

Don't throw the student out with the bathwater: online assessment strategies your class wont hate

Stuart Johnson^a, John Maclean^a, Raymond Vozzo^a, Adrian Koerber^a, Melissa A. Humphries^a

^aThe University of Adelaide, South Australia, Australia

ARTICLE HISTORY

Compiled October 25, 2021

ABSTRACT

STEM educators have taken on a new and radically different reality: that students will have access to online tools and online knowledge as they undergo assessment. Students may look up facts that are easily available online, employ online tools, and also employ paid services to cheat.

The first wave of novel assessment strategies devised in response to this new reality were online invigilation measures that were often either ineffective, or widely hated by students. This paper considers two frameworks for effective online assessment in the modern era: a framework for question design in STEM, and a framework for the structure of assessment. The framework for question design discusses methods to discourage the use of online tools and calculators to simply obtain answers, and methods to detect contract cheating. The framework for assessment structure discourages cheating while managing to increase student satisfaction and engagement with content: we present a case study from a cohort of 113 first-year University students of adapted assessment strategies with no identified paid cheating, well-differentiated student results and large-scale positive student feedback.

Our key contribution is to identify simple and effective steps to assess students fairly in an online environment, without empowering or motivating them to cheat.

KEYWORDS

Assessment, examinations, STEM, contract cheating, online assessment

1. Introduction

University STEM courses have traditionally placed much of the weighting of the course grade on one or two major exams. This is largely motivated by the nature of closed-book, invigilated exams: instructors can be more-or-less certain that the work being assessed has been produced by the student whose grade is being determined. One of the biggest impacts that COVID-19 has had on STEM courses at university has been the loss of this method of assessment. The standard method for assessment in 2020 (and beyond) at the University of Adelaide has been *uninvigilated*, open-book, online examinations that present challenges in ensuring that students are doing their own work¹, and in writing assessments that are not trivial to answer with online tools and resources.

Contact: Melissa Humphries. Email: melissa.humphries@adelaide.edu.au

¹[17] discusses the modern forms of academic misconduct, distinguishing plagiarism, copying, collusion, and contract cheating

The shift to online assessment has measurably worsened some components of academic honesty: contract cheating has increased. Usage of Chegg, for example, increased threefold from March 2019 to March 2020² [12]. Online proctoring has been taken up as a solution and has drastically increased since 2019 [9]. Although online proctoring may seem like the perfect solution, there is some evidence that assessment design is more significant than proctoring [14]. Additionally, issues have been raised with online proctoring; students often find it invasive³, and there are concerns about data privacy [2, 18].

However cheating was not always a higher risk in online assessment: late 2000’s research reported in several cases that the prevalence of cheating in online assessment was roughly equal to cheating in face to face assessment [19]. A meta-analysis of 79 papers studying the motivations behind academic honesty and academic dishonesty finds that students who perceive assessment structure as based on developing mastery of particular skills, and who are convinced that the course is useful, are less likely to cheat [10]. This broader perspective on student honesty reveals an opportunity to reverse the arms race, of student cheating vs attempts to eliminate the possibility, described above.

This manuscript is primarily concerned with the methods and philosophy of non-intrusive assessment strategies that empower educators to assess student learning. We present a suite of solutions, with examples, to redesign course structure and/or content aimed at maximising student learning and course satisfaction, decreasing cheating and maintaining differentiability of results.

2. Background

Two case studies were carried out at the University of Adelaide. In the first study, the types of assessment question asked in three first-year mathematics courses were altered to preserve their ability to test student knowledge in an open-book format. In the second study, of a first-year statistics course, the structure of assessment was altered with the same goal. We report the goals for alterations to the mathematics courses in section 2.1 and the alterations themselves—with diverse examples—in section 3.1. The goals for the statistics case study are reported in section 2.2, and the alterations described in section 3.2. Results of these alterations are the topic of section 4, and we conclude in section 5.

2.1. Question structure

The first year mathematics courses required significant redesign of exam question style to adapt to online exams. In addition to being able to access assignment help sites, students would also have access to their course notes and to online mathematical solvers. A review of a 2019 first year mathematics exam revealed that 65% of the marks could be obtained using WolframAlpha alone.

