Using STACK to support student learning at masters level: a case study

Journal Item

Teaching Mathematics and its Applications (2019) Page 1 of 11
doi:10.1093/teamat/drnxxx

Using STACK to support student learning at masters level: a case study

T. W. LOWE†, AND B.D. MESTEL‡
SCHOOL OF MATHEMATICS AND STATISTICS
THE OPEN UNIVERSITY, MILTON KEYNES, MK7 6AA, UK

[Received on 19 November 2018]

The development of six online quizzes to support students' study of an introductory mathematics masters module at The Open University is described and their use evaluated. The quizzes were implemented using the *STACK* online e-assessment system which is powered by a computer-algebra engine. Evaluation of student feedback and an initial quantitative study of the effect of engaging with the quizzes on the final examinations marks suggest that further development of e-assessment at mathematics masters level is warranted.

Keywords: *STACK*, e-assessment, mathematics, masters programme.

1. Introduction

The use of e-assessment within higher education courses in the mathematical sciences has seen significant growth in the last 20 years, and the United Kingdom (UK) has been at the forefront of innovations in this field. There have been several e-assessment systems pioneered in the UK, among them *Numbas* at the University of Newcastle upon Tyne (Numbas.org.uk, 2018), *Dewis* at the University of Western England (Dewisprod.uwe.ac.uk, 2018), and *Maths E.G.* at Brunel University (Mathcentre.ac.uk, 2018). Sangwin (2013) provides a general overview of the use of e-assessment in higher education mathematics teaching.

One of the most versatile of these e-assessment systems is *STACK* (Stack.ed.ac.uk, 2018), developed first at the University of Birmingham and now hosted by the University of Edinburgh. *STACK* is built on the open-source *Maxima* computer-algebra system (Maxima.sourceforge.net, 2018), which facilitates both algebraic randomisation of questions and mathematical judging of student answers hence enabling immediate feedback to be given, tailored to student input. *STACK* is fully integrated into the *Moodle* Virtual Learning Environment (Moodle.org, 2018a) and so it is possible to take advantage of both *STACK*'s own sophisticated input types and those of other question types available within *Moodle* in a single quiz. At the time of writing, *STACK* has been installed on over 850 sites (Moodle.org, 2018b) since its redevelopment as a native *Moodle* question-type in 2012, with users worldwide including in Germany, Norway, Finland, Japan and Spain as well as in the UK. It is used at universities as diverse as the University of Edinburgh, Loughborough University and the Open University in the UK, Tokyo University of Science, Yamaguchi Prefectural University, and Tokyo Kasei University in Japan, and by the *ABACUS* consortium in Finland, which consists of 18 universities led by Aalto University (Abacus.aalto.fi, 2018). It has also been incorporated into the *ILIAS* learning management system, used principally in Germany (Ilias.de, 2018). The contribution of *STACK* to UK higher education has been recently recognised by the UK's Higher Education Academy (Heacademy.ac.uk, 2018).

†Corresponding author. Email: Tim.Lowe@open.ac.uk
‡Email: Ben.Mestel@open.ac.uk

Since its foundation in the late 1960s as a 'university of the air', the Open University (OU), the United Kingdom's premier distance-learning university, has been in the forefront of innovation in teaching and assessment. Within assessment, the OU has been a pioneer in the use of multiple-choice objective testing, has developed its own bespoke Java based e-assessment system *OpenMark* (Open.ac.uk, 2018), and has led the development and maintenance of the *Moodle quiz engine* (Hunt, 2012). Having fully integrated *STACK* into the University's *Moodle*-based Virtual Learning Environment (VLE), the OU currently serves over a million *STACK* questions annually, the vast majority as part of the undergraduate curriculum. For distance learning students, online assessment provides an ideal opportunity to practise mathematical methods and techniques whilst receiving immediate feedback on their answers. This is acknowledged to be essential if the feedback is to aid their learning (Gibbs & Simpson, 2004).

The use of *STACK*, and of e-assessment generally, at taught postgraduate level, and even level 3 undergraduate study, is not so well established. One module of the Mathematics MSc programme, M823 *Analytic Number Theory I*, has implemented some *STACK* questions, which appear to have been well received, but their reception by students and their effect on student performance have not been rigorously studied hitherto.

