CLASSROOM NOTES

Individualised summative assessments as used during COVID-19

Nathan Clisby and Antony Edwards

Department of Mathematics, Swinburne University of Technology, Melbourne, Australia

ARTICLE HISTORY
Compiled October 21, 2021

ABSTRACT
In this classroom note we outline a system of assessment used by the authors since 2020 to deliver individualised summative assessments to students from first- and second-year mathematics courses. Our system comprises three modular components allowing a mix-and-match of different technological approaches and mathematical question types. First is a question generation module which generates appropriate variables and question syntax, second is a delivery module to send out the individualised assessment to students, and third is a marking module to generate worked solutions and final answers for markers. The key benefits of these assessments are an ability to incorporate individualised authentic elements into assessments, to allow access to technology that would be impractical for an invigilated exam setting, whilst overall reducing the likelihood—but increasing the ease of detection—of academic misconduct and contract cheating, compared with other non-invigilated assessment protocols.

KEYWORDS
Tertiary mathematics education; assessment; COVID-19; academic misconduct

1. Introduction

In this classroom note, we present our platform-agnostic approach to generating individualised student assessments. This system allows complete flexibility for design, delivery, and feedback unconstrained by a single technology choice. It has been used at our university across multiple mathematics units, year levels and assessment types: from project-based assessments already using this approach prior to 2020 [5], to high-stakes mid-term tests and final exams included due to COVID-19 restrictions.

In common with the tertiary sector as a whole, over the last ten years at Swinburne University of Technology there has been a move towards a blended online approach to teaching and learning in tertiary mathematics classes, with the introduction of auxiliary video materials in the form of MathsCasts [12], automated lecture recordings, and live-streaming of classes accompanied with in-class polling [6]. In contrast to online teaching delivery, until the COVID-19 pandemic necessitated all assessments to be online, summative assessment in mathematics usually comprised of one or two invigilated mid-term class tests and a final invigilated exam. There is evidence that

The Version of Record of this manuscript has been published and is available in the International Journal of Mathematical Education in Science and Technology 7 Oct 2021 http://www.tandfonline.com/10.1080/0020739X.2021.1982040

CONTACT Nathan Clisby. Email: nclisby@swin.edu.au

such a reliance on assessment via high-stakes invigilated tests disadvantages women and students from marginalised (or non-traditional) backgrounds [11, 17], and does not assess what employers are looking for when hiring graduate employees [23].

Exceptions to these forms of summative assessment were made by the authors prior to 2020 to include those in the form of take-home projects where either a time limit was inappropriate or where external tools and information were required for the assessment to be meaningful or authentic to the question studied [5]. Such assessments were written for domain-specific service mathematics units such as those for students studying a Biology or Aviation Major and would be considered *authentic* in the sense of Ellis et al. [7]. We do note that although there is no consensus as to what exactly constitutes an authentic assessment [9], when appropriately embedded such assessments appear to be instrumental in helping students become good professionals [21]. It is within such projects that our present process of developing individualised assessment briefs originated and which in 2020, due to COVID-19 restrictions, was rolled out to encompass all tertiary mathematics assessment in the authors' units, including replacing high-stakes invigilated tests and exams. This classroom note concentrates on our system for authoring, delivering, and marking assessment, rather than the content of individual assessments. Finer points and various tips and tricks are illustrated by presenting appropriate examples and in the Discussion section.

2. Aims of the assessment delivery system

Individualised assessment in mathematics is constrained by a lack of flexibility in three aspects: assessment content, means of delivery to students, and the mechanism by which students submit answers. Learning and Teaching Interfaces (LTIs) set up by textbook manufacturers, paid subscription services, or open-source communities will allow different levels of control on these aspects when used in conjunction with a university's Learning Management System (LMS). At one extreme, a system will consist of a selection of pre-authored questions that can be delivered to students to answer on a proprietary software interface at a time of the instructor's choosing. Such an offering usually allows instructors no flexibility to modify questions, to change how students submit answers, or to modify how these answers are stored. Systems with sufficient flexibility to allow users to author their own questions must provide an interface for the authoring which may involve an instructor learning a suitable programming language e.g. Maxima in the case of STACK [19] or JME for Numbas [8]. Both pre-authored and user-authored systems have the benefit that an LTI provides an all-in-one system of question generation, student access, question marking, and feedback.

