwich following a **sigma** Network Resource Development Grant (2015). The resource comprises a suite of e-assessments together with supporting materials that relate to the statistical activities involved in choosing and carrying out an appropriate one sample test for location (mean or median) on a randomly generated data set. Five e-assessment modules are available, as shown in figure 1, and these may be accessed independently or can be taken sequentially mimicking the flow of a full statistical analysis using the SPSS software package. On accessing the resource a new statistical data set may be generated or an existing data set used. Each module requires the data to have been downloaded to the SPSS statistical package, relevant analysis output obtained and a few questions answered to demonstrate understanding of the results. On submission, the e-assessment system marks the responses immediately and provides full bespoke feedback for inappropriate test choices as well as other incorrect analysis. Videos and instruction pamphlets are accessible as links from each e-assessment, which give clear instructions as to how to carry out the analyses and interpret results using SPSS. These additional resources, together with repeated use of the e-assessment modules, facilitates learning how to identify and employ the correct test on a variety of data sets.

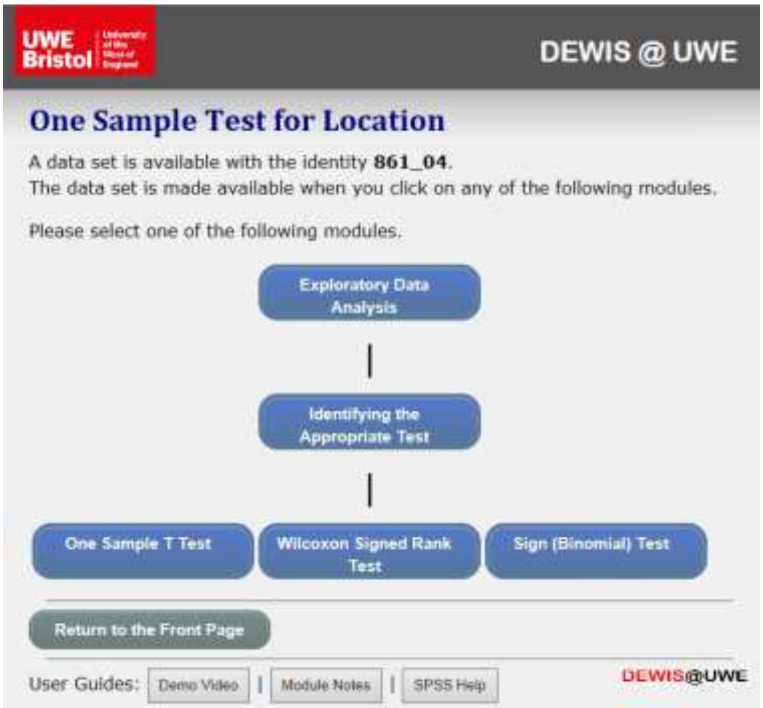


Figure 1. The five e-assessment modules that make up the resource on **stats**tutor.

2. Dewis-R Interface Project

The projects described in Section 1 demonstrate the potential of creating complex, intelligent statistics e-assessments using the Dewis system with embedded R code. However the drawback to expansion and growth of this approach has been in part the complexity of authoring questions. Not only do question authors have to be conversant with the R language they also have to be familiar with the programming of questions within Dewis. The issue of requiring familiarity with the Dewis syntax has been addressed by introducing a simplified Dewis question structure specific to this project. This simplification is possible since the statistical computations in these questions will be performed by R and not by Dewis. As such, Dewis supplies the administration of the assessment but not the computations. This approach should appeal to statisticians who are already familiar with R. The question structure is in the form of a text file partitioned into two parts, the R part and the Dewis part. The Dewis part contains the syntax of the question as presented to the student, together with a declaration of the question parameters that will be used to present the question as well as to perform the marking and feedback. This part of the question is interpreted on installation of the question. On installation of the question, the R part of the question is written as an R script into the Dewis server. This script will be executed on every occasion the assessment is run. An execution of the R script will calculate the question parameters (including the answers) and pipe these to Dewis, enabling Dewis to present the question to the student. Weir and Gwynllyw have been awarded a UWE Teaching and Learning grant to push this idea forward and to create a cross-faculty statistics e-assessment Learning & Teaching community. The goal is to create a community of Dewis-R practitioners who will be able to author and share relevant, authentic and engaging statistics e-assessments that enrich the learning experience of students.

In this case study we demonstrate the Dewis-R interface by showing two examples. It is hoped that this will generate interest from the wider statistics community to trial this new approach of authoring. The examples presented here and others can be found at the Dewis-R Statistics Resources website (2016).

2.1. Example 1: Calculating and comparing a mean

The first example is by design very simple enabling us to demonstrate a basic Dewis-R script. The example involves the calculation of a mean from a sample of IQ scores and its subsequent comparison of it with the theoretical population mean of 100. The student is required to download an Excel file that contains a random data set to analyse. The sample mean is required to be given to one decimal place and from a dropdown menu the student has to report whether it is higher than, lower than or the same as the population average of 100. Thus this e-assessment comprises solely of one numerical input and one dropdown input. Figure 2 shows a screenshot of the question that is presented and, for illustration purposes, is of a student's attempt that has one correct and one incorrect answer. Note that in the subsequent feedback we colour code correct answers green and incorrect answers red, and hence have used this convention here.

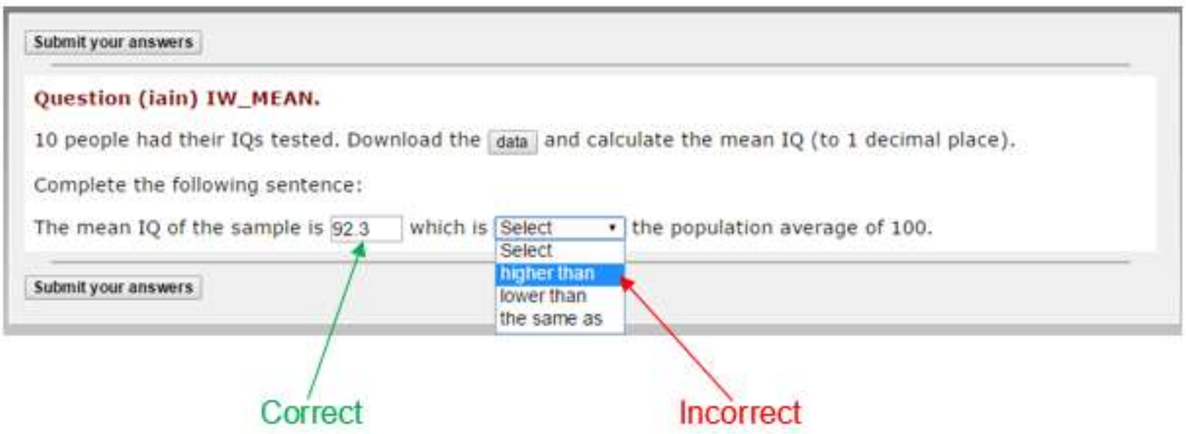


Figure 2. Screenshot of Example 1: Calculating and comparing a mean. For illustration purposes the student has entered a correct mean but has an incorrect comparison.

Figure 3 demonstrates the feedback screen. The 'Question' section reminds the student of the question and the blue text indicates where numerical values or dropdown menu choices were required. The 'Solution' supplies the correct answers to the sentence the student needed to complete. The 'Report' indicates, with colour coded marking, what the student has answered correctly or incorrectly. Incorrect answers can be supplemented with further formative feedback. In this example there is a sentence further explaining the students mistake.

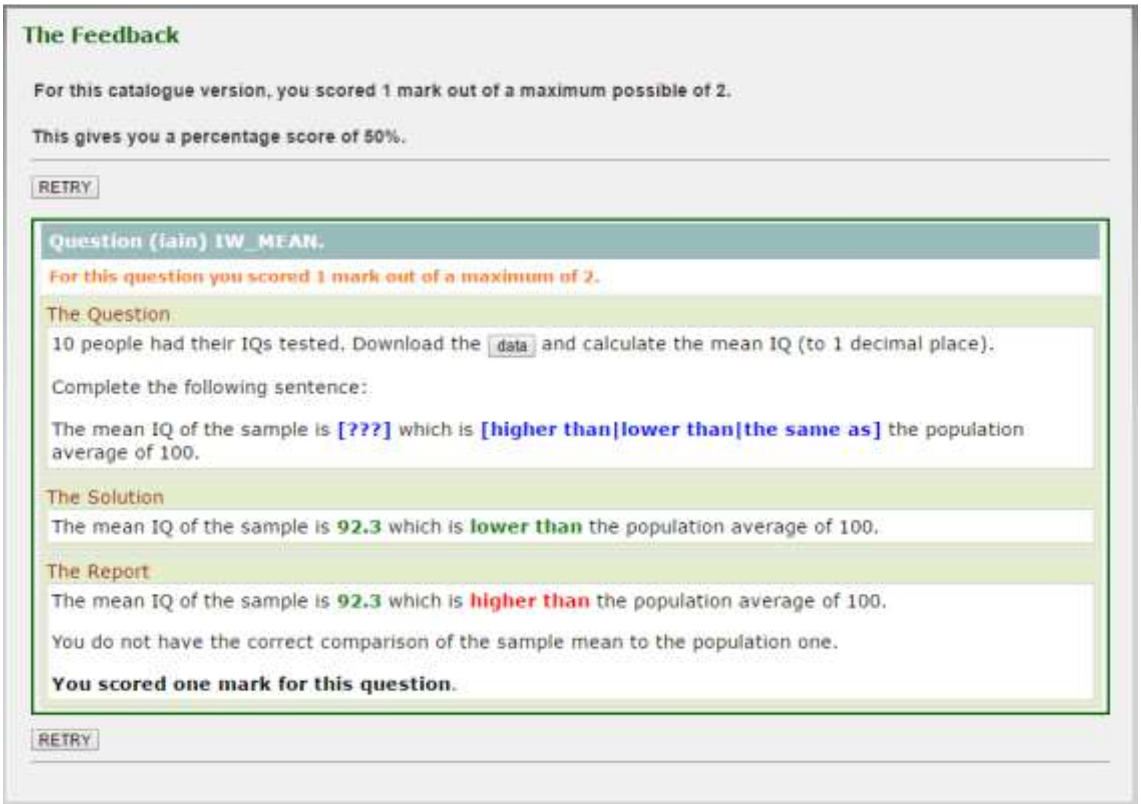


Figure 3. Screenshot of the feedback for Example 1 which comprises three sections: ‘Question’, ‘Solution’ and ‘Report’. Note that the latter indicates that the student has an incorrect answer to the mean comparison.

The Dewis-R interface requires a script file that comprises two parts; the Dewis part relates to install-time question construction and the R part relates to run-time data generation and analysis. We shall now present the script that generates Example 1. Note that the explanation of syntax in any depth is beyond the scope of this article. The syntax we are developing is in its infancy; in time, our aim is to publish a manual with full details.