Moreover, it is often thought, anecdotally at least, that e-assessment is more fitted to lower-level mathematics assessment, and in particular to mathematical methods and calculus. Additionally, the high initial cost of developing resources has weighed against e-assessment at MSc level, where there are smaller numbers of students. It was therefore of particular interest to develop e-assessment at MSc level and to study both student reaction and, subject to the usual methodological problems, to study their effectiveness.

The organisation of this paper is as follows. In Section 2 the Open University's MSc Programme is discussed in outline, and the *STACK* quizzes developed are reviewed in Section 3. Results from a statistical linear model of the effect of interaction with the *STACK* quizzes on the examination are presented in Section 4 and the results from a survey of student feedback are given in Section 5. A brief conclusion is presented in Section 6.

2. The Open University Mathematics MSc programme

The Open University's Masters Programme in Mathematics is the largest such programme in the United Kingdom. It commenced in 1986 and currently has a total student population of approximately 500 students studying part-time, with c. 50 students graduating *per annum*. Students study six 30-credit modules, starting with either M820 *Calculus of Variations and Advanced Calculus* or M823 *Analytic Number Theory I*, and culminating in a project module. The programme enables students to steer a path focussed either on pure mathematics or applied mathematics and provides a Postgraduate Certificate (60 credits) and a Postgraduate Diploma (120 credits) in addition to the full Master of Science (180 credits).

The principal entry module into the OU's MSc programme is M820 *Calculus of Variations and Advanced Calculus*, an applied mathematics 30-credit module focussed on developing calculus techniques through the extensive study of the classical calculus of variations.[1] With 100 – 150 students registering annually, its success is important for the health of the whole MSc programme since retention in the module has a knock-on effect for the whole programme.

As would be expected for an OU module, the students enrolled on M820 are rather diverse. As an illustration to set the context of the module, figures for the 2017–2018 cohort were: 91 male: 41 female,

[1]The module provides a full set of notes, in the style of Gel'fand and Fomin's famous text (Gel'fand, Fomin & Silverman, 2000).

22.6% of students with a declared disability, and a unimodal age profile ranging between 22 and 65+ with a mode of 30–39 years. Geographically, most students were based in the UK, but 21.2% were domiciled outside of the UK, principally in European Union countries. The ethnic composition of the 130 students self-identifying was 110 White, 4 Black, 12 Asian, 1 Mixed and 3 Other.

In common with other modules of the Mathematics MSc programme, M820 has a compulsory continual assessment component. To pass this component, students must obtain at least 30% in three out of the four Tutor Marked assignments (TMAs) in the module. Students receive extensive written feedback on their assignments, a hallmark of the Open University's so-called *open learning* methodology. The TMAs are not subsequently used in the formal assessment of the module, except in borderline cases.

In addition to the continual assessment component, students must also pass the module examination. This is a three-hour paper with marks from the best four out of six equally weighted questions counting towards the total examination mark. In common with other OU modules, students are permitted to bring into the examination an annotated copy of the Module Handbook (which gives a summary of the principal results as well as useful reference material).

Teaching materials consist of extensive notes, of more than 600 A4 pages including a large number of exercises and their solutions, the Module Handbook, five synchronous online tutorials to assist with the TMAs and with revision for the examination, and sundry recorded lectures and short screencasts. In addition to an assigned tutor, students may, at extra cost, attend two face-to-face weekend residential events, one in the middle of the module and the other, a revision event, towards the end of the module. In addition, for the 2017–2018 cohort, six *STACK* e-assessment quizzes were introduced to assist with student learning and revision. These six *STACK* quizzes are the focus of this present paper.

As is the case for other universities, student retention can be a major challenge for the Open University. In common with other distance-learning universities, the OU has a diverse student population, which includes *inter alia* students with full-time jobs and family or caring responsibilities. The OU has a greater percentage of students with additional needs than most other UK institutions and also supports students in a variety of secure environments. These can present significant further problems for educators. The University has recently established a special programme to finance innovative projects to assist with student retention.