However, there are limitations in using an all-in-one approach. As noted by the creator of a popular computer-aided assessment tool in mathematics— STACK [19]—questions that are authored, implemented and marked by a computer algebra system may have "fundamental problems of . . . (i) question distortion, (ii) assessment only of lower-order skills, and (iii) strategic learning [of content]." [18, p. 987–988]. We note there has been much progress in this area, in particular in the areas of assessing students' understanding of proof [1], but it is still fundamentally the case that for higher-stake summative assessment, where approaches to answering questions are often as important as final answers, such systems are inflexible and unsuitable.

Wishing to allow complete flexibility for question authoring, and with the starting position that students will eventually be submitting written answers that would be manually marked, we developed a system that would not be reliant on a specific LTI or programming language. We wished to use whichever tool felt the most appropriate at any given time, whilst ensuring that others could follow the procedure we used (e.g. in the case of handing across to a colleague to teach). As assessments would be delivered for students to complete in their own time, we wanted to produce individualised assignments in which each student had their own (or at least a randomised set of) variables in a question to reduce the likelihood that students would see value in copying answers [c.f. 13]. Another requirement of the system was that, aside from the time for initial setup, we needed our assignments to be scalable for deployment and marking in large classes, with at most a small constant increase of marking time as the cohort size grew.

3. Assessment-independent modular structure

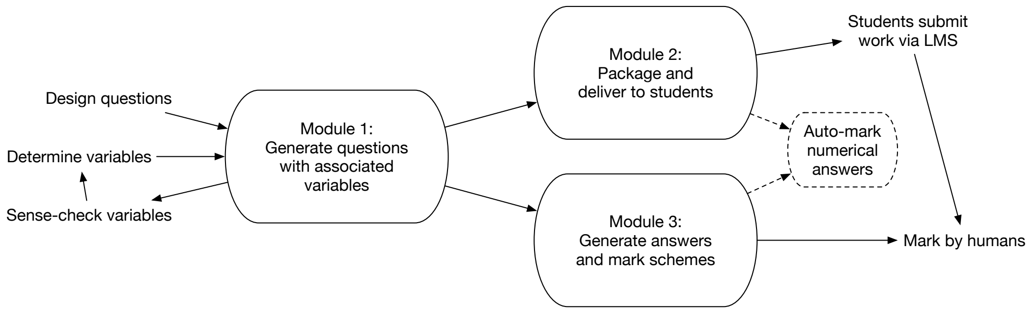

Figure 1. A flowchart describing the relationship between each stage of the assessment-independent structure. Dashed lines indicate possible extensions of the structure not yet realised by the authors.

The key feature of our system is to completely decouple the generation of questions, the delivery to students, and marking, and feedback (see Figure 1). By decoupling, we allow flexibility for an instructor to deliver the assessment that they deem appropriate for their students (within the proviso that it be an online assessment) using tools that they are familiar with. Within this overarching approach, in different assessments we have used Microsoft Office, MATLAB, Bash scripts in a Linux environment, and the programming language Julia at the various stages. Such an approach was essential during COVID-19 teaching semesters in 2020 when there was insufficient time to train in a new language or software approach. The best tool for a job is the one that you can use most effectively.

3.1. Generation of questions

For structured activities individualised questions require parameters to be embedded. Before this can happen, a question frame must be authored allowing for variable-based input, represented in the examples below by $\boxed{\text{boxed text}}$. Due to an increased potential to identify academic misconduct later, it is worth including esoteric phrases at this stage. Such questions may involve a standard setup with a quirky feature added, this itself may be a variable that changes alongside the in-question mathematics variables:

> A $\boxed{\text{LEGO monorail}}$ has a straight section of track along the negative x-axis. At the origin, the track connects to a transition curve with the equation $y = \boxed{3}x^3$ which connects

to a circular track at (x, y) where $x = \boxed{1}$ cm. Find the radius of the circular track.