Open resource exams in mathematics appear to be rare prior to 2020. Examples include service teaching for specific applications [8], or involve further forms of assessment such as an oral exam [1]. Open resource exams with sophisticated computer

²While Chegg markets itself as a site for “homework help”, it is commonly used by students to obtain test or assignment solutions.

³One report finds seventeen major petitions against their use [7]

algebra capabilities available to students were studied when CAS calculators first became available, see [16], but their use does not seem to have been widely adopted. A more recent study, [13], was open resource but employed invigilation to prevent unauthorised communication.

Section 3.1 reports redesigned question strategies employed—on short notice—in 2020 to mitigate the problems with open resource exams.

2.2. Assessment structure

The first year statistics course changed from a single, 70% exam to a series of assessments that were designed to circumvent the issues itemised above. Using principles of continual assessment (see recommendations, and warnings, in [4, 6]) to increase engagement with the course and retention of concepts, the course assessment was split into two broad categories:

- (1) Low-weight, collaborative assessment and
- (2) Higher-weight, independent assessment.

The low-weight assessment was used at regular intervals to encourage both engagement with the course and engagement with their peers. This aimed to build a sense of community, and extend learning by encouraging safe and careful exploration of difficult concepts. To differentiate marks on individual understanding and application, higher-weight assessment was used. These were staggered across the semester to provide regular points of consolidation of course concepts, and limited in time to reduce the opportunity to cheat [15].

3. Methods

3.1. Question design for assessing students with access to online mathematical solvers

Below we will illustrate some of the strategies used in the mathematics courses, referring to examples of questions used prior to and during 2020 which are given in Appendix A. Crucially, the questions test similar skills and knowledge and are of a similar difficulty.

3.1.1. Graphical presentation

As students would have graphing tools available to them, we switched from asking for graphs or answers which would be found by graphing, to testing the same skills by asking students to interpret graphs.

An example showing an old and new question created with this approach is given in Figure A1.

Figure A2 shows a question on qualitative methods for differential equations where the function is given graphically rather than as a formula so that they cannot simply use online tools to numerically solve the equation.

3.1.2. Partial working for questions that cannot be solved using online tools

Typically, an exam question that is too complicated for an online tool is also too complicated for a student. However it is possible to test very specific aspects of a method

on such a problem. For example, Figure A3 demonstrates an online-appropriate integration by substitution, on an integral that is not possible to calculate in closed form. Thus an online calculator will give up and give a numerical answer, yet it is possible for a student to demonstrate basic knowledge of the method of substitution.

3.1.3. Giving incomplete information

Figure A4 shows that the methods of solving a DE may be assessed without giving an equation which can be solved online by instead giving the integrating factor, making the question a straightforward test of the method for those who understand, yet impossible using an online calculator. Figure A5 shows contrasting approaches in testing understanding of results on eigenvectors and diagonalisation, without simply giving an explicit matrix to diagonalise.

3.1.4. Working backwards

For some problems it is possible to start with the solution and ask for the problem, for example asking for a differential equation with a given solution, see Figure A6.

3.1.5. Asking for working or specific methods

Figure A7 shows polynomial division questions from 2019 and 2020, respectively. Since this is very standard and can easily be done by online tools, asking for a specific layout can help to determine if the student has studied the course materials.

3.1.6. Question design to deter and detect contract cheating

The first year maths exams included both a written and a computer entered (Mobius) section.

The chief strategies were:

Use of distinctive wording. Many common questions are online already, and may already be solved there, so detection of cheating is much harder. Including local references can be effective: for example instead of “A factory produces ...” we used “A South Australian factory produces ...”.

Referencing specific details of the course. To reduce the ability to get answers from contract cheating websites specific course references were used, for example “Use Theorem 15.3 to explain why ...” or “You must set out your calculations in the way used in lectures”.

Randomisation. Using Mobius we have been able to randomise questions to the extent that students receive a unique version. Any upload to a website can then be uniquely matched to a student.