Figure 4 presents the Dewis install-time question communication script; it comprises HTML code and our own coding tags which we shall briefly explain. The <PARAMETERS> tag code is where the number and type of each question input is defined, each of which is assigned an ID number; here we have two. The <INPUT> tags define the first (ID=1) to be numerical and the second (ID=2) to be a dropdown with three choices. The name of the R run-time function that will generate data and input answers are also defined. The <ON_SCREEN> tag code is where the question text is composed. Question and feedback reporting are in a format that has three sections; ‘Question’, ‘Solution’ and ‘Report’. The <SHOW_IN> and <HIDE_IN> tags dictate what text appears in each of the sections. <IF_WRONG> tags allow appropriate feedback relating to any wrong answers.

```
#####################################################################
################# DEWIS install-time question construction ###########
#####################################################################

<DEWIS_INSTALL>

#===================== Define run-time function and define inputs ===

<PARAMETERS>

 <RUN_TIME_FUNCTION='dewis_run()'/>

 <NUM_IDS=2>

 <INPUT TYPE=NUMERICAL ID=1 NAME='the mean IQ'/>

 <INPUT TYPE=DROPDOWN ID=2 NAME='the mean comparison'>
  <OPTION>higher than</OPTION>
  <OPTION>lower than</OPTION>
  <OPTION>the same as</OPTION>
 </INPUT>

</PARAMETERS>


#===================== On screen question text =====================

<ON_SCREEN>

<SHOW_IN='QUESTION'>
 10 people had their IQs tested. Download the <DATALINK>data</DATALINK> and
 calculate the mean IQ (to 1 decimal place).
 <p></p>
 Complete the following sentence:
 <p></p>

<SHOW_IN='REPORT SOLUTION'>
 The mean IQ of the sample is <INPUT ID=1/> which is <INPUT ID=2/> the
 population average of 100.

<HIDE_IN='QUESTION SOLUTION'>
<IF_WRONG ID='1'>
 <p></p>
 You have not supplied the correct sample mean value.
</IF_WRONG>
<IF_WRONG ID='2'>
 <p></p>
 You do not have the correct comparison of the sample mean to the population one.
</IF_WRONG>

</ON_SCREEN>
</DEWIS_INSTALL>
```

Figure 4. Example 1 script: Dewis install-time question construction.

Figure 5 presents the script that comprises the R function that is called at run-time. This function deals with all data generation and calculations plus a written communication of the correct input answers and data values for Dewis to read.

```
#<R>

######################################################################
################# R run-time function ################################
######################################################################

dewis_run=function(){

#===================== Data generation and calculations =============

# get mean from 10 integer observations from N(100,15^2)
 IQ=round(rnorm(10,100,15))
 meanIQ=round(mean(IQ),1)
 IQdata=data.frame(IQ)

# comparison to mean of 100 1="higher"/2="lower"/3="the same"
 if (meanIQ>100){
  meanComp=1
 }else if(meanIQ<100){
  meanComp=2
 }else{
  meanComp=3
 }


#===================== Assign and communicate correct answers =======

 cat("
 <DEWIS_PARAMS>

# Assign correct answers for each input
 <CORRECT ID=1>",meanIQ,"</CORRECT>
 <CORRECT ID=2>",meanComp,"</CORRECT>

# Printing out of generated data
 <DEWIS_DATA COLS=1>
 ")
 print(IQdata)
 cat("
 </DEWIS_DATA>

 </DEWIS_PARAMS>
 ")

}
```

Figure 5. Example 1 script: R run-time function.

2.2. Example 2: Reporting a correlation

The second example is more adventurous and relates to reporting a test of Pearson's correlation coefficient. We shall restrict ourselves merely to screenshots that demonstrate some of the other tags that we are developing. Figure 6 is a screenshot of the question with various tags identified. In this example <DATALINK> provides a button for downloading the bivariate data of random sample size that has a correlation that is of a random strength and sign. The file format in this case is a CSV file, which is appropriate considering the wide range of operating systems used by students. Dynamic graphics generated via R during run-time can be included in questions using the <IMAGE> tag; in this example a scatterplot of the data is displayed. Any HTML tags may be included in a question; here we use <a> to include a link to a supporting video. Depending upon a significant or non-significant test result, different reporting statements are required; the <AREA_CHOICE> tag allows the student to pick the appropriate statement to report. In figure 7 the two competing report statements are displayed; it can be seen that extra inputs are required when reporting a significant correlation.

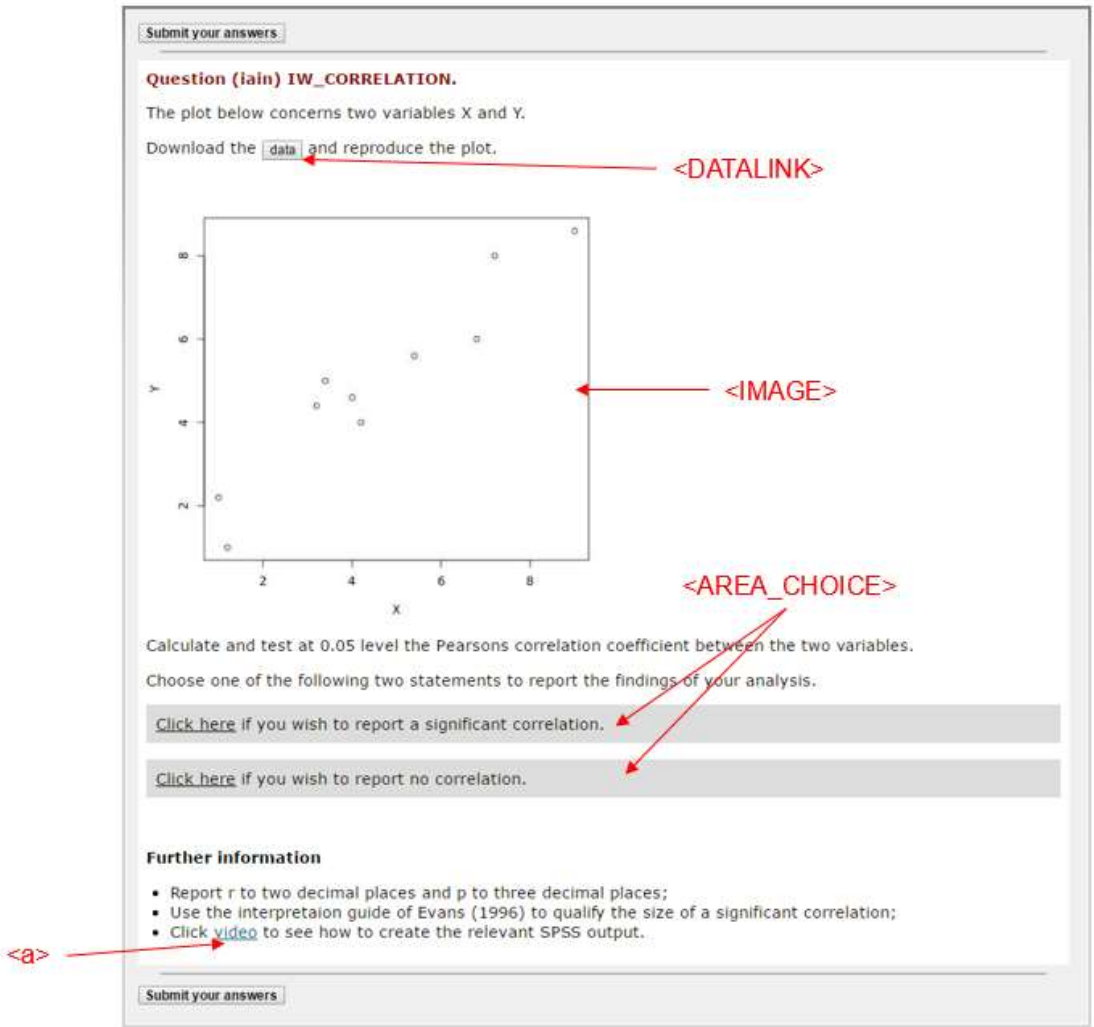


Figure 6. Screenshot of Example 2: Reporting a correlation. The use of various script tags is indicated.

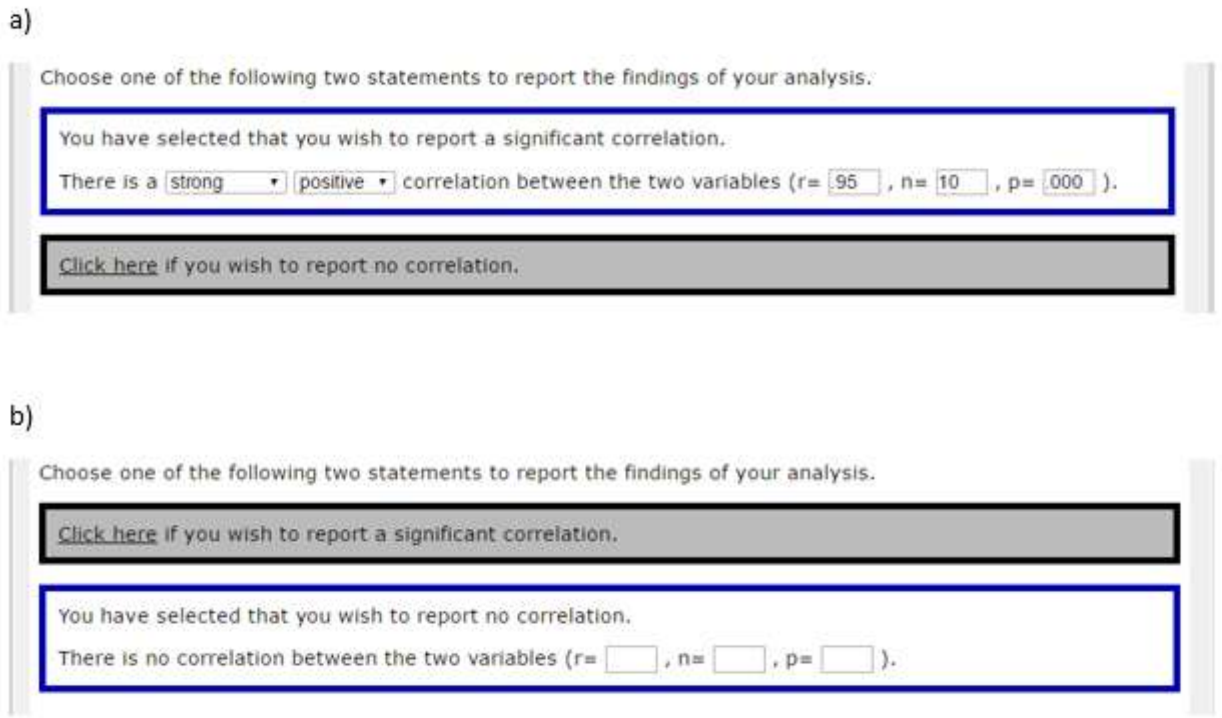


Figure 7. Example 2 competing correlation reporting statements. In a) additional information about the strength and sign of a significant correlation is required when reporting a significant correlation that is not necessary in b) when reporting that there is no correlation.

3. Discussion

Although early in the development stage, we have illustrated, through two examples, the power and ease with which meaningful statistics e-assessments can be created. This approach has allowed other statisticians within UWE to author questions and use e-assessment in their modules for the first time. We would welcome collaboration from the wider statistics community to trial this new approach of authoring and to become involved in the development of the interface functionality. We anticipate that the creation of a Dewis-R statistics e-assessment community will help the dissemination of this methodology and thus enhance the student experience. Staff in the community will benefit both from the sharing of learning & teaching innovation and the statistics practice aspects.

4. References

Dewis resources on **stats**tutor, 2015. Available at: http://www.statstutor.ac.uk/search/?q=dewis [Accessed 14 May 2016].

Dewis welcome page, 2012. Available at: http://dewis.uwe.ac.uk [Accessed 14 October 2016].