3. STACK quizzes in M820

Supported by the OU's special programme to assist with student retention, six extended exercises in the Calculus of Variations were developed using *STACK* between December 2017 and January 2018.[2] The topics chosen spanned a wide range of the M820 curriculum: Q1: Solving the Euler-Lagrange equation; Q2: the use of the first-integral equation to solve for the stationary paths; Q3: changing the dependent variables; Q4: Noether's theorem on the invariants arising from symmetries of the underlying functional; Q5: the use of the Jacobi equation to classify stationary paths; and Q6: the use of the Rayleigh–Ritz method to approximate the least eigenvalue of a Sturm–Liouville system. Apart from the final exercise, each topic was developed as a multipart quiz, building up to the solution via a sequence of steps, each of which was a *STACK* question in its own right. Students were given three opportunities to give a correct answer to each step following which a worked solution was given. The question encapsulating the next step in the solution method was then made available. This staging of questions was facilitated by the important ability to link questions with the same values of randomised variables within *Moodle*, and in particular, *STACK*.

[2] The exercises were called 'quizzes' in the nomenclature of the Moodle-based Virtual Learning Environment at the OU.

The design, implementation and checking of *STACK* e-assessment questions is in general a non-trivial task. The interactive approach taken and some of the design problems encountered are discussed in (Erskine & Mestel, 2018), where the Jacobi-equation quiz design is discussed in some detail.

Following their design and implementation, the quizzes were made available to students in February 2018. The students were recommended to use them as aids to learning and revision for the June 2018 examination. Student opinion on the six questions was tested via an online survey, the results of which are discussed in Section 5.

4. Analysis of the examination results

For the 2017–2018 student cohort, anonymised data on student performance in the online quizzes, the tutor marked assignments and the examination were analysed using the `lm` linear model library of the *R* statistical package (R Core Team, 2018). In order to elicit the effect of the *STACK* quizzes on student performance in the end-of-module examination, a linear model was fitted to the data through least-squares regression, as detailed below.

Of the 132 students who were registered on M820 in October 2017, 93 (70.5%) sat the examination in June 2018. Of the 93 students taking the examination, 79 (84.9%) interacted with the *STACK* quizzes prior to the examination date. In this study we count students as having interacted if they engaged with quizzes sufficiently to have obtained a non-zero score on at least one of the six quizzes. The total number of students who are recorded as having accessed the quizzes is 85 (91.4%). The large proportion of students engaging with the quizzes might seem unexpected since it is widely thought that students only engage with additional activities if the involvement is seen by the student to directly contribute towards the assessment of the module (Gibbs, 1999). Perhaps this statistic alone indicates students' appreciation that the quizzes would be of benefit to their exam preparation.

It should be noted that the research focussed entirely on the recorded mark in the examination before moderation by the examination board. No mitigating factors were taken into consideration and in one or two instances the final outcomes for the students concerned may have differed from that indicated by the examination mark. However, no large-scale moderation of the examination marks were made by the examination board, so the examination mark is a reliable gauge of student performance.

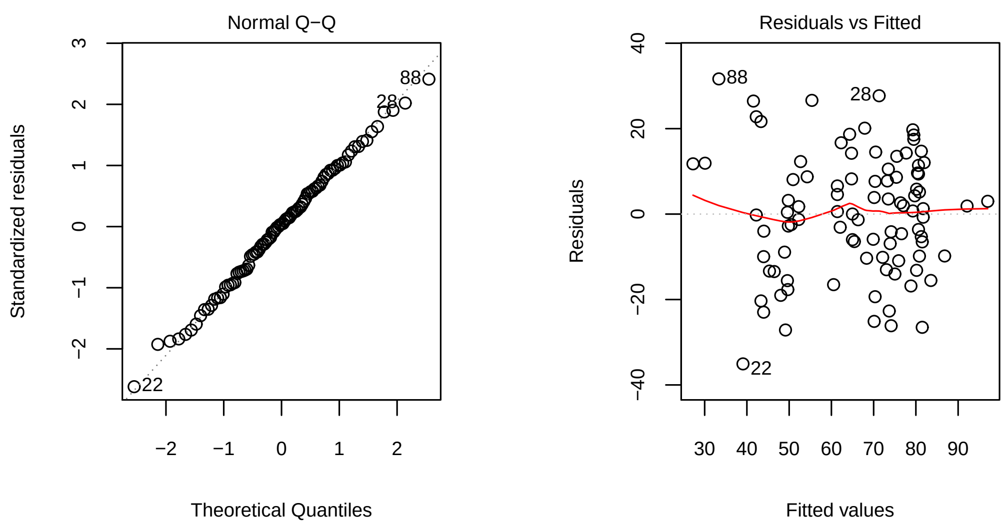
A linear model was fitted to the examination marks, relating them to quiz participation and TMA marks. This can be expressed as

