

Reflections on six years of computer-aided assessment at the Jinan–Birmingham Joint Institute

ROBERT LEEK[†], DANIEL JONES, JOHN MEYER

School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

In this article, we reflect on our experience teaching on transnational Mathematics programmes at the Jinan-Birmingham Joint Institute, particularly the use of the Möbius computer-aided assessment system throughout all modules and levels taught by the University of Birmingham. The particular context of our institute, including its transnational educational model used and the rationale for our heavy reliance on these assessments, is described in detail prior to discussing how we design and implement these assessments. The effect of the COVID-19 pandemic on academic integrity forced us to reconsider our assessment diet, and this paper concludes with a reflection on our teaching experience, focusing on the capabilities of computer-aided assessment systems and how they are used in practice by academics with a wide variety of experience.

1. Introduction

In 2017 the first cohort of Jinan-Birmingham Joint Institute (J-BJI) students commenced their studies. Throughout the following years, we have led in delivering computer-aided assessment (CAA) at the J-BJI in various roles. Following the graduation of the 2017/18 cohort of J-BJI students in 2021, we believed it was important to share our experience of and reflections on delivering CAA via the Möbius platform at the J-BJI with the broader mathematics education community. We use CAA to refer to assessments with questions automatically marked by a pre-written grading code. This paper covers a 6-year period from Summer 2017, when we had no CAA content, to the start of the 2023/24 academic year, when we have overseen 188 Möbius assessments, spanning twelve lecture courses and all three stages of our degree programmes.

Support within the literature for wider dissemination of lecturers' perspectives on their engagement with CAA of mathematics can be found in a recent research agenda for e-assessment (Kinnear, et al., 2024), specifically Questions 32 and 33. About a decade ago (Broughton, et al., 2013) and more recently (Davies, et al., 2022; Davies, et al., 2024) have reported on lecturers' opinions of implementing CAA in a UK HE setting. As observed in the latter, we also agree that there is not a substantial literature on lecturers' opinions concerning the implementation of CAA in Mathematics programmes at scale. We also note that discussions of CAA implemented using Möbius are infrequent in the literature. This article intends to address both topics and contribute to this body of research.

Transnational education (TNE) is a significant form of higher education provision from UK universities, and as of the 2021/22 academic year there are over 500,000 students registered at TNE providers, with approximately 40% of these studying under collaborative provision (like at J-BJI), compared with approximately 2.86 million students registered at UK Higher Education Institutions (HESA, 2023; Universities UK International, 2023).

As TNE and CAA may be unfamiliar to many readers, we provide a detailed exposition of our teaching and assessment arrangements (sections 2–3) to provide the appropriate context for the

[†] Corresponding author. Email: r.leek@bham.ac.uk

rest of the paper. In later sections will discuss how we develop content in Möbius (section 4), the changes made to our assessment resulting from the COVID-19 pandemic (section 5), and finally we conclude with reflections on our experience (section 6).

2. An overview of the J-BJI

The University of Birmingham (UoB) uses a fly-in/fly-out model (QAA, 2024) for staff that teach mathematics at the J-BJI, whereby academic staff travel to Jinan University (JNU) in Guangzhou, China, to deliver their teaching. The ‘flying faculty’ are responsible for the delivery of core mathematics modules across four undergraduate dual-degree programmes (University of Birmingham, n.d.); these are B.Sc. Applied Mathematics with:

- Economics (Econ),
- Information Computing Science (ICS),
- Mathematics¹ (MAM), and
- Statistics (Stats).

Specialist J-BJI staff deliver English language education and JNU faculty deliver discipline-specific content for the four dual-degree programmes. The flying faculty have expertise in various mathematical subjects, primarily linked to subjects taught at the J-BJI. Notably, staff have vastly different experiences and familiarity with CAA, whether as students or instructors.

Each academic year consists of two semesters, each containing sixteen teaching weeks, split into four blocks of four weeks. During each block, academics delivered ten course credits² of material. Under ordinary proceedings, UoB academic staff travel to Guangzhou to provide teaching in person. However, due to travel restrictions caused by the COVID-19 pandemic, education has been delivered remotely (Jones, et al., 2021) from February 2020 to June 2022, with in-person teaching resuming partially in Autumn 2022 and fully by Spring 2023. Further context for J-BJI can be found in (Du Croz & Morris, 2024), particularly on our summer school and study abroad programmes.

| Academic Year | Year 1 | | Year 2 | | Year 3 | | Total | |
|---------------|-------------------------|---------------------------|-------------------------|---------------------------|-------------------------|---------------------------|-------------------------|---------------------------|
| | Approx. No. of students | No. of Möbius assessments | Approx. No. of students | No. of Möbius assessments | Approx. No. of students | No. of Möbius assessments | Approx. No. of students | No. of Möbius assessments |
| 2017/18 | 100 | 16 | | | | | 100 | 16 |
| 2018/19 | 200 | 16 | 100 | 16 | | | 300 | 32 |
| 2019/20 | 210 | 16 | 200 | 16 | 100 | 16 | 510 | 48 |
| 2020/21 | 260 | 15 | 210 | 14 | 200 | 15 | 670 | 44 |
| 2021/22 | 210 | 8 | 260 | 8 | 210 | 8 | 680 | 24 |
| 2022/23 | 270 | 8 | 210 | 8 | 250 | 8 | 730 | 24 |

TABLE 1. *Number of Möbius assessments and students (based on assessment data) from the first 6 years of delivery at the J-BJI. Approximate numbers of students are to the nearest 10 for years 1–3, with the total in the last column being the sum of these approximations and not necessarily the true approximate total.*

¹ ‘Mathematics’ here refers to Pure Mathematics.

² These credits are worth the same as other degrees and modules provided by UoB but differ from the credit system in use at JNU.

For the academic years listed in Table 1, continuous summative assessments for each 20-credit module typically comprised of four assessments, and students were assessed every two weeks, national holidays permitting.³ During the 2017/18–2020/21 academic years, all of these assessments were computer-aided, delivered using Möbius (DigitalEd, n.d.), formerly known as MapleTA. From the 2021/22 academic year, we converted half of these assessments to hand-written assessments, and we will be discussing our rationale for this change in section 6. A distinct requirement of our teaching arrangements is that a sufficiently sizeable proportion of each assessment must differ by at least 20% between the two streams of students: Econ & Stats vs MAM & ICS.