Dewis-R Statistics Resources website, 2016. Available at http://dewisprod.uwe.ac.uk/resources/stats/index.html [Accessed 19 October 2016].

Gwynllyw, D.Rh., Weir, I.S., and Henderson K.L, 2016. Using DEWIS and R for multi-staged statistics e-assessments. *Teaching Mathematics & Its Applications*, 35(1). pp.14-26.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/ [Accessed 23 December 2016].

Sangwin, C., 2013. *Computer Aided Assessment of Mathematics*. Oxford University Press.

Sigma Resource Development Projects Funded, 2015. Available at: http://www.sigma-network.ac.uk/resource-development-projects-funded/ [Accessed 14 May 2016].

Weir, I., Gwynllyw, R. and Henderson, K., 2015. Using technology to inspire and enhance the learning of statistics in a large cohort of diverse ability. In: IATED, ed. (2015) *Edulearn15 Proceedings*. IATED.

RESEARCH ARTICLE

STACK with state

Matti Harjula, Dept. Mathematics and Systems Analysis, Aalto University, Helsinki, Finland. Email: matti.harjula@aalto.fi
Jarmo Malinen, Dept. Mathematics and Systems Analysis, Aalto University, Helsinki, Finland.
Antti Rasila, Dept. Mathematics and Systems Analysis, Aalto University, Helsinki, Finland.

Abstract

The question model of STACK provides an easy way for building automatically assessable questions with mathematical content, but it requires that the questions and their assessment logic depend only on the current input, given by the student at a single instant. However, the present STACK question model already has just the right form to be extended with state variables that would remove this limitation. In this article, we report our recent work on the state-variable extension for STACK, and we also discuss combining the use of state variables with our previous work on conditional output processing. As an outcome, we propose an expansion to the STACK question model, allowing the questions to act as state machines instead of pure functions of a single input event from the student.

We present a model question using the state variable extension of STACK that demonstrates some of the new possibilities that open up for the question author. This question is based on a finite state machine in its assessment logic, and it demonstrates aspects of strategic planning to solve problems of recursive nature. The model question also demonstrates how the state machine can interpret the solution path taken by the student, so as to dynamically modify the question behaviour and progress by, e.g., asking additional questions relevant to the path. We further explore the future possibilities from the point of view of learning strategic competencies in mathematics (Kilpatrick et al., 2001; Rasila et al., 2015).

Keywords: STACK, state machine, interactive question.

1. Introduction

Various types of e-learning systems have been used in teaching mathematics for a long time. Currently, there is a wide selection of both commercial and open source systems (such as Maple T.A., Numbas, SOWISO, STACK, etc.) that are capable not only of automatic assessment but also giving useful feedback according to students' inputs. In some of the older systems, the assessment logic is realised rather restrictively by using multiple choice questions or by string comparisons. More advanced technologies make use of symbolic computations with the aid of a Computer Algebra System (CAS) (such as Maple or Maxima) for the evaluation of students' inputs. These systems make it a relatively straightforward task to create teaching materials consisting of a series of mutually independent questions, or a series of questions with a common theme. It is, however, difficult to produce more multifaceted materials that are able to approach an underlying narrative from many directions in a game-like manner. This is a serious shortcoming in the age of simulations-based, interactive e-learning technologies.

One way of introducing modest game-likeness to current systems is to use a series of independent questions so that the selection of the next question depends on student's earlier success. This amounts to applying Computer Adaptive Testing (CAT) techniques in material design. Coupling individual questions in this way is a rather restrictive form of storytelling since it only allows queuing pre-made questions in a changing order. It does not allow the questions to adapt to the student's answers at a finer level.

The next step is to allow the question logic to do adaptation within the question itself, resulting in a more direct, more focused, and more immediate adaptation. In order for this to be possible, the question logic needs some kind of memory (i.e., *internal state variables*) for the current and past forms of the question. Another related desirable feature is an ability to access the memory of the ambient system (such as Moodle for STACK) through *external state variables*.

STACK (Sangwin et al., 2016) has been used at Aalto University since 2006 (Sangwin, 2013), and it has been extensively improved to meet the needs of material developers during that time. However, adaptation of the question based on the student's previous answers has been difficult to implement. Adding the required memory features changes the earlier stateless question model to a stateful model. At the moment, we have an improved STACK question model (and the system executing it) where both the stateless and the stateful questions can coexist (Harjula, 2016). The *stateful question model* adds complications to the implementation and design of materials as well as to the technical side of the executing system.

2. Abstraction of the stateful question

A stateful question can be understood as a set of traditional parametric questions that **(i)** can transfer the student to another such question based on the student's actions, and **(ii)** set the parameters of the consequent question accordingly. Within a stateful question, the component parametric questions are called *scenes* that must be merged to a whole. In other words, the scenes are described in terms of traditional parametric questions. As opposed to coupling of similar questions in traditional CAT, the parameterisation of scenes can be carried of using anything that the CAS can construct (based on the student's actions and the earlier state) as parameters. Traditionally, most question coupling approaches only care about the grade, success, or failure in the preceding questions.

Scenes can be considered as states of a Finite State Machine (FSM). However, the scenes are parametric and parameter values may change during transitions between scenes or in transitions within a single scene. This leads to situations where each distinct scene (i.e., state of the FSM) and the associated parameter values give rise to what we could call the (full) state. In particular, a question with only one scene can be stateful if it has parameters that it can change based on the inputs from the student, and, hence, modify the output it presents to the student. Having just one scene is enough, e.g., for implementing delayed feedback and conditional hints.

The concise terminology of stateful questions is as follows:

state defines the values for all parameters of all scenes and keeps track of the active scene.

scene is a parametric component question that may transfer the student to another scene depending on the input received.

transition is the actual act of transferring between the states, consisting of the scenes and the associated parameter values.

transition condition defines when the inputs from the student and the current parameter values lead to a state transition.

path is the listing of previously visited scenes in sequential order, including the current scene.

In the context of the stateful STACK, the merging of scenes to a single stateful question can be done by using conditional rendering of the question presentation (i.e., text and graphics) where only the details relevant to the current state are presented. The triggering of state transitions can be done using CAS level coding within the free-code portion of the normal response generation.

3. Modifications to the question model of STACK

The original STACK handles an user input event along the following lines:

1. Reconstruct the *question variables* based on a (potentially earlier) generated, original random number seed. Using the original seed, generate all the random parameter values required by the question.
2. Identify from the input of the student which ones of the (potentially many) input fields of the question have received valid values.
3. Find those *potential response trees* (PRTs) within the question that have received valid input values for all of their input variables required for processing the PRT. Then for each such PRT:
    a. Evaluate the *feedback variables*, i.e., the free-code portion of the PRT to preprocess the input values. This portion and the PRT nodes may reference the question variables and are in that sense parametric.
    b. Traverse the nodes of the PRT starting from the root and evaluating each the binary *answer test* of each node in order to decide to which branch of the PRT to continue, until the branch ends. At each node, modify the points given and the penalties accrued for this question part for this attempt, based on the binary answer test result. Also, each node may append to the feedback presentation connected to this PRT, again based on the test result.
4. Render the question presentation based on the variables reconstructed at step 1, inserting the feedback generated by the PRTs and input validation inline to it, if required.

As this description is about a stateless model, the processing of the answer does not change the way the forthcoming processing rounds would act. Here, the general presentation of the question stays the same, and only feedback or validation messages are added to it.

The stateful extension modifies the above described outline so that the free-code portion of PRTs is allowed to write values into the state. These values can later be referenced in all other stages of question processing, including the rendering of the question presentation. In rendering, it is now possible to select the parts to be displayed based on the state. Moreover, the PRTs may evaluate the properties of the student's answer based on the state constructed from previous answers, contrary to the original STACK.

Note that the original random number seed is still in control of all random parameter values. This means that these parameter values are “frozen” to the same even if the question had cycles and the author would like to construct new random values for every time a specific scene is returned to. It would be more desirable for the author to be able to decide, based on the path, whether one would present the scene using the original random number seed or a new seed constructed from the path so as to avoid repetition.

To identify which questions have state, the proposed extension requires that all state variables (both local to this question instance and external) must be declared by the author. The declaration must also define if the variable is read-only, and set the variable's default value should it not have been initialised elsewhere. Questions without writable state variables are by definition stateless; i.e., a stateless question may have references to an external state that is used similarly as the random number seed in question initialisation. If the question, however, has a writable state variable, all CAS evaluations performed by the question must be augmented with calls that transfer the current state to the CAS. Conversely, the relevant output from CAS must be stored to a state variable.

When a question references external state variable, its value is stored as a local variable at the initialisation of the question. All references to this external variable are then rerouted to the local variable, thus eliminating problem due to changing external state variable values while the question instance is being used. This, however, makes it difficult to change the external state as the current value is not shown directly to the question. To overcome these complications, the external state is typically handled as a data structure that can only be updated (e.g., a counter that gets incremented, a set that gets augmented, etc.) rather than entirely rewritten.

4. Three classes of stateful questions

The stateful question type makes it feasible to produce materials with which higher abilities – such as strategic competencies – can be trained and assessed instead of drilling. As examples, consider these classes of stateful questions having narratives:

1. The *constrained-path* questions demonstrate the progress of some algorithm or a more general collection of fixed rules and verify that the student's answers match the steps. The well-defined states and steps can be easily reproduced by CAS for verification. Collections of rules may leave the student some freedom of choice so as to use strategic thinking for attaining the desired goal. Erroneous values are typically not accepted in answers but the student is allowed to go in a wrong direction for a while.
2. In *sudoku* questions, some initial setting is given with incomplete details. The student is asked to gradually fill in the missing details by logical reasoning from what is already regarded as known. The next detail to fill in is chosen randomly, based on a dependency graph that describes what can be concluded by relatively few logical steps from what is already known. One may also consider accepting mistakes and allowing them to propagate in reasoning without letting the student know, but then there has to be a way for the student to fix the errors afterwards.
3. A dynamical mathematical (state space) model of some phenomenon lies in the core of *model-based* questions. The student's answer is a control input, and the state transitions are computed by the model. There are means for evaluating whether (and at what cost) the desired target state has been achieved, and there may even be dynamically generated new targets.

All these types can be realised as stateful STACK questions but their designs may get very complex in the two latter classes since many states are required to handle the storytelling logic and to separate it from the state of the mathematical model. The designs may benefit from a narrative backbone such as Interactive Storytelling (IS) taken from the game design context (Crawford, 2012). Within mathematics e-learning, storyboards have not been widely used even though various requirements for successful high-level system designs have been proposed in, e.g., Devlin (2011). We emphasise that the storytelling aspect is more important in authoring than programming skills since a stateful question must react sensibly and reasonably to students' various answers.

The three classes introduced above are not to be understood as exclusive or canonical but they help in characterising and comparing different types of possible stateful questions. Obviously, stateless questions are one end of the scale, and we could call them the *no-path* type in comparison to constrained-path questions where allowed transitions between scenes are very limited; however, not excluding infinite loops. Sudoku type questions have transitions between scenes that are best described as non-trivial directed graphs without cycles. Similarly, model-based questions are expected to lead to undirected, dense graphs as transition paths between highly parametric scenes.

The logical end point of this scale could be called the *free-path* question type where so little restrictions in transitions between scenes are posed that the student may not even notice any restrictions of freedom. As we move towards the model-based and free-path designs, we expect that methods from IS become particularly useful, starting from constrained narrative generation (Porteous and Cavazza, 2009) and leading to full-on drama management. We further expect that the design of a pedagogically relevant free-path question in mathematics to be very difficult.