$$E = a + b\delta_Q + cT(1-\delta_Q) + dT\delta_Q \tag{4.1}$$

where E is the total examination mark (on a scale 0 – 100), δ_Q is a binary variable taking the value 1 if the student interacted with the online quizzes and 0 if they did not, T is the mean of the three (out of four) highest scoring TMAs (on a scale of 0 – 100), and a, b, c and d are parameters of the model which are determined by fitting to the available data. The estimated best-fit values of these parameters together with their standard errors are given in Table 1. The residual standard error was 14.34 on 89 degrees of freedom with an adjusted R^2 of 0.5251. This indicates the fit is acceptable for higher-education research. The Normal Q–Q plot on the left in Figure 1 shows reasonable agreement in the central region, with some departure from normality in the tails. This must be expected since the examination marks are confined to the range 0–100. The standardized residuals vs. fitted values plot is shown on the right-hand side of Figure 1. There is clear clustering on the right resulting from a wide variation in examination performance amongst students with high TMA marks.

TABLE 1. *The estimated model parameter values, and their standard error.*

Parameter	Estimate	Standard error
a	-24.99321	14.30076
b	40.19479	15.75700
c	1.22410	0.19776
d	0.66974	0.08363



(a) (b)

FIG. 1. Diagnostic plots for the linear model (4.1). On the left is the Normal Q-Q plot showing reasonable agreement with normality except at the tails. On the right is shown the standardized residuals vs. fitted values.

The coefficients of the fit reveal useful information on the effect of interaction with the quizzes. Dividing the students into two groups, those who interacted with the online quizzes and those who did not, then to two decimal places the regression model may be written as $E = 15.20 + 0.67T$, for those who interacted, and $E = -24.99 + 1.22T$, for those who did not interact. Both coefficients of T are positive, indicating a positive correlation between TMA and examination marks, as expected.

Figure 2 shows a scatter plot of the examination marks E against T, the mean of the best three TMA marks, together with the two regression lines. The points (shown as circles) and the line corresponding to interaction with the online quizzes are coloured blue, while the points (shown as triangles) and the line corresponding to non-interaction are coloured red. The regression lines meet at a point with (T, E) = (72.5%, 63.8%) which, coincidentally, is not far from 70%, the threshold for a merit grade. For $T <$ 72.5%, the blue regression line gives a higher examination mark E, for a given continuous assessment mark T, than does the red line. So there is a higher predicted examination mark for those who interacted with the online quizzes than for those who did not. Conversely, for $T > 72.5\%$ the opposite holds, and

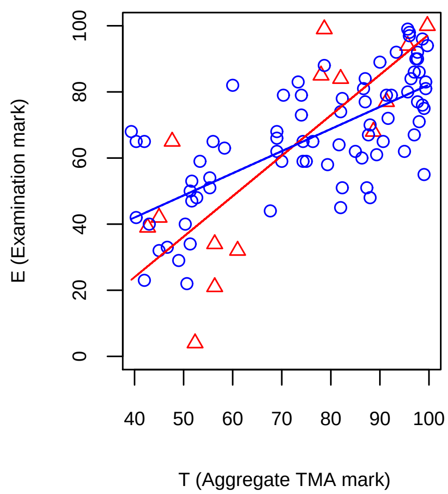


FIG. 2. Scatter plot of examination mark E against continuous assessment average T together with two regression lines. The circular points and the line in blue correspond to students who engaged with the online quizzes; the triangular points and the line in red to students who did not engage. Note the cluster of points corresponding to very high TMA marks with widely differing examination marks.

there is a higher predicted examination mark (for a given TMA performance) for those students who did not interact with the online quizzes than for those who did.

It must be stressed that it would be wrong to infer from this result that the interaction with the quizzes is 'good for weaker students and bad for stronger students', since no causality is implied by the model. Indeed, there are likely to be several confounding factors at play here. For example, a probable reason for the phenomenon discussed in the previous paragraph is evident from Figure 2. As might be expected, those students not engaging with the online quizzes (the red triangles) are clearly divided into weakly and strongly performing students, with few students in the central region. Conversely, the distribution of students who did interact with the online quizzes is more even, although the bimodality of the distribution is also evident, although much less pronounced. It is not unreasonable to speculate that more weakly performing students who did not interact were falling behind with their studies and had little time to engage with additional materials, while the more strongly performing students who did not engage were on top of their studies and perceived little need for further learning opportunities. This observation alone would be sufficient to account for the observed crossover in regression lines. Moreover, there is a significant group of students who did interact and who also achieved very high marks in the TMAs, but with highly variable examination marks, as can be seen from the cluster of blue circles on the right-hand side of Figure 2. This feature (partially an effect of the 0–100 range for TMA marks) is sufficient to lower the regression line for the students who did interact with the quizzes.