As students were uploading typed or handwritten answers, more flexibility in question types was possible. Following the pedagogical approach of Watson and Mason [22], we found it particularly effective to ask students to use their own examples of objects subject to randomised constraints:

> Give an example of a vector $\vec{w} = (a, b)$ where $|\vec{w}| = \boxed{a}$, or explain why it is impossible.

We also developed more general free-form questions where students gathered their own objects with which to construct an analysis, subject to the variables in the question:

> Collect data for a recently completed flight between $\boxed{\text{Melbourne}}$ and $\boxed{\text{Perth}}$ on flightaware.com and model the direction at the flight's midpoint by a 2D unit vector. Why is this unit vector unlikely to be in exactly the same direction as the midpoint of the great circle connecting these cities?

A similar approach was taken in a statistics component where we asked students to perform analyses and formulate appropriate hypotheses regarding temperature and rainfall data at locations around Australia. Each student was emailed a unique pair of locations and provided with access to a CSV file with an extensive climate dataset collated from the Bureau of Meteorology [4], thus ensuring that their results would be unique to them.

To produce suitable rules for variables, it can be helpful by starting with the format of the answer, a process described well by Sangwin [18, ch 4]. Once the question frame is ready, the variables should be coded. Various approaches here are sensible, including manually listing items and using a random selection (Excel: `index` and `randbetween` commands, MATLAB: `datasample`), or using random selections based on logical rules (Excel and MATLAB: `rand`). For our Julia questions we looped over all possible combinations of parameter values, then subsequently randomised the assignment of these questions to students. Depending on the complexity of the question more advanced logic may be necessary, but this can be simplified by "pasting values" in Excel to remove issues where one random choice precludes the selection of another (e.g. two different zeros of a function), or using set differences in MATLAB. Finally, a CSV file is produced (Excel: `Save As`, MATLAB: `array2table` then `writetable`).

3.2. Delivery to students

In this stage variables are fitted into questions and then sent to students. This can be achieved via a mail merge in Microsoft Office, insertion of variables into LaTeX code, or direct insertion into an LMS where allowed.

For a mail merge approach, full questions can be authored in Word with variables added as mail merge fields (including in equations). Careful checking is required to confirm that variables are correctly formatted after the merge; field codes (Alt-F9) should be used to correct formatting (useful commands are # for decimal place rounding or currency options and @ for various date settings). If an Adobe Acrobat plugin is installed, individual PDF files can then be generated and automatically sent to each student email address during the mail merge.

If a university allows SMTP access to its email server, other approaches are possible. For instance, using the programming language Julia variables can be entered directly into individualised LaTeX documents, which can then be compiled into PDF files and emailed to students via appropriate scripts (we used Bash scripts in a Linux environment). One benefit of this approach is that it allows one to leverage the high quality mathematical typesetting of LaTeX documents.

3.3. Marking and feedback

We created marking guides for each human marker based on the randomised variables. Our goal was to make the marking process as efficient, fair, and consistent as possible. Note that the nature of the questions and the free-form responses of students necessitated a human marker. It would be possible to automate some aspects of the marking process within our system (see Figure 1), in particular numerical correct-incorrect answer types, but for us this would have involved students submitting answers to a system outside our LMS. Furthermore, given that human markers were necessary to assess free-form question parts, the relative ease of marking numerical answers at the same time would mean that the efficiency gain from a partially automated marking system would be minimal.

To create the marking guides we took the variable list and constructed milestone calculation points. The number of milestones depended on the complexity of the question; first-year calculus questions can often be checked by markers without an answer for comparison, while more sophisticated second-year questions may require half a dozen milestones to ensure that the marker can check the students' work without requiring laborious calculations. With additional time allocated to the development of the marking system it would be possible to have a separate milestone entry system for students where marks could be given for milestones, then verified by human markers. The human marker could then concentrate on assessing the mathematical communication skills of students.

Student submission of answers via an LMS also allowed rudimentary checks and measures such as automated Turnitin reports to be generated to identify any highly similar (typed) solutions. We found a high rate of false-positives in this approach; Turnitin tends to weigh more heavily question sentences over displayed mathematics yielding high similarity scores for any student who included the question text before their answers, for instance.