Initially in 2017/18, the continuous assessment arrangements for each module consisted of weekly, group-submitted solutions to problem sheets, and fortnightly, invigilated class tests lasting 45-minutes using Möbius. The former had no contribution to the final module grade, whilst the latter contributed 20%. There are sound pedagogical bases for students to work together, and work individually, to gain understanding of mathematics (D'Souza & Wood, 2003). However, the practical constraints for providing these assessments were the marking time required from module leads for formative assessments, timeliness of feedback, and availability of suitable computer laboratories and invigilators (Robinson, et al., 2012, pp. 115-116). Until our first experience teaching at the J-BJI in the Spring of 2018, the total capacity of computer laboratories at the J-BJI was unknown to the authors.

These short class tests at the J-BJI typically contained up to 10 brief questions, which adhered to principles advocated in (Smith, et al., 1996; Greenhow, 2015). Given that most academic staff had little-to-no experience with CAA either as students or lecturers, this arrangement allowed staff to familiarise themselves with Möbius where requirements on randomisation and conceptual intricacy of questions were limited by the format of the assessment (Robinson, et al., 2012, p. 114); for example, invigilation meant that randomisation was less important, albeit still implemented via 3–5 variants of a randomised question, and questions testing recall were reasonable.

Following discussions in the flying faculty team in the Summer of 2018, we decided that our continuous assessments should be *for* learning (Wiliam, 2011) rather than *of* learning. End-of-semester examinations were deemed a suitable assessment of learning to determine if students had met learning outcomes and, consequently, progression requirements. Combined with the projected growth in student numbers, we decided to replace the class tests for future cohorts with noninvigilated assessments open for at least 24 hours. However, we maintained class tests for the 2017/18 cohort, as these students were already familiar with this format and student numbers were sufficiently small to manage. For newly appointed academic staff, this also provided a more manageable introduction to assessment authoring in Möbius.

From the 2018/19 cohort onwards, we granted more time for students to complete their continuous assessments, so we could:

- provide feedback to students on a broader variety of material than the 45-minute class tests permitted;
- incorporate more types of questions (see the scheme of (Pointon & Sangwin, 2003, p. 675) for a suggested classification of question types);
- give students more authentic assessments (Wiggins, 2011), as outside of an educational setting, one is rarely asked to answer questions with stringent time limits comparable to that of a class test; and
- encourage students to not only answer questions but also consider how they can justify their correct answers.

Practicalities were also a driver for this change since weekly student timetables differed between the four dual degree programmes, and thus finding periods of time where students were all

³ The number of assessments in the 2020/21 academic year are slightly lower, due to the COVID-19 pandemic.

comparably free was only guaranteed over sufficiently short (< 1 hour) or sufficiently large periods (> 24 hours) of time. This did result in staff having to spend more time and effort designing and implementing their assessments, which we discuss this further in section 4.

3. Procurement and management of Möbius

The decision to use Möbius to deliver CAA was based on several factors:

- existing institutional expertise in the use of MapleTA at UoB;
- access to professional technical support from DigitalEd (formerly Maplesoft) to keep the CAA system running;
- ability to deliver assessments based on a server physically hosted in China; and
- the price-point.

We agree with the observations of other UK-based practitioners that for the advantages one gains from commercial solutions, one compromises their control and ability to develop the system. We find that without the following essential enhancements designed ad-hoc by UoB's Engineering and Physical Sciences (EPS) EdTech team and RL (re-grader; batch question testing; batch file downloader; wrapper for defining MapleTA variables in algorithm section of question authoring; see section 4.2 for details), we could not effectively use Möbius to author and test questions, assess students, or employ suitable quality assurance processes when grading assessments.

Physically locating the server in China is required to provide reliable access to Möbius for students, as well as security considerations for UoB. This creates manageable but frustrating syncing issues since academic staff typically develop Möbius content in the UK, on another instance of Möbius hosted within the EU that is used for Birmingham-based students. Due to differing timetables at UoB and J-BJI, updates to the Chinese Möbius server typically occur at different dates/times compared to the EU server, albeit sometimes during J-BJI teaching periods, which on occasion has caused disruption. Software updates are a general requirement of any CAA setup; disruption to staff and students can be avoided via clear communication with all stakeholders, however since 2021 DigitalEd automatically upgrades Möbius instances and no longer notifies of planned maintenance periods, as part of their move to cloud infrastructure. Moreover, Maple library locations for UoB pre-written grading code require manual updates whenever questions are transferred between both versions of Möbius.

When students' access to assessments/feedback in Möbius has been unintentionally limited, the primary causes have been:

1. the LTI link (Blackboard, n.d.) between Möbius and the virtual learning environment Blackboard being broken, for example, due to required/ongoing maintenance of Blackboard, or inconsistent timestamps between servers;
2. human error, for example, due to academic staff setting incorrect assessment policies (DigitalEd, n.d.);
3. bugs within Möbius; and
4. students using unsupported browsers (e.g. Internet Explorer; Microsoft Edge; 360 Secure Browser; outdated versions of supported browsers) and issues with 3rd party cookies and our LMS (for example, Safari blocks these by default).

Maintaining clear communication with JNU IT support and the UoB administrative team based at the J-BJI, who gather student feedback, has been key to mitigating item 1. To address item 2, a guide for setting assessment policies has been provided for staff, explicitly stating which options must be set and when (i.e. before/after the assessment). All authors confirm that no matter how much experience staff gain using Möbius, human error can still limit access to Möbius for J-BJI students. On item 3, we typically need to observe the bug before mitigating it. Bugs were reported to DigitalEd, and a practical workaround was created to avoid situations where the bugs arise. This

often leaves bugs unfixed but mitigated for our purposes (for instance, only allowing specific staff to use the proctor tools to re-open assessments for individual users). Finally, item 4 is addressed by reminding students to use supported browsers (DigitalEd, n.d.) which currently consists of Google Chrome and Safari (Firefox support was dropped in August 2022).

4. Developing content in Möbius

At the start of the 2017/18 academic year, teaching staff (including RL and JM) were trained in developing content for Möbius. This consisted of working through the Möbius School classes on the Möbius Community forum (DigitalEd, n.d.) and attempting to create simple questions, including *multiple-choice* and *Maple-graded response areas*. As lecturers were busy preparing their teaching materials for the first cohort of students, as well as helping manage J-BJI, student interns were hired during the summer of 2017 and the winter of 2017/18 to code questions written by the lecturers. The interns were UoB students based in the UK and did not have contact with students at the J-BJI, and they have blogged about their experiences in (Johnson, et al., n.d.). This allowed lecturers to focus on the design of questions and feedback rather than their implementation, as well as providing valuable programming experience to interns, not just in Maple but also HTML, CSS, and JavaScript for the Möbius HTML response areas.