5. Stateful question design principles

One example of a stateful question is given below. A prototypal design process for this kind of questions is outlined as follows:

1. Design the narrative for the question, consisting of scenes that can be represented as usual STACK exercises.
2. Describe the narrative as a FSM diagram such as figure 1 where each scene corresponds to one or several potential tasks. Enumerate all scenes and describe their scene transitions in response to the student's answers and parameters.
3. Specify what information the scenes need as parameters. Define the necessary internal state variables to convey this information.
4. Produce the rendering code of the user interface for each scene.
5. Realise each scene like an usual STACK exercise whose response tree is activated by the student's answers. Add transition conditions and transitions to the free-code portion of the PRTs to produce transitions between the scenes.

The author should minimise the number of scenes and state variables without making the code unreadable. The generation of the question texts is challenging as the text depends on the full state, leading to many variants.

6. An example question of constrained-path type

We present an example of a constrained-path question involving integration by parts where the integrand includes a monomial. The student is allowed to split the integrand into two parts at will, and then use integration by parts to compute the antiderivative. Some strategic competency is required in addition to computation skills. Indeed, the degree of the monomial may increase as the result of a silly choice when splitting $f = uv'$, and the student gets further away from the right solution.

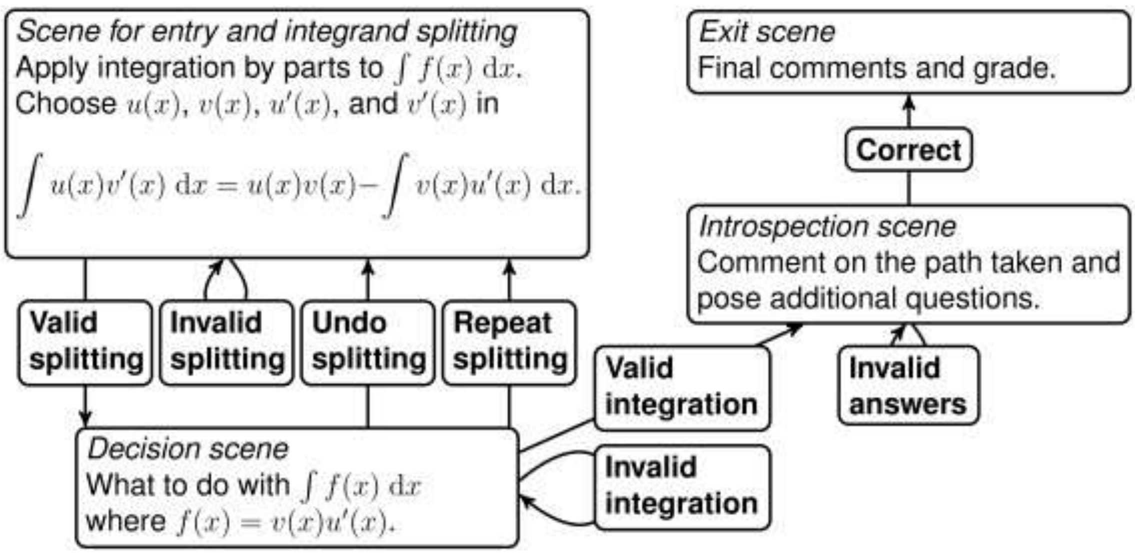


Figure 1. The state transitions of the example question having four scenes.

The question consists of four scenes as described in figure 1, two of which are being visited repeatedly. The user interfaces associated to these scenes are given in figures 2-4 below. More precisely, the scenes are as follows:

- **Entry and integrand splitting:** The initial or current integration by parts exercise is displayed by this scene, and the student is asked to split the integrand for the integration by parts process.
- **Decision:** In this scene, the student is shown the valid splitting given earlier. Now it is possible to revert the earlier splitting, choose applying integration by parts again, or just compute the integral by other means.
- **Introspection:** The student arrives at this scene after having produced the correct answer. Comments are given on the solution path of the student, and relevant control questions are asked concerning the solution strategy.
- **Exit:** This scene shows the points and penalties with final comments.

The state transition rules of the question are described in figure 1. These allow the student to go on integrating by parts as far as they wish (within storage/memory limitations). The question is initialised with a randomised degree of the monomial in the integrand. The question state only keeps track of the two expressions generated by the process, the ejected part and the remaining integral, as a list consisting of all the values generated by previous splittings, the current scenes name, and some flags signalling which warnings have been given. The most difficult part in the authoring of this question is describing the logic for checking if the given splitting of the integrand leads to the desired direction in integration by parts process.

The stateful question type makes it possible to give specific feedback to the student after complex conditions are met. In the example question, we start to warn about going in the wrong direction only after multiple steps of integration by parts in that direction. The warnings get more direct if the student continues even further. It is also detected if the student returns to the original expression after many steps, or if the student tries to integrate by parts an expression that does not require it. A simple way of constructing the final grade is to keep track of the warnings that have been given, and this is the approach used here. More complicated evaluations can be based on the trajectory that is stored as a list of uv and vu' expressions.

Any of the four scenes in the example question can be implemented without using states. The point of the example question is in joining those scenes as a dynamic storyline and evaluating the path the student takes. This is impossible to achieve without having the kind of memory that the state variable extension provides.

7. Conclusions

The state variable extension to STACK was proposed. The extension can be used to produce e-learning materials that respond more adaptively and intelligently than earlier such materials. An elementary example involving integration by part was given to show some of its technological and pedagogical potential.

As *STACK with state* gets technically more mature, the next big step is to build advanced tools to simplify the question authoring process. Drawing diagrams such as figure 1 and manually translating them into conditional branching statements is not the way to go when building large scale applications – it is better to have a more refined “method in madness”. The state dependent randomisation of questions is highly nontrivial, and its framework remains unspecified at the moment. Furthermore, there is need for analytical systems for **(i)** evaluating the structure of a stateful question, (i.e., static code analysis), and for **(ii)** analysing in bulk the solution paths (i.e., the state trajectories) generated by the students. The latter requirement relates to learning analytics.

In this article, we concentrated on finite states (scenes) and internal state variables. External state variables are used for accessing the global state in an ambient system (such as Moodle), providing an interface for an external learning analytics system, or used for sharing student specific information between different stateful STACK questions.

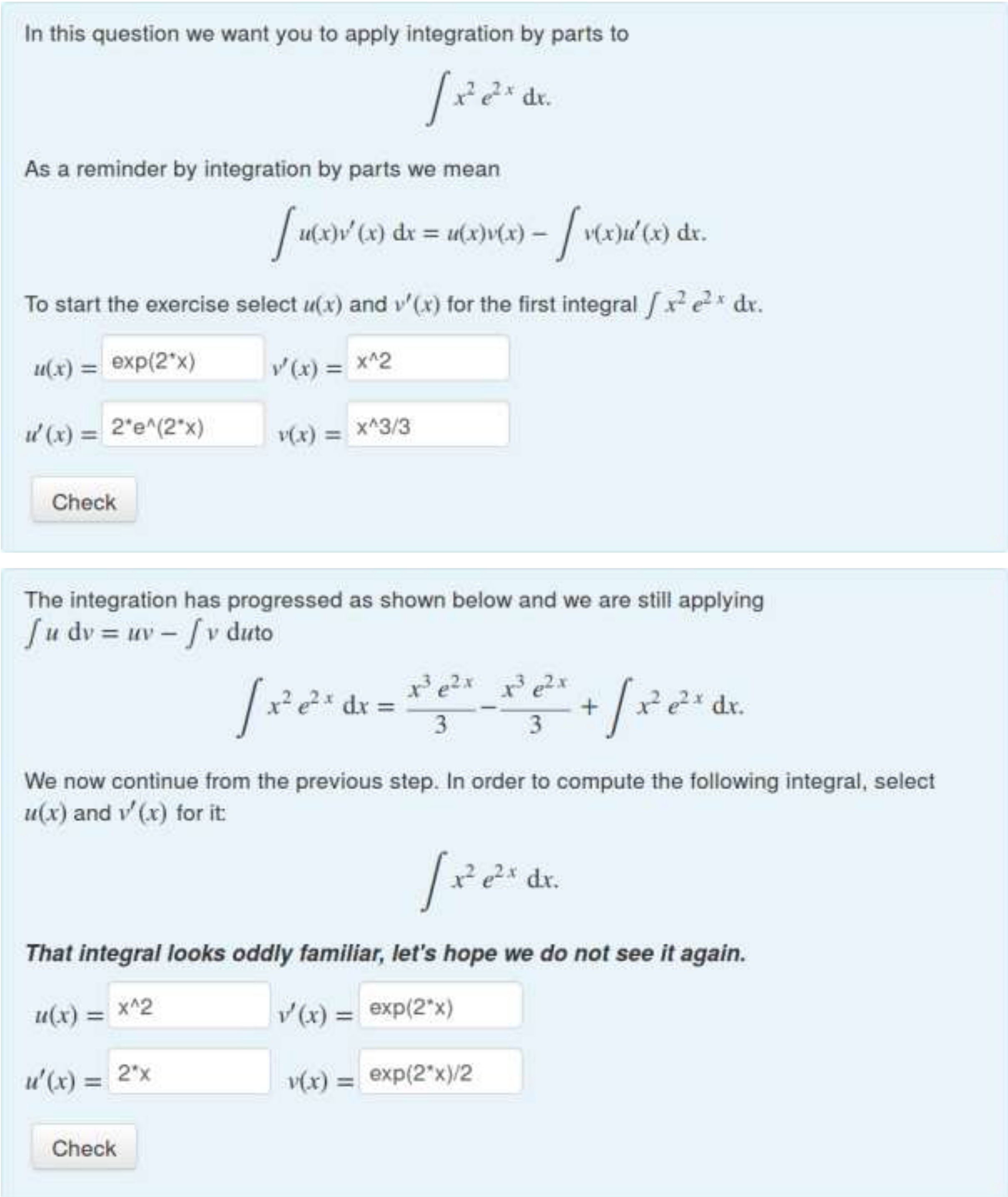
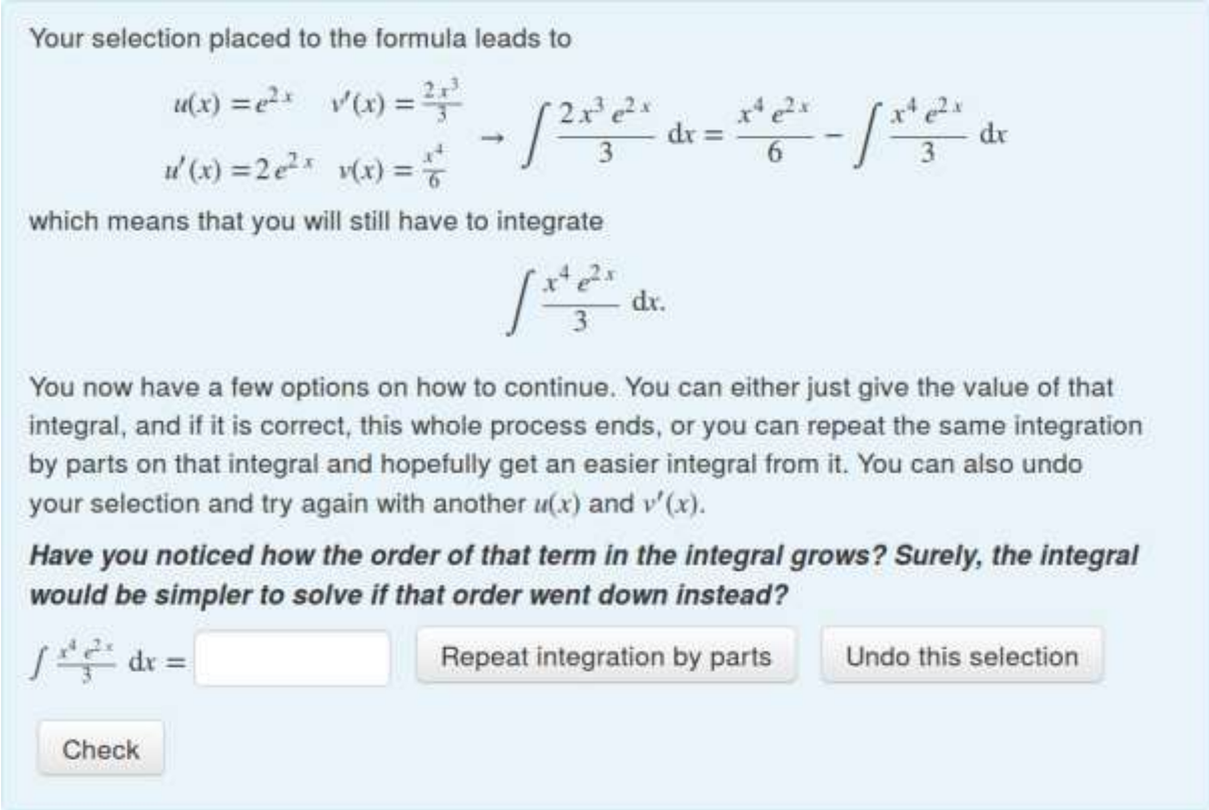