4. Discussion

As with any assessment containing variables that may take different forms for different students, attention must be paid to ensure variables give every student a meaningful question of an appropriate standard. A good way of doing this is first to test edge cases, then to sense check 10–15 randomly generated questions, taking into account that it is normal for fringe cases to throw up unexpected pairings where several variables are dependent on each other.

Designing effective parametrised questions can be challenging, especially if one wishes for the questions to be unique to each student. For a structured assessment for a second-year maths course we asked students to calculate Fourier series for an individualised piecewise linear function. Choosing a family of functions and associated parameters which were (a) unique, (b) of the same difficulty, and (c) pedagogically useful, was quite challenging. On reflection, the challenges of accommodating (a) and (b) led to the questions being more fiddly than was ideal for pedagogical utility. In particular, it may sometimes be necessary to sacrifice uniqueness in parameter choice, and instead have a variety of questions. One consequence of parametrised questions is that simple choices of parameters, which might be desirable for confidence-building questions, must generally be excluded. For instance, it is unfair to ask some students to compute $\int \sin x \, dx$ and other students to compute $\int \sin 7x \, dx$, as the latter integral is more challenging. There are other reasons for excluding the former: if a mail merge is used, it would result in the unwieldy $\int \sin 1x \, dx$.

Our markers reported that the additional time to mark a randomised question was negligible. This was partly because care was taken when generating marking guides to ensure that the order of solutions matched the order in which the LMS presented assessments. Furthermore, although the variables changed between students the context of the answers did not, so the usual speeding up of marking occurred as the ways in which students could err within a question became more established. Some markers opted themselves to construct more sophisticated guides that were then used by other markers. We note that it would be desirable to quantify the extra time required for marking individualised assessments, but as we did not implement any control group the best we can do is rely on qualitative feedback from markers.

The combination of esoteric phrases and randomly generated mathematical variables allowed for the identification of individual instances of academic misconduct. In one early major assessment, over 15% of students had either requested or viewed a solution from one single homework subscription website (and these were just those who did not obfuscate their online identity). As there is reasonable evidence that using individualised assessments decreases academic misconduct [13], we possibly encountered so many because it was much easier to detect via Turnitin and searches on popular homework-help websites. One approach that proved extremely useful, after we discovered that academic misconduct was occurring, was to create Google Alerts [10] with identifiable snippets of questions. In this way we would receive email alerts of instances of questions being posted to contract-cheating websites, typically within hours of these questions appearing. This measure significantly decreased the cognitive load, and worry, of continual monitoring for academic misconduct.

Our surprise is perhaps naive in the context of studies that found 6% [2] and 70% [3] of students admit to some form of cheating at university, and that authentic features in assessments may actually increase instances of academic misconduct [7]. As noted by Seaton [20], there appears to be little research in academic misconduct in tertiary mathematics, with the exception of some research on automated detection [14–16]. Although we observed a high rate of contract cheating in early assessments, it dramatically declined when it became clear to students that cheating could and would be detected, and once we established a dialogue around issues of academic integrity.

5. Takeaway

A system to deliver individualised assessments has been developed by the authors during the period of COVID-19 restrictions. This system incorporated flexible components of generation, delivery, and marking, so as to make it highly scalable. We have found that the risks of academic misconduct, and in particular contract cheating, can be substantially mitigated by having a dialogue with students about academic integrity, and through invigilation by monitoring the appearance of test questions online via Google Alerts. These assessments are more authentic than the time-limited in-class tests they have replaced and will continue to be used by the authors when the COVID-19 crisis ends.

Acknowledgements

We thank Emily Cook for the use of her LEGO monorail example question.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

The authors are happy to provide code and parametrised questions on request.

ORCID

Nathan Clisby https://orcid.org/0000-0001-7068-8347
Antony Edwards https://orcid.org/0000-0002-8816-5488