After the questions were written, interns tested them and provided there were no issues, these questions were passed back to lecturers to provide final checks. Ultimate responsibility for assessments rested with the relevant lecturers. Questions had to be occasionally revised by either interns or lecturers, sometimes due to imprecise specifications from lecturers. In later years, lecturers typically reworked these questions and wrote new questions using the interns' question generation and grading code as a template.

This working arrangement allowed for a separation of duties between design and implementation, and with slight variations, continued till the start of the 2020/21 academic year, under the supervision of the authors. We have not hired interns to develop questions for us since, as our team of lecturers were familiar with the design and implementation of assessments in Möbius, and we had sufficient assessment resources that could be reused with minor modifications (e.g. changing randomised parameters). Several colleagues could also provide administrative support for our assessments, adding resiliency to our operations.

Most applications of CAA are seemingly aimed at lower-level courses, as noted in (Lowe & Mestel, 2020, p. 68), and use a core group of 'designers' to write and code the questions. In contrast, at J-BJI we use CAA throughout all levels of our programmes. With the former approach, it would be a formidable to write and design questions for all courses, in particular for advanced statistical and optimisation topics, which is why we have opted for staff to write questions themselves. This approach also gives staff additional agency in the delivery of their modules.

4.1. Question design and regrading

As part of their training, lecturers are provided with guidelines to aid in the design of their questions. The main points are summarised below:

- Lecturers should specify the layout of the question and response areas, as well as the expected format of answers (e.g. whether to include brackets and which type, if the answer is a list/vector).
- Questions should ideally be randomised (e.g. drawing parameters from a specific range), ensuring that different instances are of comparable difficulty. The choice of suitable ranges should be made with due consideration when question difficulty is particularly sensitive to parameter choices (e.g. finding highest common factors using the Euclidean algorithm).
- General solutions should be provided to each question. If the lecturer intends for students to

apply algorithms, interns should be notified so that this algorithm can be implemented using Maple. This helps future-proof the question against changes to parameters.

- The mark scheme should specify when partial marks are to be awarded, and lecturers are encouraged to carry forward errors to avoid penalising students multiple times for the same mistakes.
- Lecturers should avoid long, multi-part questions where possible, as it makes regrading harder (see below for details about our regrading process).
- Question parts could be implemented as separate questions if the lecturer wishes.

Many of our lecturers first encounter CAA with the expectations of paper-based assessments, where students can give free-form answers, and mark schemes are revised for any edge cases that arise. This has also been observed in the literature (Kinnear, et al., 2024, pp. research question 23, p. 13) and can cause issues, mainly when the form of a solution is essential. For example, consider the following question: *give the second-order Taylor series expansion of $\exp(x)$ about the point $x = 1$* (Meyer & Leek, 2020). Ideally, we would like to award partial marks for correct coefficients; however, if a student responds with ' $1 + (e/2)x^2$ ', it is unclear how to interpret this answer. Did they calculate the expansion about the point $x = 0$ (in which case only the constant term is correct), or did they expand about the point $x = 1$ but miscalculate the constant term, and if so, did they intend to omit the linear term?

In a paper-based assessment, we would typically have their working on their answer and be able to infer the student's intent. Whilst we do request students upload their working for academic integrity purposes (see section 5 for more), this does not aid the automated marking. By designing these questions carefully and including intermediate steps, we aim to avoid such issues. Unfortunately, DigitalEd has yet to implement linked answer boxes (Watkins, 2018) despite this being the most popular feature request on DigitalEd's Ideas Portal. If we want to create multi-part questions or questions that award marks dependent on multiple response areas, we either must incorporate the entire question with an HTML response area (which hinders our capability to regrade as loading hundreds of student responses for this question type takes an excessive amount of time, hence our advice to avoid this where possible) or ask students to provide all responses as a list. This creates unwanted tension between wanting to assess comprehensively, simplifying regrading, and reducing friction in the student experience by prompting for 'natural' input (for example, the 2D math response area allows students to input matrices, but if an HTML response area is required to link multiple response areas together, then such an input would need to be manually coded).

Much of the advice pertains to making the 'regrading' process more accessible and providing an improved student experience. Once an assessment has concluded, a regrader will check for any issues before releasing the results and feedback to students. This was initially someone other than the module lead and in later years this role involved moderation of our assessments (particularly important when continuous assessments contributed 50% towards module marks during the 2020/21 academic year, as discussed in section 5), before finally allowing module leads to regrade their own assessments as our Möbius questions became more stable and robust. Most problems are caused by incorrect syntax, however understanding Maple syntax is not part of our learning outcomes for our modules and hence should not be a factor in our marking. To mitigate these issues, we provide syntax hints (see figure 2) and during their orientation week students undertake a 'student-readiness test' that teaches them how to use Möbius and input answers. Nevertheless, students will still occasionally make mistakes in future assessments. Unfortunately, as Möbius must output a mark for each response area (DigitalEd, n.d.) instead of a more detailed data format, if Maple encounters an error, then students will be awarded 0 marks. We feel very strongly that this situation should be avoided as much as possible, and we have developed several tools to improve the student and staff experience of Möbius.

4.2. *Improvements to Möbius from UoB*

To aid our error-detection, we have a standardised template for the grading code within a response area (called ‘error-catching code’) that assigns 1% in case an error is encountered in Maple so that these issues will be visible in the Gradebook. This method is not 100% fool-proof: whilst Maple expects asterisks when multiplying, it also treats numerals as constant functions, so ‘ $(x+1)^2$ ’ raises an error, but ‘ $2(x+1)$ ’ does not and outputs ‘2’. This could be fixed by using custom parser to achieve comparable functionality with STACK, which can report back in real-time to students when they have syntax errors (Sangwin, 2013, pp. 111-112). Möbius provides an optional response preview, which is insufficient from our perspective, and a feature request to implement a syntax checker was turned down (Watkins, 2018). This ultimately results in more correction of marks due to syntax errors, resulting in more staff time spent on regrading and lower student confidence in the Möbius platform and/or their lecturers.

Another common issue with regrading is using Unicode characters that Maple/Möbius does not correctly interpret. As we teach Chinese students, they will often input characters using their Chinese keyboard layouts, such as fullwidth and local punctuation characters (Unicode, 2022; Unicode, 2022), as seen in the following string: ‘【 1 , 2 , 3 】’. The error-catching code identifies these issues for regrading, and was updated in 2022 to automatically substitute common characters and check for any remaining characters that are not printable ASCII characters (Unicode, 2022).