Figure 2. The entry scene of the example question presents the currently relevant integral and asks to split the integrand. In the lower panel, the student has returned to the starting point after some unlucky choices.



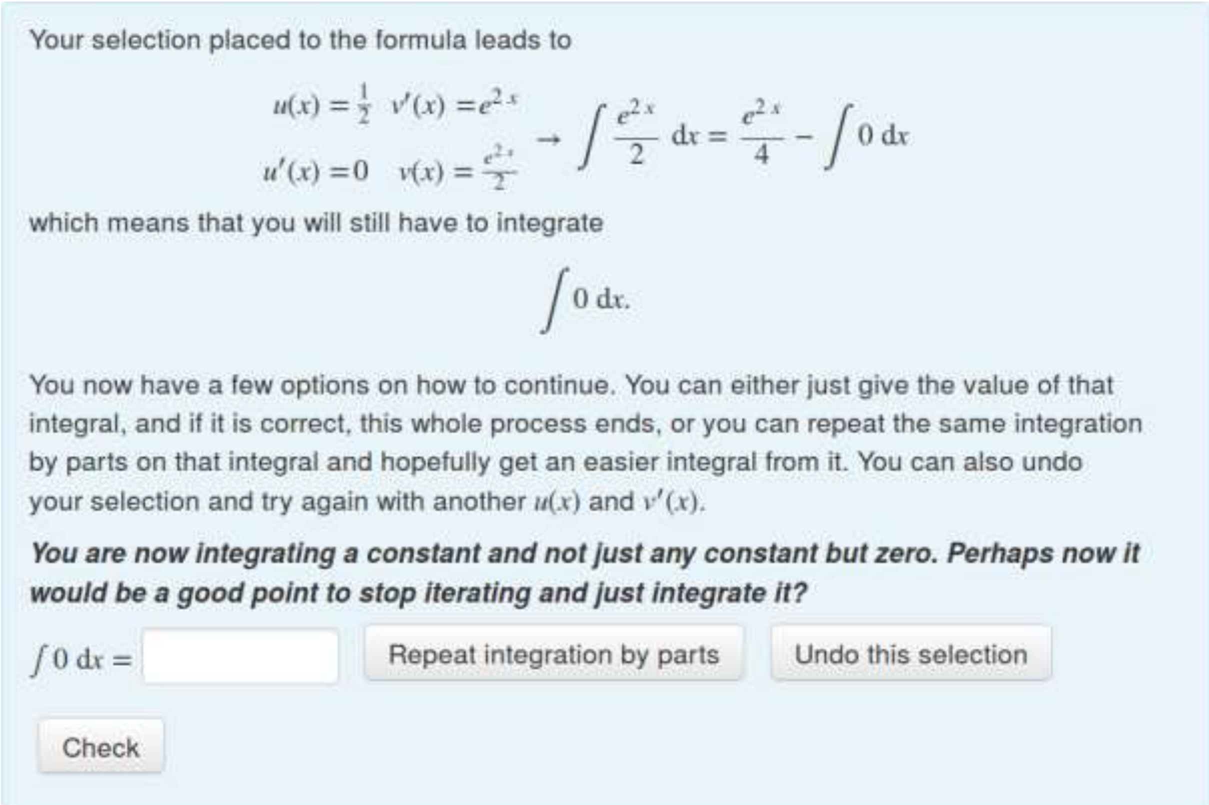


Figure 3. The decision scene gives information of the current situation and allows the student decide how to proceed. The student may continue to integrate by parts even if the integrand is zero.

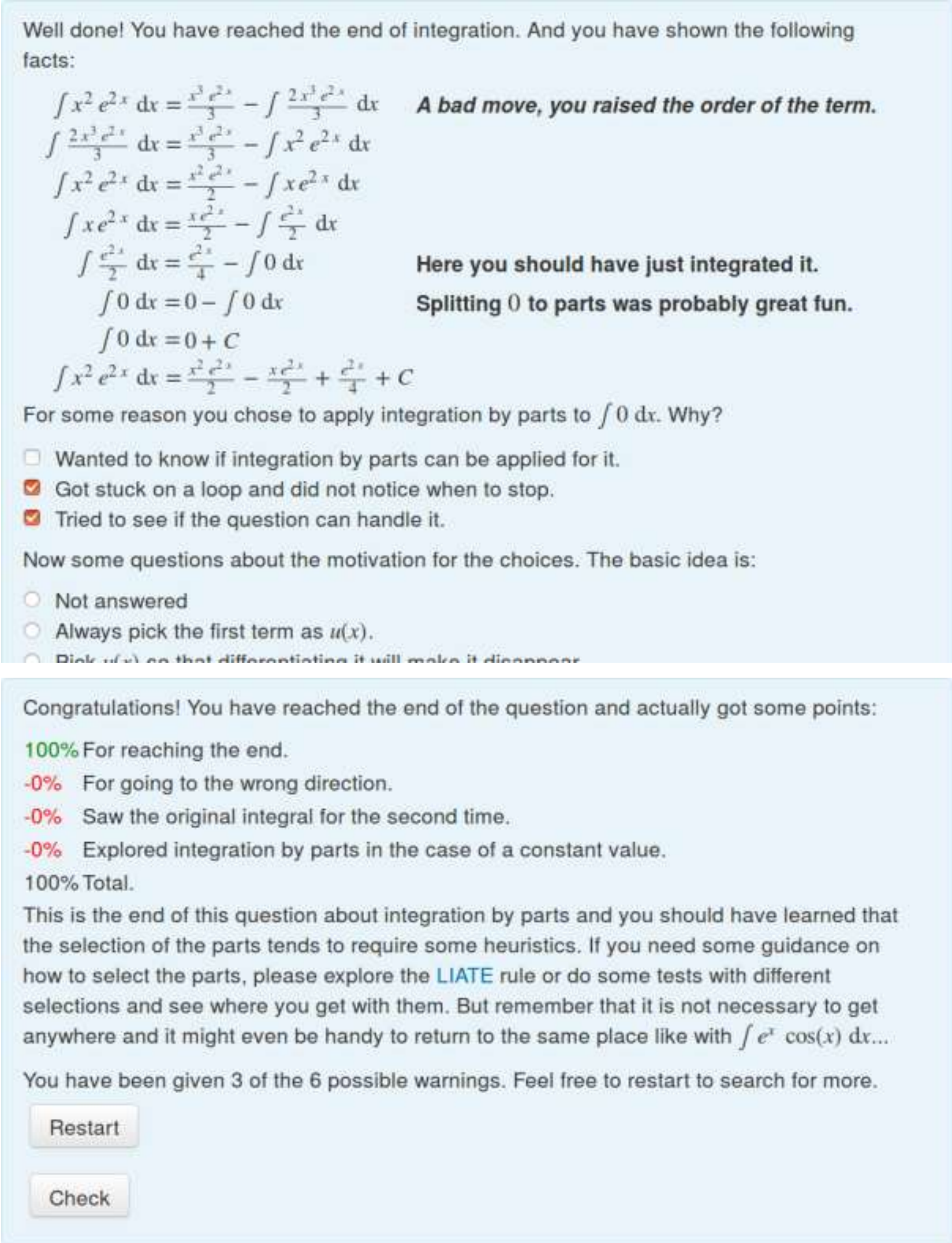


Figure 4. The introspection scene presents multiple choice questions about matters that should have become apparent during the solution process. The exit scene gives the grading with comments.

We conclude that the most exciting aspect is the exploring the new types of mathematical questions made possible by the proposed *STACK with state*, its future modifications, and other improvements. We believe that a practically limitless *terra incognita* of possibilities opens up for game-like e-learning materials in mathematics, challenging the current technological and pedagogical paradigms.

8. References

Crawford, C., 2012. *Chris Crawford on Interactive Storytelling*. New Riders.

Devlin, K., 2011. *Mathematics Education for a New Era: Video Games As a Medium for Learning*, 1st Edition. A. K. Peters, Ltd., Natick, MA, USA.

Harjula, M., 2016. Stateful extension, frozen proof of concept code version. Available at: https://github.com/aharjula/moodle-qtype_stack/tree/EAMS-frozen-state [Accessed 12 December 2016].

Kilpatrick, J., Swafford, J., Findell, B., 2001. *Adding It Up: Helping Children Learn Mathematics*. Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press.

Porteous, J., Cavazza, M., 2009. Controlling narrative generation with planning trajectories: the role of constraints. In: *Joint International Conference on Interactive Digital Storytelling*. Springer Berlin Heidelberg, pp.234–245.

Rasila, A., Malinen, J., Tiitu, H., 2015. On automatic assessment and conceptual understanding. *Teaching Mathematics and its Applications*, 34 (3), pp.149–159.

Sangwin, C., 2013. *Computer Aided Assessment of Mathematics*. Oxford University Press.

Sangwin, C. J., Hunt, T., Harjula, M., et al., 2016. STACK, code of the master version. Available at: https://github.com/maths/moodle-qtype_stack [Accessed 12 December 2016].

RESEARCH ARTICLE

MathTOUCH: Mathematical Input Interface for E-Assessment Systems

Shizuka Shirai, Department of Human Environmental Sciences, Mukogawa Women's University, Nishinomiya, Japan. Email: shirai@mukogawa-u.ac.jp
Tetsuo Fukui, Department of Human Environmental Sciences, Mukogawa Women's University, Nishinomiya, Japan. Email: fukui@mukogawa-u.ac.jp

Abstract

In 2012, we developed a new mathematical input interface with Java, named MathTOUCH. The interface facilitates the acceptance of a mathematical expression as input by interactively converting from a colloquial-style mathematical text (string). This input method enables users to input almost any mathematical expression without learning a new language or syntax. However, the user requires a Java-compliant device for mathematical input. In this study, we developed a reconstructed version of MathTOUCH by using JavaScript (HTML5) to enable the use of MathTOUCH across various devices. The result of our evaluation showed that students are able to practice their mathematical work on STACK using the reconstructed MathTOUCH as well as the previous version of MathTOUCH.