The question in Figure A8 has randomised matrix entries, including some that do not affect the solution but aid in randomisation for detection purposes.

The question in Figure A9 is more theoretical (which can be harder to randomise), but in this case the parent question contains over 200 million versions. We randomised the statements given and the parameters contained within them. In order to efficiently match any question on Chegg to the student that had posted it, we included hidden text in the question that lists the parameter values in html source. When a copy of the question was posted online, we deduced the list from the parameter values in the question, and then used the list to immediately find the matching exam question and prove that only one student had been given that precise question.

3.2. Assessment Design

As an alternative to retaining a single, high weight assessment, a first-year statistics course of 113 first-year Mathematics students sought to address assessment issue by changing the assessment structure.

The course had previously had 5 assignments and a 70% final invigilated exam. This was adapted to embrace continual assessment; utilising quizzes, statistical written reports and smaller, regular maths tests across the semester. Assessment was split into low-weight, collaborative components that formed 30% of the total grade while the remaining majority of students marks came from independent assessment pieces; major quizzes, statistical reports and maths tests.

3.2.1. Low-Weight, collaborative assessment

There were two low-weight assessment styles used throughout the course, aimed at encouraging engagement with the material, providing opportunities to practice learning, and increasing collaboration, fostering a sense of community. The first of these styles was weekly, low-weight quizzes to incentivize regular engagement and provide a safe environment for testing learning. Timing was generous to enable students time to look through their notes or collaborate to find answers. The questions were heavily randomised using the `exams` package in R, so that even with collaboration, students would need to apply the methods learned to their own scenario.

The second low-weight assessment was special assignments that students were encouraged to do together. Due at three points across semester, the assignments included questions that most students would be incapable of answering alone. The questions were designed to encourage unique thought and exploration and required collaboration and research to achieve a result. These special assignments led students through proofs and examples that underpin some of the more difficult concepts in the course, providing depth of understanding.

3.2.2. High-weight, independent assessment

Major quizzes, similar in content to the low-weight quizzes, were held at mid and end of semester. In order to reduce collaboration, collusion, and contract cheating, the major quizzes were randomised using R/`exams` and had a reduced available completion time relative to the number of questions being asked [15].

Restricted timing was also utilised for the three maths tests, run in class, across the semester. These tests were semi-invigilated, open-book tests that ran for both face to face and online only cohorts. For the online cohorts, the test was completed via zoom. Students showed their student ID and were requested to keep their video on during the test. The cohort was split into break out rooms of no more than ten students with a tutor monitoring each break out room. As the tests were run in class, students had 50 minutes to complete them, reducing their ability to apply to contract cheating sites for answers. Students were primed to expect questions based on the topic videos from the course, increasing the likelihood that students went first to the course notes or videos for information, than using other online sources.

The final high-weight assessment pieces were three written reports and a prac test, based on statistical analysis completed in prac classes. Reports consisted of a technical section, detailing their statistical analysis, and an executive summary to communicate their findings. As this was a worded assignment, it could be run through a plagiarism checker, encouraging students to complete their own report even if they shared analy-

Subject	2019		2020	
	Mean (<i>sd</i>)	n	Mean (<i>sd</i>)	n
Statistics Exam	69.9 (<i>21.16</i>)	106	70.6 (<i>22.19</i>)	104
Introductory Mathematics	48.7 (<i>22.67</i>)	88	50.2 (<i>22.10</i>)	76
Intermediate Mathematics	59.1 (<i>17.29</i>)	351	57.1 (<i>21.10</i>)	320
Advanced Mathematics	60.2 (<i>20.19</i>)	475	60.7 (<i>18.71</i>)	555

Table 1. Means, standard deviations (*sd*) and class sizes (*n*) for the 2020 statistics course and the 2020 mathematics exams discussed in this paper, alongside their 2019 counterparts. All students with valid grades were counted. It is clear that there is a good distribution of grades across all courses and that there is little difference in students outcomes between 2019 and 2020, despite the stresses associated with lockdown and changing assessment.

ses. To test for students' capability in running analyses, the report was supplemented with a final prac test (in class). Students were given a question related to their current analysis to answer and incorporate into their report within 50 minutes. This tested computational ability in a time pressured environment but with data, and a project, they were familiar with. Using this technique reduces the desire to cheat by building on their current work [11].