UoB’s EPS EdTech team wrote a JavaScript bookmarklet (figure 1) that collates student responses on the currently viewed question and modifies marks for all students that submitted a particular response. This improves the speed of regrading, but it can still take a significant amount of time, especially as the rendering of MathJax and HTML response areas in the Möbius gradebook is slow. The EdTech team also developed additional JavaScript bookmarklets to simplify the algorithm section of question authoring by using the Maple language and converting it automatically into the Möbius language, and RL has developed another bookmarklet to batch download PDFs submitted by students in the gradebook view (again, not available within Möbius).

4.3. *Improvements to Möbius from DigitalEd*

Since DigitalEd was spun-off from Maplesoft to handle Möbius product development, there have been a few notable improvements to the Möbius user experience, from our perspective:

- Question regrading (rerunning grading code) was introduced in version 2019.1. This allowed staff to alter their grading code after an assessment was released to correct any bugs in the code and/or adjust the code to reflect an updated mark scheme.
- Policy Sets were introduced in version 2019.2. There have been several instances where staff have incorrectly set up their assessments, and we tried to use this feature by creating two policy sets (‘Before and during test’ and ‘After regrading’) that would handle all the settings needed for our assessments. However, when we tried to use this, we found we could not unset specific options and required an extensive workaround to get it working again. We have not used this feature since.
- Activity Grading View (purple manual grade button) was introduced in version 2020.1. This has made it slightly easier to regrade HTML question types as it only loads a single student’s response at a time.

| Count | Question Index | Student Answer | Correct Answer | Raw Response | Mark |
|-------|----------------|--------------------------------|--------------------------|--------------|-------|
| 1 | 0 | 45405360*a^2*b^5*c^8 | 135135 $a^2 b^5 c^8$ | | 0 / 1 |
| 1 | 0 | 25025*a^8*b^3*c^4 | 225225 $a^8 b^3 c^4$ | | 0/1 |
| 1 | 0 | 1175720*a^4*b^3*c^12 | 1763580 $a^4 b^3 c^{12}$ | | 0/1 |
| 1 | 0 | 17325 | 225225 $a^8 b^3 c^4$ | | 0/1 |
| 2 | 0 | 135135 | 135135 $a^2 b^5 c^8$ | | 0.5/1 |
| 1 | 0 | 1763580*a*b*c | 1763580 $a^4 b^3 c^{12}$ | | 0.5/1 |
| 6 | 0 | 225225*a^8*b^3*c^4 | 225225 $a^8 b^3 c^4$ | | 1/1 |
| 3 | 0 | 1058148*a^2*b^12*c^5 | 1058148 $a^2 b^{12} c^5$ | | 1/1 |
| 2 | 0 | 1763580*a^4*b^3*c^12 | 1763580 $a^4 b^3 c^{12}$ | | 1/1 |
| 2 | 0 | 1763580*(a^4)*(b^3)*(c^12) | 1763580 $a^4 b^3 c^{12}$ | | 1/1 |
| 1 | 0 | 135135*a^2*b^5*c^8 | 135135 $a^2 b^5 c^8$ | | 1/1 |
| 1 | 0 | 135135*(a^2)*(b^5)*(c^8) | 135135 $a^2 b^5 c^8$ | | 1/1 |
| 1 | 0 | 225225*a^8*b^3*c^4*x^8*y^3*z^4 | 225225 $a^8 b^3 c^4$ | | 1/1 |
| 1 | 0 | 255255*a^8*b^3*c^4 | 225225 $a^8 b^3 c^4$ | | 1/1 |

FIG 1. An example output from the regrader script. Different responses have been collated (see Count column) so that marks can be changed simultaneously.

5. COVID-19, academic standards, and plagiarism

The UK government's then foreign secretary Dominic Raab's statement on 4 February 2020 advised against 'all but essential' travel to China (Foreign & Commonwealth Office and Raab 2020). Shortly afterward on 7 February 2020, a 'closed management of communities' policy was implemented in Guangzhou (Foreign Affairs Office of the People's Government of Guangzhou Municipality, 2020). Hand-written, group-submitted assessments at J-BJI were cancelled for the remainder of the academic year and end-of-year examination were significantly modified for open-book, online assessment. This move from 'class tests' to 'take-home exams' (Bengtsson, 2019) became necessary due to the logistical challenges of remote invigilation (see Jones, Meyer and Huang 2021 for more details on our remote teaching experience during this time), and inevitably increased the risk of plagiarism (Lancaster & Cotarlan, 2021; Comas-Forgas, et al., 2021; Institute of mathematics & its applications, London Mathematical Society, Royal Statistical Society, 2022).

In the 2019/20 academic year, this was less of a concern due to the limited contribution (20% per module) of these assessments towards overall module marks and the randomisation of parameters in most questions; however the Framework for Education Resilience introduced at UoB for the 2020/21 academic year increased this contribution of continuous assessment for each module to 50% (Armour & Senior Education Team, 2020) so that students could progress/graduate even with disrupted end-of-semester exams, and incidentally provided further incentive for plagiarism. It was difficult to anticipate how, and to what extent, this would affect module marks and the student experience. It was not straightforward to create CAA questions that were sufficiently challenging relative to the exams of previous years, as expected by (QAA, 2020), although we improved over time and helpful, and practical ideas began to appear soon after (Bickerton & Sangwin, 2022).

It was decided that increasing the frequency of summative assessments was not appropriate to accommodate the increase in continuous assessment weighting, but instead to increase the difficulty and scope of CAA assessments. Despite our efforts, most modules had very high average marks for these assessments, in many cases over 90%, which resulted in many students approaching

their final exam having already passed that module, as well as little differentiation between students. Combined with limited feedback on their hand-written work, this likely led to misaligned expectations so students believed they would perform exceptionally well on exams, which were more challenging than usual due to their open-book nature and the School of Mathematics exam preparation guidance for that year. We list below several types of questions and examples used in CAA that were more successful in maintaining academic standards and providing a sufficient challenge to students:

1. Questions with increased complexity (e.g. requiring more intermediate steps to solve) that only award marks for the final answer (see figure 2 below).
2. Questions requiring a proof or a counterexample to determine the answer but don't specify what the student should do. Multiple-choice can be a suitable format for these questions, as discussed in (Lyakhova, 2023, pp. 12-13) and demonstrated in figure 3 below.
3. Questions with increased conceptual difficulty, including those that assess proof comprehension, such as the proof fallacy example in (Bickerton and Sangwin 2022, figure 6).
4. Questions which defeat WolframAlpha⁴ (see figure 4 below).