Keywords: math input interface, mathematics, human-computer interaction, STACK.

1. Introduction

In recent years, e-assessment systems, for example STACK, WeBWorK, Maple T.A., Numbas, and Math on Web (Osaka Prefecture University, 2004), whereby learners are able to answer mathematical questions by entering mathematical expressions, have been used for instructing students at many universities. However, current standard interfaces that accept mathematical expressions as input are either text-based or structure-based interfaces, and are cumbersome for novice learners to use. For example, text-based interfaces accept input according to the CAS command syntax. Furthermore, it is difficult for the users to imagine the desired mathematical expressions because the input is not in WYSIWYG format. On the other hand, the advantage of structure-based interfaces is that learners are able to operate in WYSIWYG. In addition, users are able to input mathematical template icons with the help of a GUI. Therefore, they do not need to remember CAS command syntax as for the text-based interface. However, users have to understand the structure of the mathematical expressions they require and should be able to select the mathematical template icons from the GUI in the correct order (Pollanen, Wisniewski and Yu, 2007). Furthermore, it is troublesome to make corrections later (Smithies, Novins and Arvo, 2001).

Fukui (2012) attempted to overcome these shortcomings by proposing a new mathematical input interface, named MathTOUCH. MathTOUCH facilitates interactive conversion from a colloquial-style mathematical text to the desired two-dimensional mathematical expressions. The results of a previous study of ours (e.g. Shirai and Fukui, 2014) showed that novice mathematics learners found MathTOUCH to be user-friendly. However, the users have to use a Java-compliant device in order to use MathTOUCH.

In this study, we reconstructed MathTOUCH using JavaScript to make MathTOUCH available not only on Java-compliant devices but also on various other devices. We evaluated the effectiveness of the reconstructed version of MathTOUCH (hereafter abbreviated as RMT) by investigating whether students are able to practice mathematical work using the reconstructed MathTOUCH at the same learning rate as with MathTOUCH based on Java (hereafter abbreviated as PMT).

2. Proposed interface

2.1. MathTOUCH

MathTOUCH is a mathematical input interface developed using Java. MathTOUCH enables users to input the desired mathematical expressions by converting colloquial-style linear strings (Fukui, 2012). For example, if users would like to enter $\frac{1}{a^2+3}$, they only have to enter “1/a2+3” (see step 1 of figure 1). Neither do they need to input a power sign (e.g. a caret symbol) nor parentheses for the delimiters. In other words, users do not need to enter symbols that are not printed. Next, users perform the conversion to the desired two-dimensional mathematical expressions on an element starting from the left. They have only to select interactively the desired elements and/or the operands from the candidates by using MathTOUCH (see step 2 of figure 1). Finally, after they fixed all the elements, they are able to obtain the mathematical expressions in the desired format. The available output formats are LaTeX, MathML, PNG, JPEG, EPS, Maxima, Maple, and Mathematica. Table 1 presents examples of mathematical expressions and corresponding linear strings that can be entered into MathTOUCH. This input method enables users to input almost all mathematical expressions without learning a new language or syntax.

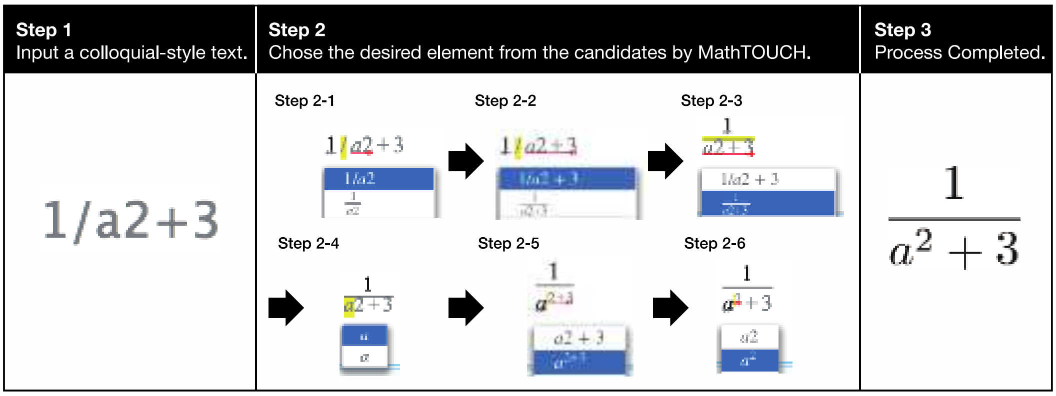


Figure 1. MathTOUCH input procedure.

Table 1. Example mathematical expressions using MathTOUCH.

Example	MathTOUCH
$5x^2+2$	5x2+2
$\sqrt{2}$	root2
$sin^2 x$	sin2x
$log_{10} x$	log10x
$e^{\pi x}$	epx
$\sum_{k=1}^{n} k^2$	sumk=1nk2
$\lim_{x \to 1} \frac{x}{2}$	limx-->1x/2
$\frac{df}{dx}$	df/dx
$\int_0^1 x(1-x)dx$	int01x(1-x)dx

2.2. Advantages of MathTOUCH

In our previous study, we conducted two experiments, i.e. a performance survey (Shirai, Nakamura and Fukui, 2015) and an eight-week learning experiment (Shirai and Fukui, 2014), to evaluate the efficacy of MathTOUCH. In this section, we summarise these studies.

The performance survey focused on the following research question: Are students able to input mathematical expressions using MathTOUCH more smoothly than with the standard interfaces found in current e-assessment systems? In this study, we carried out a mathematical entering test to compare MathTOUCH, a text-based interface, and a structure-based interface regarding three elements of usability: effectiveness, efficiency, and satisfaction. The number of participants was 108 including 54 high school students and 54 university students. The ratio of humanities courses to science courses and males to females among participants was almost the same. They were asked to enter mathematical expressions using one of the three interfaces assigned to them. The results of a Mann-Whitney test showed no significant difference for the task-performance rates among the three interfaces. Regarding the task-performance times, MathTOUCH enabled participants to enter mathematical expressions approximately 1.2 to 1.6 times faster than the standard interfaces. Moreover, our system was shown to have a high level of user satisfaction in regards to mathematics input usability.

The eight-week learning experiment focused on the following research question: are students able to practice mathematical work using MathTOUCH on STACK at the same learning rate as with the current interface on STACK? In this intersubject study, 84 students, who showed no significant difference regarding mathematical skill based on a previous paper test, practiced mathematical work on STACK using their assigned mathematical input interface for eight weeks. The results showed that students were able to practice using MathTOUCH at the same learning rate. Furthermore, the results of the questionnaire revealed a higher level of satisfaction regarding Memorability, which was significantly higher than with the current interface.

These above results indicate that MathTOUCH is more effective in terms of input performance than current standard interfaces. However, users needed to use a Java-compliant device in order to interact with MathTOUCH.

2.3. Reconstructed MathTOUCH

In this section, we describe how we reconstructed MathTOUCH such that the interface is not only available on Java-compliant devices but also on various other devices. We reconstructed MathTOUCH using JavaScript (HTML5) because JavaScript is compatible with web applications. Developers are able to incorporate MathTOUCH into their own web applications and only have to include information of the header and body for MathTOUCH. Figure 2 shows a screenshot of the interface of the new version of MathTOUCH implemented on STACK. We used MathJax to display the conversion candidates. Furthermore, we added an edit function and enhanced the support function. The procedure for entering expressions is the same as with MathTOUCH using Java.

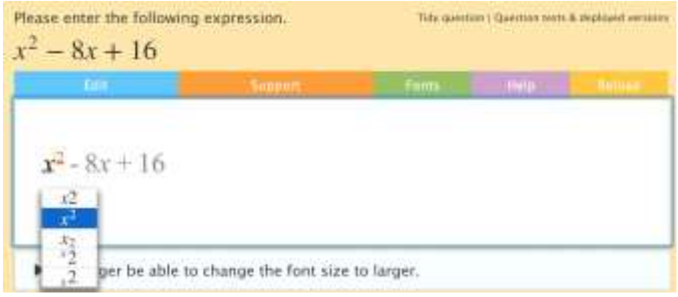


Figure 2. Reconstructed MathTOUCH on STACK.

3. Evaluation

3.1. Purpose and Procedure

Our prior work showed that the input performance of MathTOUCH exceeds that of current standard interfaces. We conducted a five-week learning experiment to test whether the RMT version offered stable input performance in comparison to the PMT version. The experiments are intended to investigate whether students are able to practice mathematical work using RMT on STACK at the same learning rate as PMT.

This experimental study involved 30 students as participants who were assigned to two groups. One group was assigned RMT (N=16), whereas the other group was assigned PMT (N=14). We conducted a pre-survey to investigate whether any differences in terms of typing skill and basic mathematical knowledge existed between the groups. The results showed that the difference between the groups was not significant. Every week for 5 weeks, they practiced mathematical work on STACK using their assigned mathematical input interface. Table 2 contains examples of the questions. The mathematical content involved basic calculations with square roots. They practiced 10 questions once a week.

3.2. Measures

We measured the solving times, the percentage of correct answers, and the learning rates. Moodle automatically measures the total amount of time spent on quizzes from the moment the button is clicked to start the quiz until the button is clicked to submit all the answers. We took the mean time for each group to be the solving time. After they completed all the mathematical work, we administered a questionnaire to determine students' subjective satisfaction with each interface. We used a 5-point rating scale from 1 (strongly disagree) to 5 (strongly agree). The contents of the questionnaire are shown in the first column of table 3.

Table 2. Mathematical example of mathematics drill questions.

Example of questions	Answer
Simplify the expressions. $\sqrt{20} \times 2\sqrt{2} \div \sqrt{5}$	$4\sqrt{2}$
Simplify the expressions. $\sqrt{35} \times \sqrt{5} - \frac{14}{\sqrt{7}}$	$3\sqrt{7}$

Table 3. Results of the questionnaire regarding subjective satisfaction.

Contents of the questionnaire		PMT	RMT
It was easy to master the use of this UI.	**Learnability**	3.94 (0.85)	3.86 (1.03)
Mathematical expressions could be inputted smoothly using this UI.	**Efficiency**	3.81 (0.91)	3.71 (0.99)
It was easy for me to correct miss-entered operations.	**Error**	3.63 (0.89)	3.14 (0.86)
Even after the second week, I remembered how to use this UI I was instructed on in the first week.	**Memorability**	4.31 (0.70)	3.79 (1.19)
Would you like to use this UI when you enter the mathematical expressions?	**Loyalty**	3.88 (0.62)	3.64 (1.34)

Numbers in parentheses denote SD.

3.3. Results

Figure 3 shows the result of the solving time and percentage of correct answers. We ran a Mann-Whitney test for each week. This analysis yielded no significant interaction between PMT and RMT for each week.

We also calculated the learning rate using the log-linear model by progressive average. We excluded the data for the first week because the students had been given instruction on how to use the system on that same day. The five-week coefficient of determination (R^2) for PMT and RMT is 0.910 and 0.801, respectively. On the other hand, when the first week is excluded, the four-week R^2 for PMT and RMT is 0.998 and 0.991, respectively. These results support the validity of use for the 4 weeks of data for evaluating learning rate. The results of the learning rate show that students are able to practice mathematical work using RMT with the same learning rate as with PMT with 92.4% and 89.9%, respectively.