We can conclude, however, that for students below the crossover point (corresponding roughly to students scoring below a merit grade in the continuous assessment) interaction with online quizzes leads to an improved predicted examination mark. That the students themselves found positive benefit in their interaction with the online quizzes will be clear from Section 5.

5. Analysis of student feedback

Students registered on M820 were given the opportunity to complete an online survey giving feedback on the quizzes developed. This survey was completed by 28 students (35% of those interacting with the quizzes). Of these, 24 students reported using all 6 quizzes, 1 student used the first 5 quizzes and 2 students reported using the first two quizzes only. One student did not report having used any quizzes, but nonetheless gave positive feedback on them, which has been included in the analysis below.

As well as being given the opportunity to give open comments, the students were asked to rate the usefulness and their enjoyment of the online quizzes together with how confident they felt about the topics covered after having used the quizzes. These responses were recorded on a five-point Likert scale for each individual quiz and also for an overall impression. The results are summarised in Figure 3. The figure shows a diverging stacked bar chart [following (Robbins & Heiberger, 2011)] with neutral responses centred on zero, positive responses on the right, and negative responses on the left.

It is clear that students generally found the quiz both useful and enjoyable. This is reinforced by open comments given, where the most common view expressed was how useful the quizzes were for learning the material, revision and practice. The detailed solutions provided were also praised. As one student stated:

> The quiz has been very helpful in strengthening key skills, especially since it is scaffolded and the solutions are clearly explained.

The randomised nature of the questions was also appreciated. This enables students to repeat the quizzes and be given different instances of the questions each time. One student commented how this randomisation helped him abstract the general method for solving the problem from the specific details of an individual question:

> The good thing about these quizzes is the possibility to take them more than once, with slightly different variable values each time. I think that answering them repeatably allows to get a view on a broader picture behind a problem, instead of focusing on the calculation techniques, which become more obvious then.

The most common criticisms related to the inability to return to a previous sub-question within a quiz, which was a constraint resulting from the method of implementation.[3] Three students expressed some frustration with the need to enter mathematics using a typed linear syntax.

After taking the quizzes, students generally felt confident in the topics covered, although in the main this was rated “confident” rather than “very confident”. This is consistent with the hypothesis drawn from the quantitative data that the quizzes lead to higher predicted examination marks for the weaker students.

When asked how they use the quizzes, 28 students stated it was to revise for the exam, and 13 to learn the subject material more generally. (There is some overlap in these groups.) Additional online quizzes were requested by 24 students, and when asked on what topics these should be, the most common answer was all (or any) topics covered in the module.

[3] This constraint, which was a design feature, was chosen so that students were encouraged to tackle the quizzes in a coherent manner, although a summary of the results of earlier parts was given at the start of each quiz part so it was not necessary to revisit earlier parts. The constraint will be reviewed in the light of student feedback.