One should acknowledge that the difficulty of a question is not intrinsic to the question; the setting in which a question is answered matters. For example, inverting a given 3-by-3 matrix is more prone to error in an invigilated, closed-book setting without a calculator than in an open-book, noninvigilated setting without time pressure and access to online tools.

möbius

Preview

A group G contains three elements a, b, c satisfying the following equations:

- $cb = abc$,
- $ca = a^2c$,
- $a^3 = e$,
- $ab = ba$.

Express the element $cb^{-1}ab^{-1}c^{-1}$ in the form $a^i b^j c^k$ for some integers i, j, k .

Grade

How Did I Do?

Refresh

Close

FIG 2. A question from RL's first-year Algebra & Combinatorics course, which asks students to convert a word into a normal form using the provided group relations. By choosing words which require the same number of applications of these relations, different versions of this question can be created with comparable difficulty.

⁴ChatGPT, one of the first publicly available, consumer-friendly AI chatbots based on a large language model, was only released in November 2022 (OpenAI, 2022) and as such AI chatbots were not considered a serious plagiarism threat at the time.

Preview

Suppose that the linear system $\mathbf{A} \cdot \underline{x} = \underline{b}$ has a unique solution and $\mathbf{A} \in \mathcal{M}_{mn}(\mathbb{R})$ with $m, n \in \mathbb{N}$ for some non-zero $\underline{b} \in \mathcal{M}_{m1}(\mathbb{R})$. Select all of the statements that necessarily follow:

- ☐ $m \leq n$.
- ☐ There does not exist a basis of $\text{col}(\mathbf{A})$ containing \underline{b} (considering $\underline{b} \in \mathbb{R}^m$).
- ☐ $\text{rank}(\mathbf{A} \mid \underline{b}) = n$.
- ☐ There exists a basis of $\text{col}(\mathbf{A})$ containing \underline{b} (considering $\underline{b} \in \mathbb{R}^m$).
- ☐ $\text{rank}(\mathbf{A}) = m$.
- ☐ If $m = n$ then \mathbf{A} is a transition matrix (on \mathbb{R}^m) for some bases B_d and B_{cd} in the domain and co-domain respectively.
- ☐ $\text{row}(\mathbf{A}) \neq \mathbb{R}^n$.
- ☐ If $m = n$ then $|\mathbf{A}| \neq 0$.
- ☐ If $m = n$ then \mathbf{A} is not a transition matrix (on \mathbb{R}^m) for any bases B_d and B_{cd} in the domain and co-domain respectively.

Grade

How Did I Do?

Close

FIG 3. A question from JM's first-year Vectors, Geometry, & Linear Algebra course. To determine the correct answers, students should prove or construct a counterexample for each statement, connecting many different concepts in this course.

Preview

State the value of $\lim_{n \rightarrow \infty} \sin(2e\pi n!)$ or input NE if the limit does not exist.

Grade

How Did I Do?

Close

discrete limit of $\sin(2 \cdot e \cdot \pi \cdot n!)$ as n tends to infinity

NATURAL LANGUAGE



MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Discrete limit

$$\lim_{\substack{n \rightarrow \infty \\ \mathbb{Z}}} \sin(2e\pi n!) = \lim_{\substack{n \rightarrow \infty \\ \mathbb{Z}}} \sin(2e\pi n!)$$

$n!$ is the factorial function

$\lim_{\substack{x \rightarrow h \\ \mathbb{Z}}} f$ represents the discrete limit of f as x approaches h

Download Page

POWERED BY THE WOLFRAM LANGUAGE

FIG 4. A question from Yuzhao Wang's first-year Real Analysis and Calculus course. WolframAlpha was unable to answer this question, even when specified as a discrete limit.

5.1. Academic integrity measures

Plagiarism is a clear threat to take-home exams (Bengtsson 2019, section 4.2) such as our CAA setup; in particular, if questions lack randomisation, responses to plagiarism are not sufficiently severe, or the risk of getting caught is not strong enough. Inevitably, some students will commit plagiarism and not be caught, but this will be true for any system of detecting plagiarism. Various ideas were implemented to address the above concerns, as outlined below.

Randomisation can be difficult and time-consuming to implement. A suggested method of randomisation, typified by its simplicity in implementation, utilises single instances of parameterised values which modify the solution to a problem. This is considered desirable when plagiarism is a concern. We can easily detect a student providing a solution that is not correct for their question but valid for a different instance, as in figure 5 below, or even when multiple students submit the same peculiar response. Such questions should be designed so that it is unlikely students would accidentally submit the correct answer to a different question instance purely through calculation errors, to minimise the risk of genuine attempts being unfairly labelled as plagiarism. Ideally, students are unaware that any randomisation is taking place, although it is optimistic to think that students willing to collude will not notice randomisation.

möbius

Preview

Let $X = [-1, 1]$, $f_1(x) = -2x$, $f_2(x) = (5 - x^2)^{1/2}$. Solve the tradeoff model

$$\max_{x \in X} \mu f_1(x) + f_2(x)$$

for the three given values of μ . In each case, give x^* to 2 decimal places.

$$\mu_1 = 0.1$$

$$\mu_2 = 0.25$$

$$\mu_3 = 0.3$$

Grade

How Did I Do?

Refresh

Close

FIG 5. A question from DJ's third-year Game Theory & Multicriteria Decision Making course. Each student is presented with three values for μ , chosen at random from six possible values. For $\mu = 0.05$, the student should return the answer -0.22 . The regrader would raise a concern if a student responded with -0.22 for the value $\mu = 0.25$, if $\mu = 0.05$ was not given to that student.

At the end of each assessment, we included a file upload which required students to submit their hand-written work as a PDF file, worth 5% to encourage students. We also notified students of

our intent to consider handwritten submissions as evidence of their original work and we performed cursory checks to see if they contained sufficient working before awarding the 5% marks. Failing to provide an upload could be penalised more harshly (by scoring the student 0 for the entire assessment), but the intermittent server issues hinder this approach. These file uploads provided reliable evidence to manually check student's work and decide if they have arrived at suspect answers via genuine errors or collusion. If evidence of plagiarism was sufficient, students were referred to the J-BJI academic integrity officer, who would arrange a meeting with the student to discuss the evidence and determine whether plagiarism occurred.