4. Results

We report three areas,

- (1) Whether the course was able to differentiate ability between students,
- (2) student satisfaction,
- (3) the ability to detect cheating.

4.1. Ability to differentiate

One of the measures of success for a course is its ability to differentiate between the ability and understanding of students. Table 1 shows the mean and standard deviation for the 2020 courses discussed in this manuscript, along with their 2019 counterparts for comparison. A good distribution of grades was maintained across all courses, evidenced by the standard deviations around 23 marks.

Comparing the results in 2020 to those in 2019 gives a measure of consistency of outcomes across the two years. Using 2-sample t-tests (equal variance not assumed), we found no significant difference between the 2019 and 2020 grades for Statistics ($t(207) = 0.25$, $p = .803$), or between any of the reconfigured exams for Introductory Mathematics ($t(160) = 0.45$, $p = .656$), Intermediate Mathematics ($t(618) = 1.35$, $p = .178$), or Advanced Mathematics ($t(976) = 0.44$, $p = .658$). In these calculations a value of p larger than 0.05 indicates that there is no observable difference between the years, and hence there is no evidence that the changes to assessment structure changed student outcomes. In other words, the radical alterations in assessment delivery did not alter the difficulty of the assessment or the differentiability of the student cohort.

4.2. Student feedback

At the University of Adelaide, students are offered the opportunity to complete a student evaluation questionnaire at the completion of each course. Questionnaire responses are recorded before the exams are taken; student feedback is therefore irrelevant when considering assessment design in the mathematics courses, for which the design of exams was the key focus. However the continual assessment structure of the statistics course, laid out in section 3.2, meant that all assessment had been concluded by the time of the questionnaire, and we therefore report on student feedback.

For the first-year statistics course, 95% of 38 respondents broadly agreed with the statement, “*The assessment tasks in this course help me learn*”. With a maximum score of 7, this course scored a mean of 6.3 and median of 7. This was a higher mean than both the School of Mathematical sciences (6.0) and the University in general (5.8). In addition to the Likert responses, 28 individuals gave positive written feedback on the course and, of those, 22 ($\sim 79\%$) mentioned their appreciation of the adapted assessment.

As an example of the written student feedback, one student said:

“The ability to have three tests throughout the semester allowed for me to grasp concepts better due to the constant revision as well as taking off some major stress of an end of year exam. Because the weighting of each of these is still high, they carried significance but the knowledge that bombing one didn’t mean you would necessarily fail was a weight off of my shoulders”

Other comments included an appreciation that they were, “...forced to keep up with the work as we go” and a student, “...was able to really absorb the material as opposed to cramming.”, providing evidence that the goals of the changed assessment were met in this dimension.

4.3. Detection of integrity breaches

The mathematics course sought to reduce the opportunity to cheat, discussed in section 3.1. As discussed in section 3.1.6, moreover, the maths course included randomisation of questions, and instances of contract cheating could be matched to the precise student who had uploaded the question.

The statistics course included some of the strategies described in section 3.1 but, as described in section 3.2, extensively altered assessment design to focus on repeated, short-duration assessment. For this course Turnitin found no examples of plagiarism across the semester. Regular checks across known contract cheating sites returned only one posted question (of one special assignment - low-weight, group work) that went unanswered.

5. Conclusion

We have discussed two complementary approaches to course (re)design for STEM educators in a world of online assessment. Both approaches were motivated by the need to alter assessments without dramatically changing the difficulty of questions or the skills being tested.

It is fair to mention here that it is not always a realistic goal to eliminate cheating in an online course. Primary motivations for students to cheat relate to extrinsic and intrinsic motivations and goal structures (see [10] for a meta-analysis), and some of

these factors will always apply to certain student cohorts; for example, [3, 5] discuss perspectives of engineering students regarding mathematics. Nonetheless, we highlight the following key points.