References

[1] R. Bickerton and C. Sangwin, *Practical online assessment of mathematical proof*, International Journal of Mathematical Education in Science and Technology (2021), pp. 1–24.
[2] T. Bretag, R. Harper, M. Burton, C. Ellis, P. Newton, P. Rozenberg, S. Saddiqui, and K. van Haeringen, *Contract cheating: a survey of Australian university students*, Studies in Higher Education 44 (2019), pp. 1837–1856.
[3] M.A. Broeckelman-Post, *Faculty and student classroom influences on academic dishonesty*, IEEE Transactions on Education 51 (2008), pp. 206–211.
[4] Bureau of Meterology, *Climate Data Online*. Available at http://www.bom.gov.au/climate/data/.
[5] A. Edwards, *Individualised summative assignments that can be scaled sustainably*, First Year in Mathematics Group, University of Melbourne, 2018. Available at https://fyimaths.files.wordpress.com/2018/08/edwards-2018-fyi-maths-handout3.pdf.
[6] A. Edwards, *Student polling: It's the taking part that counts*, Proceedings of the 12th Delta conference on the teaching and learning of undergraduate mathematics and statistics, Fremantle, Western Australia, 2019. Available at https://researchdirect.westernsydney.edu.au/islandora/object/uws%3A52645.
[7] C. Ellis, K. van Haeringen, R. Harper, T. Bretag, I. Zucker, S. McBride, P. Rozenberg, P. Newton, and S. Saddiqui, *Does authentic assessment assure academic integrity? Evidence from contract cheating data*, Higher Education Research & Development 39 (2020), pp. 454–469.
[8] B. Foster, C. Perfect, and A. Youd, *A completely client-side approach to e-assessment and e-learning of mathematics and statistics*, International Journal of e-Assessment 2 (2012). Available at https://ijea.org.uk/index.php/journal/article/view/35
[9] B.B. Frey, V.L. Schmitt, and J.P. Allen, *Defining authentic classroom assessment*, Practical Assessment, Research, and Evaluation 17 (2012), p. 2. Available at https://doi.org/10.7275/sxbs-0829
[10] Google Alerts. Available at https://www.google.com/alerts.
[11] M.M. Jameson and B.R. Fusco, *Math anxiety, math self-concept, and math self-efficacy in adult learners compared to traditional undergraduate students*, Adult Education Quarterly 64 (2014), pp. 306–322.

[12] B. Loch, O. Gill, and T. Croft, *Complementing mathematics support with online MathsCasts*, ANZIAM Journal 53 (2011), pp. C561–C575.

[13] S. Manoharan, *Personalized assessment as a means to mitigate plagiarism*, IEEE Transactions on Education 60 (2016), pp. 112–119.

[14] N. Meuschke, M. Schubotz, F. Hamborg, T. Skopal, and B. Gipp, *Analyzing mathematical content to detect academic plagiarism*, in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 2017, pp. 2211–2214. Available at https://doi.org/10.1109/JCDL.2019.00026

[15] N. Meuschke, V. Stange, M. Schubotz, M. Kramer, and B. Gipp, *Improving academic plagiarism detection for STEM documents by analyzing mathematical content and citations*, in *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 2019, pp. 120–129.

[16] M. O'Malley and T.S. Roberts, *Plagiarism on the rise? Combating contract cheating in science courses*, International Journal of Innovation in Science and Mathematics Education 20 (2012). Available at https://openjournals.library.usyd.edu.au/index.php/CAL/article/view/5803

[17] D.L. Oswald and R.D. Harvey, *Hostile environments, stereotype threat, and math performance among undergraduate women*, Current Psychology 19 (2000), pp. 338–356.

[18] C. Sangwin, *Computer aided assessment of mathematics*, OUP Oxford, 2013.

[19] C. Sangwin, *Computer aided assessment of mathematics using STACK*, in *Selected regular lectures from the 12th international congress on mathematical education*. Springer, 2015, pp. 695–713.

[20] K.A. Seaton, *Laying groundwork for an understanding of academic integrity in mathematics tasks*, International Journal of Mathematical Education in Science and Technology 50 (2019), pp. 1063–1072.

[21] V. Villarroel, D. Boud, S. Bloxham, D. Bruna, and C. Bruna, *Using principles of authentic assessment to redesign written examinations and tests*, Innovations in Education and Teaching International 57 (2020), pp. 38–49. Available at https://doi.org/10.1080/14703297.2018.1564882

[22] A. Watson and J. Mason, *Mathematics as a constructive activity: Learners generating examples*, Routledge, 2006.

[23] M. Yorke, *Employability in higher education: what it is - what it is not*, Vol. 1, Higher Education Academy York, 2006.