There are potential problems with the file upload approach. Suppose a student is aware that randomisation is taking place in a question and can find another student who has the same instance of this question and submits the same response. In such cases, the file upload could contain no genuine attempt to show the student's work (either through omission, uploading the wrong document, or copying the working from the same student) and would go unnoticed by the regrader if the response is correct. These false negatives are concerning, and it is unclear to what extent they are taking place. The time available to a student to complete an assessment is also a relevant consideration in plagiarism mitigation (refer back to our discussion of logistical constraints in section 2), and question reuse leads to an increase in undetected plagiarism attempts.

There are ways to automatically detect certain types of plagiarism. RL programmed a Python script to extract pages and images from students who submitted handwritten work, hash the resultant files (using the MD5 algorithm) and checked for collisions (images/pages with the same hash). We performed this check at the end of each semester, resulting in a few hundred collisions for all modules from approximately 2500 submissions (a reasonable number to consider as these collisions were exact matches and most were easily dismissed as underlying background images or logos/watermarks) and several penalties were ultimately applied for collusion. RL also experimented with perceptual hashes (Buchner, 2022), which hashed the visual representation, as opposed to digital data, in such a way that visually similar images should have similar hashes. However, this resulted in too many false positives to be of practical use: one particular plagiarism case involved 2 students submitting essentially the same work, with one student having added extra written annotations on top of the other student's work; this was detected using the MD5 method as the annotations were stored as a separate embedded image (the original image was unmodified), but the Hamming distance for the perceptual hash between the original and annotated image was not small enough to make checking collisions up to that distance feasible.

Ultimately the laziest of plagiarists were typically those identified with this method as they either submitted the same file or their PDF files contained shared common parts with other students; many cases involved students sharing identical images in their handwritten work, which whilst not direct evidence of plagiarism, did prompt further investigation. The time invested in manual plagiarism detection could make one question the purpose of CAA since there is still a sizeable logistic advantage compared to marking handwritten work, even with our limited teaching assistant support. In those rare cases where plagiarism was confirmed, a typical outcome was to score the student 0 marks for the corresponding assessment.

6. Reflection and conclusion

Whilst CAA can provide more efficient and effective assessment, in practice there are several obstructions to optimal performance; some are related to the nature of CAA generally or more specifically the Möbius platform we use, whilst others are due to the institutional context and operational constraints we operate within. Long complex proofs have a particular artistic license for how one might construct them, with most applications of CAA to proof-writing questions limited to particular problem types, often related to relatively elementary (in a tertiary-level mathematics context) mathematical material (Sangwin, 2019; Sangwin & Köcher, 2016; Moons & Vandervieren,

2022).

A semi-automated approach to marking that utilises a rubric marking system, such as in (Graide, n.d.), appears to combine the detailed feedback students expect from humans with the efficiencies gained from automated assessment, and our university is currently trialling this software. Proof assistants/interactive theorem provers such as Coq (Coq Team, n.d.) and (Lean Focused Research Organization, n.d.) are used to create formal mathematical proofs and have been used in mathematics undergraduate courses to help teach students proof comprehension and writing skills (Blok & Hoffman, n.d.; Avigad, 2019). The use of such software for summative assessment in a module needs to be carefully considered as students would have to learn a new computer language/system as well as the ordinary mathematical content in that module (Iannone & Thoma, 2023).

Even without using proof assistants, it is still possible to assess proof-adjacent skills within our CAA setup. A typical example is when students are given proof frameworks (Selden, et al., 2018) with sections missing that the student needs to complete with justifications, which in a CAA will be provided by multiple-choice options that require the student to select only those conditions which are necessary. This can help develop student's proof comprehension skills as they do not have to create a whole proof, and in our experience have been effective at highlighting common misconceptions which can guide future study sessions. Integrating proof comprehension questions into the English language provision could help alleviate additional difficulties non-native speakers have in understanding mathematics (Barton, et al., 2005).

Rather than fight against the limitations CAA has in assessing all learning outcomes of our modules, we could instead embrace its advantages in providing practice exercises for students with rapid and efficient feedback by only using it formatively. Although such a move would admittedly require more resources initially to create assessments for each module, especially to create high-quality feedback for each assessment (Gill & Greenhow, 2008), they could easily be reused each year the module is delivered with little modification, assuming there are no issues in the implementation of the questions. Students seem to benefit from performing 'typical' examples multiple times until they gain confidence (Anderson, et al., 2000) and questions which test basic understanding of definitions (such as asking for an example of a binary relation on a small set that is symmetric but neither reflexive nor transitive) seem particularly well-suited to CAA. Using CAA in this way would reduce the burden on academic staff in preparing questions with consistent difficulty across different instances, as well as eliminating plagiarism concerns. STACK can provide response-specific feedback (STACK, n.d.) and the authors believe this would be a very beneficial feature if added to Möbius. Combined with hand-written assessments that focus primarily on higher-order skills and detailed, individual feedback on students' submissions, this approach can combine the strengths of both forms of assessment, and is also advocated for in (Lyakhova, 2023, pp. 4-5).

Whilst adapting to the logistical issues caused by the COVID-19 pandemic, it became clear that students were in desperate need of feedback on their hand-written work so that they could adequately prepare for exams and gain insight into how their lecturers would assess them, such as their proofs in linear algebra or explanations of statistical analysis. The increased difficulty of exams during the 2020/21 academic year made this issue particularly acute as students were expected to write more clearly and coherently than previously, given the open-book nature of these exams. Hand-written assessments were finally reinstated late in the 2020/21 academic year, initially submitted in groups then individually from 2021/22, replacing half of our CAAs so that each form of continuous assessment now contributed 10% towards final module marks. The authors strongly believe that proofs and explanations should be a core part of our mathematics curriculum and the associated skills cannot currently be assessed directly/adequately with fully-automated CAA without students having to also learn to use interactive theorem provers; this is not to say that we view traditional closed-book exams as necessarily the gold-standard for all our modules, and within the school we are investigating an approach in the vein of standards-based grading (Owens, 2015) whereby pre-exam assessments determine whether a student has passed a module, with a final exam

determining their passing grade.

The authors are all intimately familiar with the Möbius platform and confident in their ability to create suitable assessment content, but we recognise that was not the case at first nor is it the case currently for several academics at J-BJI. The prominence of programming varies significantly in undergraduate and postgraduate UK mathematics programmes (Sangwin & O'Toole, 2017), which may partially explain this variance. This is a significant concern given that all J-BJI academics authored CAAs, not just those who had interest/expertise. The perceived limitations of CAA, whether due to preconceptions of its capabilities or lack of experience, results in many colleagues relying upon multiple-choice questions or numerical response boxes almost exclusively. These types of questions have their place as part of our assessment diet but are unable to assess all of our learning outcomes (Sangwin, 2013, pp. 2-3) for a given module, and when heavily relied-upon (such as during the pandemic) can lead to poor differentiation between students, especially when taken as open-book assessments.