Table 3 provides the results for the subjective satisfaction. The results of the Mann-Whitney test showed no significant difference for the solving times between PMT and RMT for each of the questions.

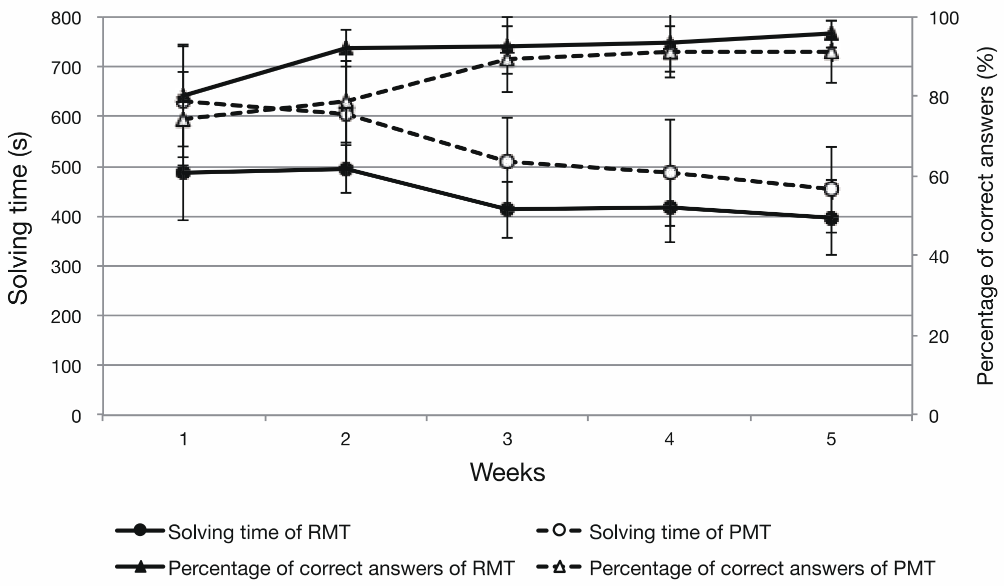


Figure 3. Progression of the solving time and percentage of correct answers.

4. Conclusions

This paper presents our reconstructed version of MathTOUCH using JavaScript to make MathTOUCH available not only on Java-compliant devices but also on various other devices. In this paper, we presented the results of the five-week learning experiment for testing the stability of reconstructed MathTOUCH. The results showed that students are able to study using reconstructed MathTOUCH on STACK as well as the previous version of MathTOUCH. We have made a web trial version of MathTOUCH available to everyone (Fukui, 2016).

The most important avenues for future research are to make the conversion prediction of MathTOUCH intelligent using machine learning to prevent students from having to convert each element individually. In 2015, we proposed a predictive algorithm for converting linear strings to an entire mathematical expression (Shirai and Fukui, 2016). In the future, we plan to implement this algorithm to enhance MathTOUCH. Furthermore, we additionally aim to develop a MathTOUCH interface for smart devices.

5. Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP26330413 and 16K16178.

6. References

Osaka Prefecture University, 2016. MATH ON WEB. Available at: http://www.las.osakafu-u.ac.jp/lecture/math/MathOnWeb/ [Accessed 8 October. 2016].

Pollanen, M., Wisniewski, T., Yu, X., 2007. Xpress: a novice interface for the real-time communication of mathematical expressions. *Proceedings of MathUI 2007*. Available at: http://euclid.trentu.ca/math/marco/papers/3_Pollanen-Xpress.pdf [Accessed 5 October 2016].

Smithies, S., Novins, K., Arvo, J., 2001. Equation Entry and Editing via Handwriting and Gesture Recognition. *Behaviour and Information Technology*, 20(1), pp.53-67.

Fukui, T., 2012. An Intelligent Method of Interactive User Interface for Digitalized Mathematical Expressions (in Japanese). *RIMS Kokyuroku*, 1780, pp.160-171.

Shirai, S. and Fukui, T., 2014. Improvement in the Input of Mathematical Formulae into STACK using Interactive Methodology (in Japanese). *Computer & Education*, 37, pp.85-90.

Shirai, S., Nakamura, Y., and Fukui, T., 2015. An Interactive Math Input Method for Computer Aided Assessment Systems in Mathematics (in Japanese)*. IPSJ Transactions on Computers and Education*, 1(3), pp.11-21.

Shirai, S. and Fukui, T., 2016. Evaluation of a Predictive Algorithm for Converting Linear Strings to Mathematical Formulae for an Input Method*. Mathematical Aspects of Computer and Information Sciences*, *Lecture Notes in Computer Science*, 9582, pp.421-425.

Fukui, T., 2016. MathTOUCH. Available at: http://math.mukogawa-u.ac.jp/en/ [Accessed 9 November 2016].

RESEARCH ARTICLE

A New Mathematics Input Interface with Flick Operation for Mobile Devices

Yasuyuki Nakamura, Graduate School of Information Science, Nagoya University, Nagoya, Japan. Email: nakamura@nagoya-u.jp
Takahiro Nakahara, Sangensha LLC., Chitose, Japan. Email: nakahara@3strings.co.jp

Abstract

Developing online test environments for e-learning on mobile devices will be useful for increasing drill practice opportunities. To provide a drill practice environment for calculus using an online mathematics test system, such as STACK, we developed a flickable mathematics input interface that can be easily used on mobile devices such as smartphones and tablets. The interface developed using JavaScript and MathDox is mainly for entering mathematical expressions. When the alphabet or number keys on the interface are touched, various candidates of operation appear around the touched key. Flicking in either the leftward, rightward, upward, or downward direction performs the required operation, depending on the selected key. The number of key taps required for entering mathematical expressions on a mobile device using the proposed mathematics input interface is compared with the number of key taps required in direct input; direct input involves using the built-in keyboard of a device. The number of key taps is considerably reduced when using the new mathematics input interface. Furthermore, our new mathematics input interface is compatible with traditional keyboards. The keyboard is automatically selected based on the types of devices being used.

Keywords: STACK, math input, mobile devices, flick operation.

1. Introduction

In recent years, learning management systems (LMSs) are being used for learning courses in many educational institutions. One of the most popular features of an LMS is online testing for determining students' understanding of a subject. Typical online tests include multiple-choice, true-or-false, and numerical input, but the mathematics input online test is gaining popularity in science education. In the mathematics input test, mathematical expressions are entered as answers and are automatically assessed, usually using a Computer Algebra System (CAS). WeBWork (Baron, 2010), Maple T.A. (Zivku, 2015), MATH ON WEB (Kawazoe et al., 2013), Numbas (Perfect, 2015), and STACK (Sangwin, 2013) are examples of online assessment systems that are used in educational institutions worldwide.

Online testing is useful for determining students' understanding of a subject; it offers the advantage of instant feedback through automatic assessment, and students can practise by solving many online test questions by themselves. Furthermore, if questions are designed such that they are automatically generated with random variables, students can repeatedly practise different questions, thereby improving their understanding of a subject.

Online drill testing can be delivered not only using PCs but also through mobile devices such as smartphones and tablets to help students practise anytime and anywhere. However, the problem of mathematics input complexity arises for questions requiring the entry of mathematical expressions as answers rather than multiple-selection or number input types of answers. For example, when students answer x^2+5x+6 to the question asking for the expansion of the expression$(x+2)(x+3)$, they have to enter the expression x^2+5*x+6 in the answer space. However, when users enter the expression in which numbers and symbols are combined using smartphones, it is necessary to switch the smartphone keyboard screen many times; this requires 19 key touches.

In fact, the difficulties in entering mathematical expressions are not limited to smartphones; there are difficulties when using a tablet and a PC as well. There are some approaches to overcome these difficulties, which are discussed in the next section. In this paper, we introduce a new type of mathematics input interface with flick operations for mobile devices. This interface enables students to enter mathematical expressions easily; it also gives them more opportunities to practise through online testing in e-learning systems using mobile devices.

This paper is organised as follows: we analyse some mathematics input interfaces and identify problems with them in section 2. The flickable type of mathematics input interface is introduced in section 3, and its mathematics input efficiency is discussed briefly in this section. We summarise the paper in section 4.

2. Examples of mathematics input interfaces

As described above, to reduce the difficulties in mathematics input, several interfaces have been proposed. For example, Maple T.A. features an 'Equation Editor', and mathematical expressions are displayed in a 'two-dimensional' manner (for example, $x^2 + \frac{x+1}{2}$). The equation editor increases the efficiency of recognising mathematical expressions, particularly indices and fractions, and supports their input in smartphones and tablets. When users enter mathematical expressions through these devices, switching between letters and numbers/symbols is required, and the equation editor does not reduce the complexity of the mathematics input process.

To increase the input efficiency of STACK, MathTOUCH (Shirai et al., 2014) and interfaces utilising MathDox (Nakamura et al., 2014) were proposed. However, it is assumed that they are used mainly on PCs. MathTOUCH runs as a Java plug-in, and it is not supported by some mobile operating systems (OSs) such as iOS. MathDox was developed using JavaScript, but it is not supported by mobile devices because it is not designed for touch operations.

To reduce the difficulties in entering mathematical expressions using mobile devices, we propose the use of a flick input interface that is often used in mobile devices, particularly by Japanese people. We used STACK and developed a mathematics input interface with a flick operation, which is expected to facilitate drill practice through online testing on mobile devices. We will introduce this new type of mathematics input interface in the next section.

3. Mathematics input interface with flick operation

We decided to use JavaScript to minimise the dependency on mobile device OSs. We had already developed a conversion filter from MathDox to Maxima, and we used MathDox for describing entered mathematical expressions, which was another reason to adopt JavaScript for developing the new interface. This section provides an overview of the interface and explains the process for entering mathematical expressions with a simple example. To determine the efficiency of mathematics input using the new interface, we compared the number of key touches in the new interface with the number of key touches required with a conventional keyboard. Furthermore, in the last subsection, we discuss implementing the flick operation in traditional keyboards and the automatic selection of keyboards, depending on the devices that students use.

3.1. General specifications

Figure 1 shows the basic layout that is displayed when the interface is activated. The user can input numbers using the '123' key and alphabets or Greek letters using the 'xy' key in the left column. When the user taps the 'fx' key, options such as exponential and trigonometric functions become available. Keys for basic operations keys are in the right column.

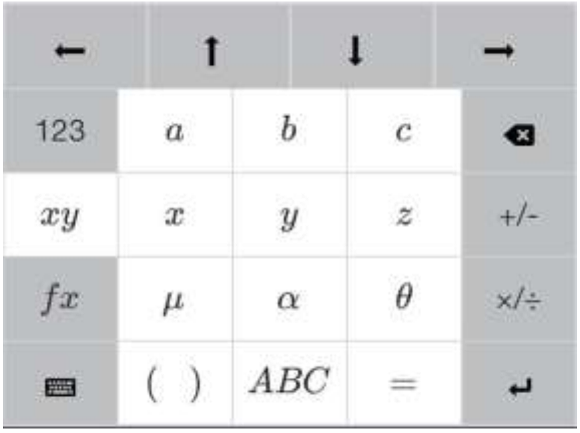


Figure 1. Basic layout of the flickable mathematics input interface that is displayed when the interface is activated.