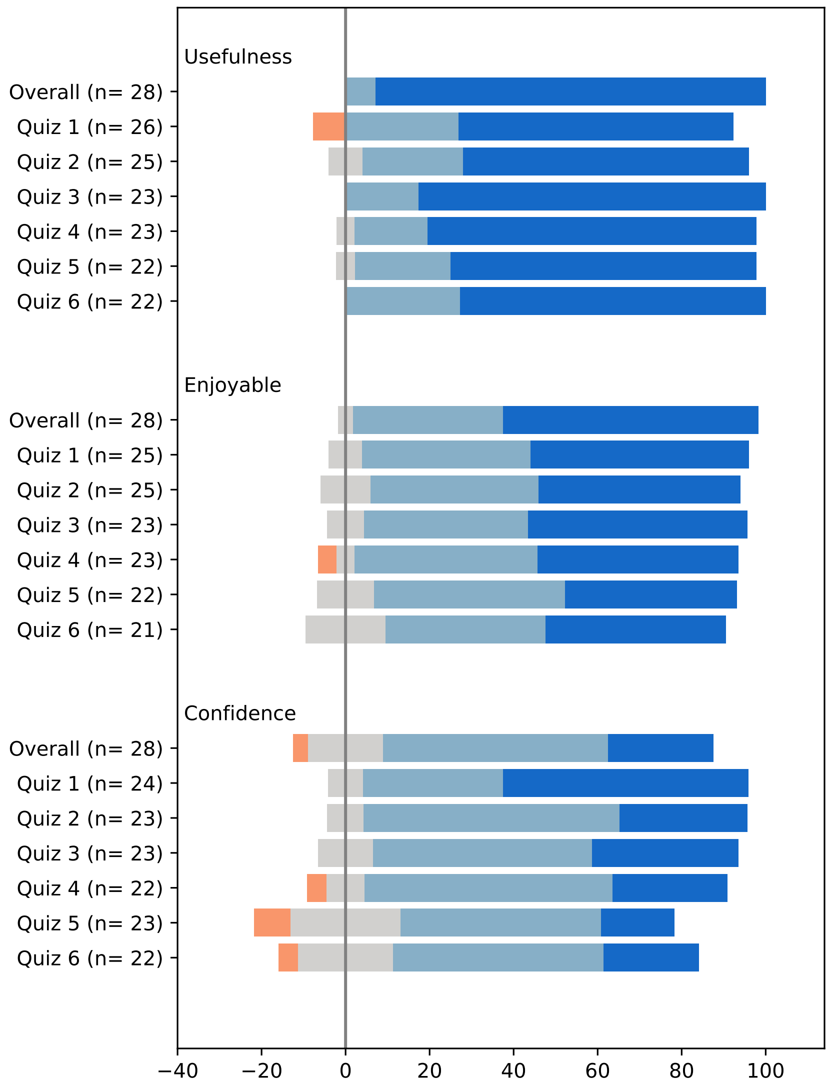


FIG. 3. Responses on a five-point Likert scale for the usefulness and enjoyment of the online quizzes, together with students' confidence in the topics after taking the quiz. Those responses to the right of the 0-line are positive while those to the left are negative. Note that no very negative opinions were given.

6. Conclusion

While there has been significant growth in the use of e-assessment in HE mathematics education, most use has been confined to (lower-level) undergraduate teaching and there has been relatively little implementation at higher levels. In this article, we have reported on one such implementation at MSc level. Whilst still under development, the initial student feedback has indicated these quizzes have the potential to improve student performance in the the final examination.

The result of the quantitative studies detailed in Section 4 are necessarily tentative, not least because of the small sample size, the absence of any (non-self-selected) control group, the clustering of continuous assessment marks at the higher end of the scale, and the only partial mapping of the topics covered in the online quizzes with the examination questions (a problem that will be alleviated in time through the provision of more online quizzes). There is also a need for a more refined definition of what ‘interaction’ with the quizzes actually entails.

While it would be incorrect to claim any direct causality, the output of the statistical modelling suggests that, for students performing below the merit grade, as indicated by their TMA marks, interaction with the online quizzes leads to a higher predicted examination mark. Since the motivation for the development of the quizzes was to assist with student retention, this result (however tentative) is positive.

It seems that on the basis of this, albeit limited, initial study, further research & development of e-assessment with *STACK* for MSc mathematics teaching is warranted. Let us give the final word to one of the more enthusiastic M820 students: ‘These quizzes made a big difference to me coming to grips with the material. They were extremely useful! Thank you!’

Acknowledgements

The authors would like to thank Grahame Erskine for his help in developing the *STACK* questions in the Open University’s Virtual Learning Environment, and Paul Garthwaite, Chris Jones and Heather Whitaker for help in the statistical modelling of the examination results. The question development was financed by a grant from The Open University’s *More Students Qualifying* programme.

References

ABACUS.AALTO.FI. (2018) *ABACUS*. `https://abacus.aalto.fi` [Accessed 31 Aug. 2018].

BUTCHER, P. (2008) Online assessment at the Open University using open source software: Moodle, OpenMark and more. *Proceedings of the 12th CAA International Computer Assisted Assessment Conference, 8th and 9th July 2008 at Loughborough University* (Khandia ed). Loughborough: Lougborough University, pp. 65-78. `https://research.moodle.net/149/1/Butcher_P_final_formatted_n1_2.pdf` [Accessed 12 Jun. 2018].