At the other end of the spectrum, despite advising against this, some staff expected students to input complicated Maple expressions with insufficient guidance from the lecturer and/or Möbius platform, that resulted in many syntax errors which had to be corrected during regrading. This results in an uneven student experience across modules, however this would be the case for other aspects of their education (for example, there may be differences in the quality of learning materials and delivery). It is unclear whether the 20% difference between the MAM & ICS / Econ & Stats assessments resulted in staff considering the comparable difficulty of questions. STACK, another CAA system, allows instructors to select desired instances of a randomly generated question (STACK, n.d.) to help ensure questions are of comparable difficulty, and again we would appreciate this feature in Möbius, as it would aid our staff in designing questions with greater consistency. Further research is needed on the implementation of CAA at scale, as well as academic's needs and training when learning to develop CAA content is required, as highlighted in questions 32 and 33 of (Kinneer, et al., 2024).

7. Acknowledgements

We would like to thank Jonathan Watkins and Richard Mason for their support over the lifetime of the joint institute and our valuable discussions on this paper and computer-aided assessment generally. We would also like to Yuzhao Wang for giving permission to Figure 4.

REFERENCES

- Anderson, J. R., Reder, L. M. & Simon, H. A., 2000. Applications and Misapplications of Cognitive Psychology to Mathematics Education. *Texas Educational Review*, 1(2), pp. 29-49.
- Armour, K. & Senior Education Team, 2020. *Framework for Educational Resilience 2020/21*. [Online]
Available at: <https://birminghamucu.files.wordpress.com/2020/05/framework-for-educational-resilience-18-may-d6-m3-and-all-staff.pdf>
- Avigad, J., 2019. Learning Logic and Proof with an Interactive Theorem Prover. In: G. Hanna, D. A. Reid & M. de Villiers, eds. *Proof Technology in Mathematics Research and Teaching*. Cham: Springer, pp. 277-290.
- Barton, B. et al., 2005. EAL undergraduates learning mathematics. *International Journal of Mathematical Education in Science and Technology*, 36(7), pp. 721-729.
- Bengtsson, L., 2019. Take-home exams in higher education: a systematic review. *Education Sciences*, 9(4).
- Bickerton, R. T. & Sangwin, C. J., 2022. Practical online assessment of mathematical proof. *International Journal of Mathematical Education in Science and Technology*, 53(10), pp. 2637-2660.
- Blackboard, n.d. *Learning Tools Interoperability (LTI)*. [Online]

Available at:

https://help.blackboard.com/Learn/Administrator/SaaS/Integrations/Learning_Tools_Interoperability

[Accessed 18 April 2024].

Blok, R. J. & Hoffman, C., n.d. *SpatchCoq blog*. [Online]

Available at: <http://spatchcoq.co.uk/spatchcoq/>

[Accessed 13 August 2023].

Broughton, S. J., Robinson, C. L. & Hernandez-Martinez, P., 2013. Lecturers' perspectives on the use of a mathematics-based computer-aided assessment system. *Teaching Mathematics and its Applications*, 32(2), pp. 88-94.

Buchner, J., 2022. *ImageHash*. [Online]

Available at: <https://pypi.org/project/ImageHash/>

[Accessed 13 August 2023].

Comas-Forgas, R., Lancaster, T., Calvo-Sastre, A. & Sureda-Negre, J., 2021. Exam cheating and academic integrity breaches during the COVID-19 pandemic: An analysis of internet search activity in Spain. *Heliyon*, 7(10).

Coq Team, n.d. *The Coq Proof Assistant*. [Online]

Available at: <https://coq.inria.fr/>

[Accessed 7 May 2024].

Davies, B., Crisan, C., Geraniou, E. & Smart, T., 2024. A department-wide transition to a new mode of Computer-Aided Assessment using STACK. *Unpublished*.

Davies, B., Smart, T., Geraniou, E. & Crisan, C., 2022. *STACKification: automating assessments in tertiary mathematics*. Bozen-Bolzano, Free University of Bozen-Bolzano and ERME.

DigitalEd, n.d. *Author a Maple-graded question*. [Online]

Available at: <https://www.digitaled.com/support/help/instructor/Content/INST-AUTHORING/QUESTION-TYPES/Author-Maple-graded-question.htm>

[Accessed 13 August 2023].

DigitalEd, n.d. *Define assignment properties*. [Online]

Available at: <https://www.digitaled.com/support/help/admin/Content/INST-ACTIVITIES/ASSIGNMENTS/Define-assignment-properties.htm>

[Accessed 18 April 2024].

DigitalEd, n.d. *Möbius*. [Online]

Available at: <https://www.digitaled.com/mobius/>

[Accessed 13 August 2023].

DigitalEd, n.d. *Möbius Community Forum*. [Online]

Available at: <https://web.archive.org/web/20220819204435/https://mobiuscommunity.com/>

[Accessed 19 August 2022].

DigitalEd, n.d. *View Möbius system requirements*. [Online]

Available at: <https://www.digitaled.com/support/help/student/Content/GENERAL/System-Requirements.htm>

[Accessed 13 August 2023].

D'Souza, S. M. & Wood, N. L., 2003. *Tertiary students' views about group work in mathematics*. Auckland, Australian Association for Research in Education.

Du Croz, R. & Morris, N., 2024. *Developing formal summer school and study abroad options within a dual degree (4+0) programme*, London: QAA.

Foreign Affairs Office of the People's Government of Guangzhou Municipality, 2020. *广州发布第三号防疫通告：全市小区实施封闭管理*. [Online]

Available at: http://www.gzfao.gov.cn/ztlm/yqfk/content/post_137139.html

[Accessed 3 July 2024].

Gill, M. & Greenhow, M., 2008. How effective is feedback in Computer-Aided Assessments?. *Learning, Media and Technology*, 33(3), pp. 207-220.

Graide, n.d. *Organising Feedback*. [Online]

Available at: <https://www.graide.co.uk/features/organising-feedback>
[Accessed 3 July 2024].

Greenhow, M., 2015. Effective computer-aided assessment of mathematics; principles, practice and results. *Teaching Mathematics and Its Applications: International Journal of the IMA*, 34(3), pp. 117-137.