Takeaways

- When carefully implemented, continual assessment across the semester results in good differentiability of student performance along with increased student satisfaction.
- With careful question design, it is largely possible to redesign mathematics exams to test course material in a manner suitable for the open resource environment.

Data sharing: The University of Adelaide does not require specific ethics approval to share the pooled student results, or anonymous student feedback, reported in this paper.

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

- [1] I Cnop and F Grandsard. An open-book exam for non-mathematics majors. *International Journal of Mathematical Education in Science and Technology*, 25(1):125–130, 1994.
- [2] Colleen Flaherty. Online proctoring is surging during COVID-19. <https://www.insidehighered.com/news/2020/05/11/online-proctoring-surging-during-covid-19>.
- [3] Jennifer Flegg, Dann Mallet, and Mandy Lupton. Students’ perceptions of the relevance of mathematics in engineering. *International Journal of Mathematical Education in Science and Technology*, 43(6):717–732, September 2012. Publisher: Taylor & Francis .eprint: <https://doi.org/10.1080/0020739X.2011.644333>.
- [4] Tony Harland, Angela McLean, Rob Wass, Ellen Miller, and Kwong Nui Sim. An assessment arms race and its fallout: high-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4):528–541, May 2015. Publisher: Routledge .eprint: <https://doi.org/10.1080/02602938.2014.931927>.
- [5] Diane Harris, Laura Black, Paul Hernandez-Martinez, Birgit Pepin, Julian Williams, and with the TransMaths Team. Mathematics and its value for engineering students: what are the implications for teaching? *International Journal of Mathematical Education in Science and Technology*, 46(3):321–336, April 2015. Publisher: Taylor & Francis .eprint: <https://doi.org/10.1080/0020739X.2014.979893>.
- [6] Rosario Hernandez. Does continuous assessment in higher education support student learning? *Higher Education*, 64(4):489–502, October 2012.
- [7] Jason Kelley. Students Are Pushing Back Against Proctoring Surveillance Apps, September 2020.
- [8] R. Nazim Khan. Assessments: an open and closed case. *International Journal of Mathematical Education in Science and Technology*, 46(7):1061–1074, 2015.
- [9] R Kimmons and G Veletsianos. Proctoring software in higher ed: Prevalence and patterns. *EDUCAUSE Review*. <https://er.educause.edu/articles/2021/2/proctoring-software-in-higher-ed-prevalence-and-patterns>, 2021.
- [10] Megan R. Krou, Carlton J. Fong, and Meagan A. Hoff. Achievement Motivation and Academic Dishonesty: A Meta-Analytic Investigation. *Educational Psychology Review*, 33(2):427–458, June 2021.

- [11] Thomas Lancaster and Robert Clarke. *Contract Cheating: The Outsourcing of Assessed Student Work*, pages 639–654. Springer Singapore, Singapore, 2016.
- [12] Thomas Lancaster and Codrin Cotarlan. Contract cheating by STEM students through a file sharing website: a Covid-19 pandemic perspective. *International Journal for Educational Integrity*, 17(1):1–16, 2021.
- [13] Wes Maciejewski. Let your students cheat on exams. *PRIMUS*, 31(6):685–697, 2021.
- [14] Harvey Mellar, Roumiana Peytcheva-Forsyth, Serpil Kocdar, Abdulkadir Karadeniz, and Blagovesna Yovkova. Addressing cheating in e-assessment using student authentication and authorship checking systems: teachers perspectives. *International Journal for Educational Integrity*, 14(1):1–21, December 2018. Number: 1 Publisher: BioMed Central.
- [15] Timothy Michael and Melissa Williams. Student equity: Discouraging cheating in online courses. *Administrative Issues Journal*, 3:30, 10 2013.
- [16] David Pountney, Carl Leinbach, and Terence Etchells. The issue of appropriate assessment in the presence of a CAS. *International Journal of Mathematical Education in Science and Technology*, 33(1):15–36, 2002.
- [17] Katherine A. Seaton. Laying groundwork for an understanding of academic integrity in mathematics tasks. *International Journal of Mathematical Education in Science and Technology*, 50(7):1063–1072, 2019.
- [18] Bonnie Stewart. Why Higher Ed Needs Data Ethics | Inside Higher Ed.
- [19] George R Watson and James Sottile. Cheating in the Digital Age: Do Students Cheat More in Online Courses? *Online Journal of Distance Learning Administration*, 13(1):14–23, 2010.