3.2. Entering mathematical expressions

Figure 2 shows an example wherein the expression $x^2 + 5x + 6$ is entered as the answer for the expansion of the expression$(x + 2)(x + 3)$. An upward flick in the direction of the 'x' key (figure 2, upper left) causes the index input state to appear (figure 2, upper middle); the user can tap the '2' key to enter the index (figure 2, upper right). Then, by tapping the '+/-' key and flicking in the upward direction (figure 2, lower left), the '+' operator is entered. To enter$5x$, the user simply taps the '5' key and flicks in the leftward direction (figure 2, lower middle). After entering '+', 6 is entered by tapping the '6' key (figure 2, lower right). As seen in figure 2, the product of a number and x or y that often appears in mathematical expressions is built into the interface; this results in a reduction in the number of key touches.

3.3. Estimation of input efficiency

Table 1 provides a comparison of the number of tap operations required in direct input using a traditional keyboard and that required in flick input using the interface keyboard. Note that the direct input starts from the alphabet keyboard, and the flick input starts from the state depicted in figure 1. In addition, the process of entering numbers by leaving the alphabet keyboard and holding down the number switching key is not adopted. As seen in table 1, the number of key touches is obviously reduced, leading to a reduction in the number of steps required for mathematics input. It is remarkable that fewer key touches are required for the input of functions, particularly when entering trigonometric functions.

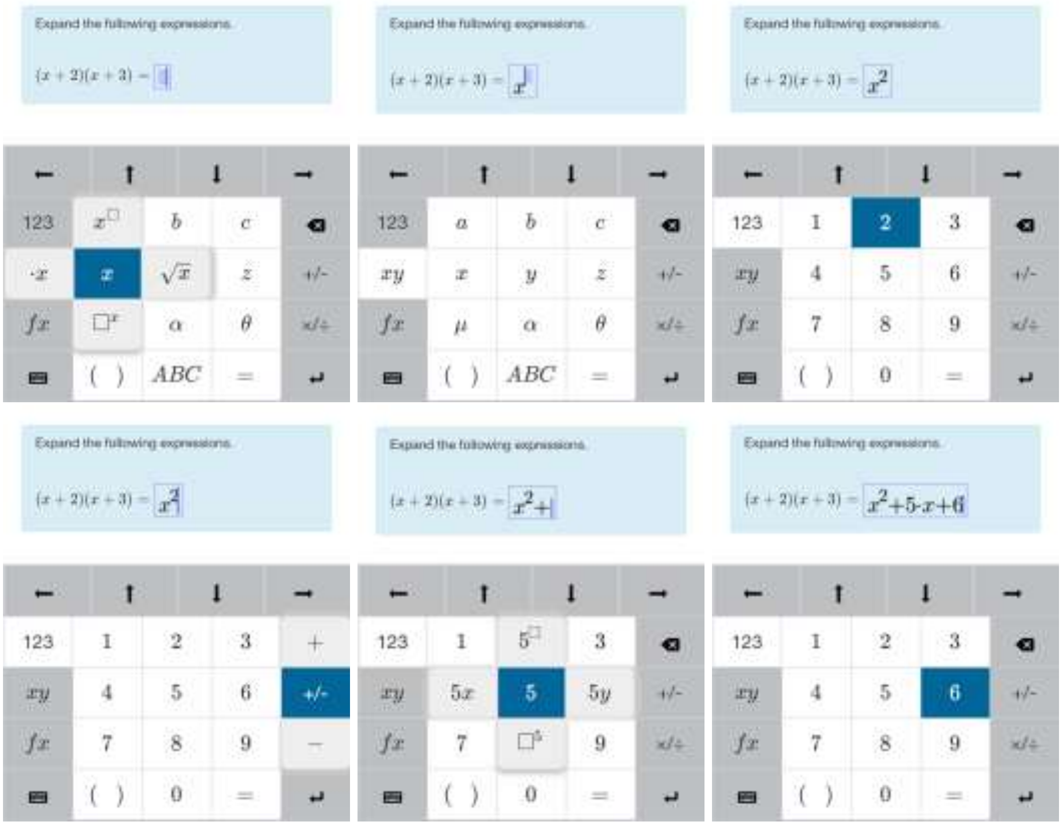


Figure 1. Procedure for entering mathematical expression $x^2 + 5x + 6$. The figure should be read from left to right and from top to bottom.

Table 1. Comparison of the number of tap operations in direct input using a traditional keyboard and in flick input using the interface keyboard.

Mathematical expressions	Number of tap operations	
	Direct input	Flick input
$x^2 + 5x + 6$	19	8
$3x^2 - \frac{2x}{(x^2+1)^2}$	36	13
$2xcosx^2$	23	7

We conducted a usability test at a university in Nagoya, Japan. We asked 29 students to enter the following five mathematical expressions by using both a traditional keyboard and the new flick input keyboard.

$$(a+b)^2, \frac{y}{x+1}, \frac{1}{\sqrt{x^2+y^2}}, sin(\pi x) + e^x, \frac{2xsinx}{(cosx+1)^2}$$

After they entered these mathematical expressions, we conducted a survey on usability and satisfaction levels. The survey questions were based on the following five parameters: learnability, efficiency, difficulty or ease in making corrections, rememberability, and the intent to reuse. These parameters are originally from Jakob Nielsen's five goals of usability (Nielsen, 1993). Table 2 lists the questions and the results of the survey.

The ratings are averages based on the responses of 29 students with the maximum rating being 5.0. The higher the ratings, higher are the usability and satisfaction levels. We can conclude that the usability and satisfaction levels are higher when the flick input method is used to enter mathematical expressions rather than the direct input method.

Table 2. Summary of the survey on usability and satisfaction levels based on Nielsen's five goals of usability.

Questions	Direct input	Flick input
It is easy to learn how to input math	3.0 (1.2)	3.5 (1.0)
I can input math quickly and easily	2.6 (1.1)	3.2 (1.3)
It is not confusing and easy to correct	3.0 (1.2)	3.1 (1.1)
I remember the method that I learnt at the rehearsal	3.0 (1.1)	3.1 (1.1)
I will use this method to input math the next time	2.9 (1.3)	3.2 (1.4)

Numbers in parentheses denote SD.

3.4. Implementation of flick operation to traditional keyboards

As can be seen, users cannot enter many letters such as d and t when they use the flickable mathematics input interface described earlier in this section. The alphabets and symbols available in our mathematics input interface cover most of the letters and symbols used in the introductory mathematics test. However, the letter t or v cannot be entered; these letters are often used as variables denoting time or velocity in physics. Therefore, we implemented the flick operation using a traditional keyboard. When the icon on the lower-left side of the keyboard displayed in figure 1 is selected, the keyboard changes to a full/traditional keyboard. Figure 3 shows an example where a mathematical expression is entered from a full keyboard using the flick operation. In this example, we consider a physics test in which the letters v and g are used to denote velocity and the gravitational acceleration constant, respectively. In a full keyboard, candidates of operation appear horizontally over the tapped key (figure 3, centre).

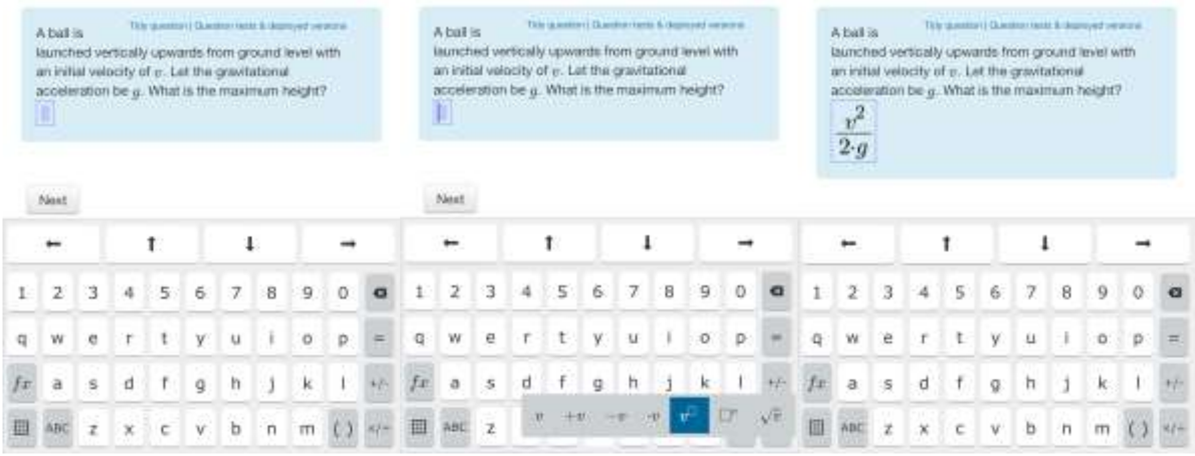


Figure 3. Procedure for entering mathematical expression $\frac{v^2}{2g}$ from full keyboard using flick operation.

3.5. Automatic selection of keyboard based on types of devices

We have demonstrated the working of our new mathematics input interface. We believe that the numerical keyboard (figure 3) is suitable for smartphones, and the traditional keyboard is suitable for tablets; the MathDox input process is suitable for PCs, depending on the screen size and usability of the physical keyboard. Therefore, by default, the numerical keyboard and traditional keyboard appear in smartphones and tablets, respectively. It is possible to switch between these keyboards.

4. Conclusion

For students taking online mathematics tests, online test environments for e-learning in mobile devices are considered to be useful for increasing drill practice opportunities. Therefore, we developed a mathematics input interface with a flick operation, using STACK, for taking online mathematics tests. The demonstrations using the interface in the preceding sections confirmed that the number of key touches is reduced; the usability survey indicated a positive response from students who used the interface. Furthermore, we implemented the flick operation in traditional keyboards and solved the problem regarding the non-availability of alphabets in numerical keyboards.

5. Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 26282033.

6. References

Baron, L. M., 2010. Helping Teachers Generate Better Homework: MAA Makes Time for WeBWorK, *MAA Focus*, 30(5), pp.18-19. Available at: http://webwork.maa.org/pdfs/webwork_focus_Oct2010.pdf [Accessed 10 October 2016].

Kawazoe, M., Takahashi, T. and Yoshitomi, K., 2013. Web-based system for after-class learning in college mathematics via computer algebra system, *Proceedings of the 6th East Asia Regional Conference on Mathematics Education.*

Nakamura, Y., Nakahara, T. and Inagaki, Y., 2014. Math Input Interface by MathDox Formula Editor for Math e-Learning System STACK, *Abstract in Proceedings of the 19th Asian Technology Conference in Mathematics*.

Nielsen, J., 1993. *Usability Engineering*. Morgan Kaufmann.

Perfect, C., 2015. A demonstration of Numbas, an assessment system for mathematical disciplines, *CAA Conference 2015*. Available at: http://www.numbas.org.uk/wp-content/uploads/2015/07/caa2015_numbas.pdf [Accessed 10 October 2016].

Sangwin, C., 2013. *Computer Aided Assessment of Mathematics*, Oxford University Press.

Shirai, S. and Fukui, T., 2014. Development and Evaluation of a Web-Based Drill System to Master Basic Math Formulae Using a New Interactive Math Input Method, Proceedings of the 4th International Congress on Mathematical Software, Seoul, Korea, *Lecture Notes in Computer Science 8592*, Springer, pp.621-628.

Zivku, J., 2015. Maple T.A.: Today and the Future, The 2015 Maple T.A. User Summit. Available at: https://webstore.maplesoft.com/taconference/PDFs/2015Pres/TATodayFuture-JonnyZivku.pdf [Accessed 9 October 2016].