DEWISPROD.UWE.AC.UK. (2018) `http://dewisprod.uwe.ac.uk/` [Accessed 6 Jun. 2018].

ERSKINE, G. & MESTEL, B. D. (2018) Developing STACK practice questions for the Mathematics Masters Programme at the OU. MSOR Connections, **17**, 37–45.

GEL’FAND, I. M., FOMIN, S. V. AND SILVERMAN R. A. (2000) *Calculus of variations*. Mineola, N.Y.: Dover Publications.

GIBBS, G (1999) Using assessment strategically to change the way students learn. *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches* (S. Brown & A. Glesner eds). Buckingham: Society for Research into Higher Education and Open University Press, pp. 41-53.

GIBBS, G. AND SIMPSON, C. (2004) Conditions under which assessment supports students’ learning. *Learn. Teach. High. Educ.*, 1, 3-31.

HEACADEMY.AC.UK. (2018) *The Open University - STACK — Higher Education Academy. Collaborative award for teaching excellence*. `https://www.heacademy.ac.uk/person/open-university-stack` [Accessed 6 Jun. 2018].

HUNT, T. (2000) Computer-marked assessment in Moodle: Past, Present and Future. *Proceedings of the 2012 CAA International Computer Assisted Assessment Conference, 10th – 11th July 2012, University of Southampton* (Whitelock et al. eds).

ILIAS.DE. (2018) *ILIAS The Open Source Learning Management System*. `https://www.ilias.de` [Accessed 29 Jun. 2018].

MACCLUER, C. R. (2012) *Calculus of variations: mechanics, control and other applications*. Mineola, N.Y.: Dover Publications.

MAPLESOFT.COM. (2018) *Maple T.A. - Online Assessment System for STEM Courses - Maplesoft*. `https://www.maplesoft.com/products/Mapleta/` [Accessed 6 Jun. 2018].

MATHCENTRE.AC.UK. (2018) *Maths E.G.* `http://www.mathcentre.ac.uk:8081/mathseg/` [Accessed 6 Jun. 2018].

MAXIMA.SOURCEFORGE.NET. (2018) *Maxima, a Computer Algebra System*. `http://maxima.sourceforge.net` [Accessed 28 Aug. 2018].

MOODLE.ORG. (2018a) *Moodle*. `https://moodle.org` [Accessed 28 Aug. 2018].

MOODLE.ORG. (2018b) *Question types: STACK*. `https://moodle.org/plugins/stats.php?plugin=qtype_stack` [Accessed 29 Aug. 2018].

NUMBAS.ORG.UK. (2018) `https://www.numbas.org.uk/` [Accessed 6 Jun. 2018].

OPEN.AC.UK. (2018) *OpenMark Examples*. `http://www.open.ac.uk/openmarkexamples/` [Accessed 12 Jun. 2018].

R CORE TEAM. (2018) *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing, Vienna, Austria*. `https://www.R-project.org/` [Accessed 12 Sep. 2018].

ROBBINS, N. B. & HEIBERGER, R. (2011) *Plotting Likert and other rating scales*. Proceedings of the Survey Research Methods Section, American Statistical Association. `http://www.montana.edu/msse/Data_analysis/Likert%20Survey%20Graphs.pdf` [Accessed 31 Aug. 2018].

SANGWIN, C. (2013) *Computer Aided Assessment of Mathematics*. Oxford: Oxford University Press.

SANGWIN, C.J. (2015) *Who uses STACK? A survey of users of the STACK CAA system, May 2015*. Loughborough University Report. `https://dspace.lboro.ac.uk/2134/18540` [Accessed 12 Sep. 2018].

STACK.ED.AC.UK. (2018) *STACK*. `http://www.stack.ed.ac.uk` [Accessed 6 Jun. 2018].

Tim Lowe is Director of Teaching in the School of Mathematics and Statistics at the Open University, UK and is a Fellow of the Higher Education Academy. His background is in applied mathematics and scientific computing and he is interested in the use of technology to support and enhance the teaching of mathematics, particularly for distance learning students.

Ben Mestel is Director of the Mathematics MSc programme at the Open University, UK and is a Senior Fellow of the Higher Education Academy. His principal research area is nonlinear dynamical systems and he has an interest in the application of mathematics to energy systems.