HESA, 2023. *Higher Education Student Statistics: UK, 2021/22*, Cheltenham: Higher Education Statistics Agency.

Iannone, P. & Thoma, A., 2023. Interactive theorem provers for university mathematics: an exploratory study of students' perceptions. *International Journal of Mathematical Education in Science and Technology*.

Institute of mathematics & its applications, London Mathematical Society, Royal Statistical Society, 2022. *Statement on Methods of Assessment in the Mathematical Sciences*. [Online]
Available at: <https://www.lms.ac.uk/news/assessment-statement-update>
[Accessed 24 May 2024].

Johnson, T. et al., n.d. *Maple TA 2017*. [Online]
Available at: <https://mapletabham2017.wordpress.com/>
[Accessed 24 May 2024].

Jones, D., Meyer, J. & Huang, J., 2021. Reflections on remote teaching. *MSOR Connections*, 19(1), pp. 47-54.

Kinnear, G. et al., 2024. A Collaboratively-Derived Research Agenda. *International Journal of Research in Undergraduate Mathematics Education*, Volume 10, pp. 201-231.

Lancaster, T. & Cotarlan, C., 2021. Contract cheating by STEM students through a file sharing website: a Covid-19 pandemic perspective. *International Journal for Educational Integrity*, 17(3).

Lean Focused Research Organization, n.d. *Lean*. [Online]
Available at: <https://leanprover.github.io/>
[Accessed 13 August 2023].

Lowe, T. W. & Mestel, B. D., 2020. Using STACK to support student learning at masters level: a case study. *Teaching Mathematics and its Applications*, 39(2), pp. 61-70.

Lyakhova, S., 2023. *On the use of technology in university mathematics teaching and assessment in STEM degree schemes*, London: Joint Mathematical Council of the UK.

Meyer, J. C. & Leek, R., 2020. *E-assessment for the Jinan University – University of Birmingham joint institute: From content development to assignment re-grading*. Newcastle: E-Assessment in Mathematical Sciences.

Moons, F. & Vandervieren, E., 2022. *Handwritten math exams with multiple assessors: researching the added value of semi-automated assessment with atomic feedback*. Bozen-Bolzano, Free University of Bozen-Bolzano and ERME.

OpenAI, 2022. *Introducing ChatGPT*. [Online]
Available at: <https://openai.com/index/chatgpt/>
[Accessed 15 June 2024].

Owens, K., 2015. *A Beginner's Guide to Standards Based Grading*. [Online]
Available at: <https://blogs.ams.org/matheducation/2015/11/20/a-beginners-guide-to-standards-based-grading/>
[Accessed 24 May 2024].

Pointon, A. & Sangwin, C. J., 2003. An analysis of undergraduate core material in the light of hand-held computer algebra systems. *International Journal of Mathematical Education in Science and Technology*, 34(5), pp. 671-686.

QAA, 2020. *COVID-19: Initial Guidance for Higher Education Providers on Standards and Quality*. Gloucester: QAA.

QAA, 2024. *Quality Evaluation and Enhancement of UK Transnational Higher Education: People's Republic of China*, Gloucester: QAA.

- Robinson, C. L., Hernandez-Martinez, P. & Broughton, S., 2012. Mathematics lecturers' practice and perception of computer-aided assessment. In: P. Iannone & A. Simpson, eds. *Mapping university mathematics assessment practices*. Norwich: University of East Anglia, pp. 105-118.
- Sangwin, C., 2013. *Computer Aided Assessment of Mathematics*. Oxford: Oxford University Press.
- Sangwin, C., 2019. *Developing and evaluating an online linear algebra examination for university mathematics*. Utrecht, Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.
- Sangwin, C. J. & Köcher, N., 2016. Automation of mathematics examinations. *Computers & Education*, Volume 94, pp. 215-227.
- Sangwin, C. J. & O'Toole, C., 2017. Computer programming in the UK undergraduate mathematics curriculum. *International Journal of Mathematical Education in Science and Technology*, 48(8), pp. 1133-1152.
- Selden, A., Selden, J. & Benkhalti, A., 2018. Proof Frameworks: A Way to Get Started. *PRIMUS*, 28(1), pp. 31-45.
- Smith, G. et al., 1996. Constructing mathematical examinations to assess a range of knowledge and skills. *International Journal of Mathematical Education in Science and Technology*, 27(1), pp. 65-77.
- STACK, n.d. *Authoring quick start 3: improving feedback*. [Online]
Available at: https://docs.stack-assessment.org/en/AbInitio/Authoring_quick_start_3/
[Accessed 14 August 2023].
- STACK, n.d. *Authoring quick start 4: randomisation*. [Online]
Available at: https://docs.stack-assessment.org/en/AbInitio/Authoring_quick_start_4/
[Accessed 14 August 2023].
- Unicode, 2022. *C0 Controls and Basic Latin*. [Online]
Available at: <https://www.unicode.org/charts/PDF/U0000.pdf>
[Accessed 13 August 2023].
- Unicode, 2022. *CJK Symbols and Punctuation*. [Online]
Available at: <https://www.unicode.org/charts/PDF/U3000.pdf>
[Accessed 13 August 2023].
- Unicode, 2022. *Halfwidth and Fullwidth Forms*. [Online]
Available at: <https://www.unicode.org/charts/PDF/UFF00.pdf>
[Accessed 13 August 2023].
- Universities UK International, 2023. *The scale of UK HE TNE 2021-22*, London: Universities UK.
- University of Birmingham, n.d. *Jinan-Birmingham dual degree undergraduate programmes*. [Online]
Available at: <https://www.birmingham.ac.uk/schools/mathematics/ug-admissions/jinan/programme.aspx>
[Accessed 16 May 2024].
- Watkins, J., 2018. *Chained / Linked Response Areas*. [Online]
Available at: <https://web.archive.org/web/20210731061806/https://ideas.digitaled.com/ideas/M-I-35>
[Accessed 31 July 2021].
- Watkins, J., 2018. *Syntax Checker for student input (1D Math/plain text)*. [Online]
Available at: <https://web.archive.org/web/20210507134859/https://ideas.digitaled.com/ideas/M-I-32>
[Accessed 7 May 2021].
- Wiggins, G., 2011. A True Test: Toward More Authentic and Equitable Assessment. *Phi Delta Kappan*, 92(7), pp. 81-93.
- Wiliam, D., 2011. What is assessment for learning?. *Studies in educational evaluation*, 37(1), pp. 3-14.