Appendix A. Example Exam Questions

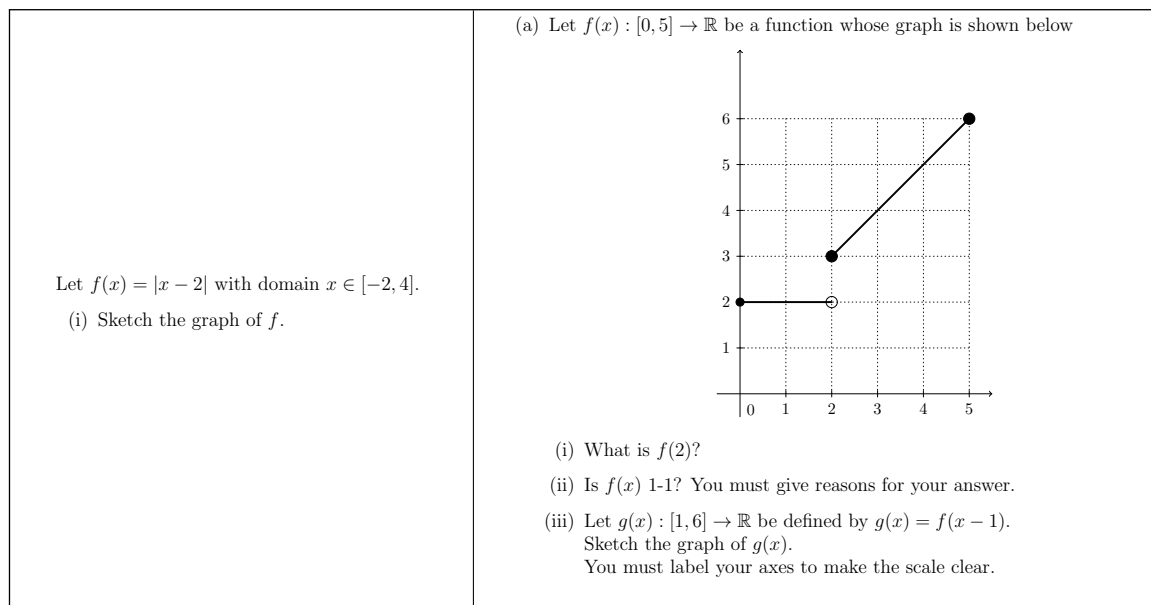


Figure A1. Question on real-valued functions. *Left:* 2019 question, easily solved with diverse online tools. *Right:* 2020 question.

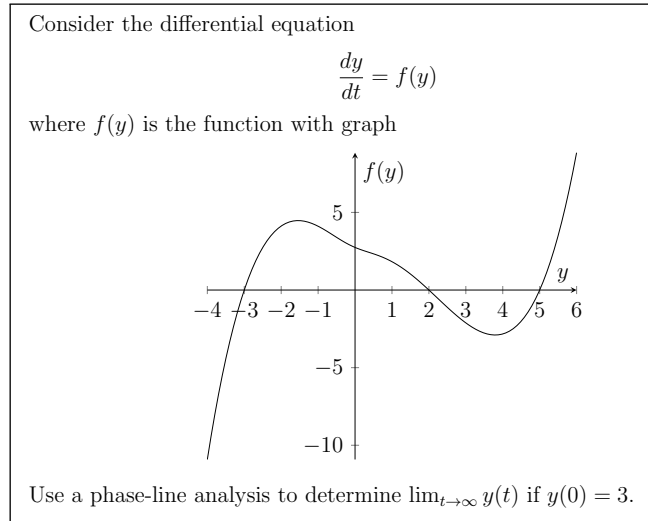


Figure A2. 2020 question on differential equations.

<p>Make the substitution $u = 1 + x^2$ to evaluate $\int \frac{x}{1+x^2} dx$.</p>	<p>Consider the integral $\int_{\frac{\pi}{2}}^{\pi} \sin(2x)e^{\sin^2(x)} dx$.</p> <p>Use the substitution $u = \sin(x)$ to formulate this as an integral in terms of u. (Do not attempt to evaluate the integral.)</p>
---	--

Figure A3. Question on integration by substitution. *Left:* 2019 question, easily answered (with working) by WolframAlpha. *Right:* 2020 question.

<p>Solve the separable differential equation $\frac{dy}{dx} = xy^2$, with $y(1) = -1$.</p>	<p>Consider the initial value problem</p> $\frac{dy}{dt} + p(t)y = \cos(e^t), \quad y(\ln(\pi/2)) = 1.$ <p>You are told that an integrating factor for the DE is $\mu(t) = e^t$. Use this information to find the solution of the initial value problem.</p>
--	---

Figure A4. Question on differential equations. *Left:* 2019 question, easily answered (with working) by WolframAlpha. *Right:* 2020 question.

<p>A 3×3 matrix B has three eigenvalues 1, 2, 3. The eigenspaces are</p> $\mathbb{E}_1 = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right\}, \quad \mathbb{E}_2 = \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}, \quad \mathbb{E}_3 = \text{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right\}$ <p>Write down a matrix P such that $P^{-1}BP = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.</p>	<p>A 3×3 real symmetric matrix C has two distinct eigenvalues. Suppose one of the eigenvalues is known to be $\lambda = 2$. If $\text{tr}(C) = 7$, then there are two possible values for the second eigenvalue. Find them, giving reasons.</p>
--	--

Figure A5. Two examples of incomplete information. *Left:* on diagonalisation of a matrix. *Right:* on eigenvalue calculations.

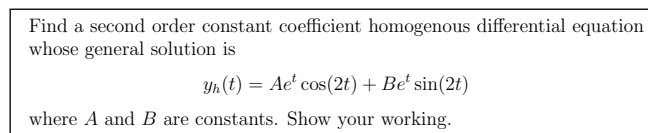


Figure A6. 2020 question asking for a DE rather than the solution

<p>Let $p(x) = 2x^3 + x - 1$.</p> <p>10(a). Use polynomial long division to divide $p(x)$ by $x + 2$.</p>	<p>Use polynomial long division to divide $x^3 + x + 7$ by $x^2 - 2x$, to find the quotient and remainder. You must set out your calculation in the way used in lectures.</p>
--	---

Figure A7. Polynomial division. *Left:* 2019 question, for which many correct solutions can be searched online. *Right:* 2020 question.

Let $A = \begin{bmatrix} 1 & 20 & -19 & 11 \\ 0 & 0 & 0 & -38 \\ 0 & 0 & 0 & -19 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Select from the following all vectors which are in $\text{Nul}(A)$:

$(-1, -1, 1, 0)^t$

$(-1, 1, 1, 0)^t$

$(0, 38, 40, 0)^t$

$(-11, 0, 0, 1)^t$

Figure A8. A randomised Mobius question with over 14 million versions (see discussion in section 3.1.6).

Select from the following statements about eigenvalues, eigenvectors and diagonalisation, all of those which are **TRUE**

All 4×4 matrices with characteristic polynomial $p(\lambda) = (\lambda - 1)(\lambda - 5)(\lambda - 8)^2$ are diagonalisable

If a diagonalisable matrix has characteristic polynomial $p(\lambda) = (\lambda - 1)^2(\lambda - 2)(\lambda + 3)^2(\lambda + 5)$, then given non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{E}_2$, the set $\{\mathbf{u}, \mathbf{v}\}$ must be linearly dependent

Every 4×4 matrix with characteristic polynomial $p(\lambda) = (\lambda - 12)^4$ is diagonalisable

If A is a 6×6 matrix with characteristic polynomial $p(\lambda) = \lambda^2(\lambda - 7)^4$ then $\det(A) = 0$

Figure A9. A randomised Mobius question addressing theory.