Using e-assessment for interactive example-generation tasks

George Kinnear[1] and Igor' Kontorovich[2]

[1]The University of Edinburgh, UK; G.Kinnear@ed.ac.uk

[2]The University of Auckland, New Zealand; i.kontorovich@auckland.ac.nz

The paper concerns the use of e-assessment systems to advance students' example spaces of mathematical concepts. We report on a pilot study with seven first-year students of linear algebra. Three students engaged with the static version of the e-task, which asked them to generate three examples that were as different as possible. The remaining students were prompted by an interactive e-task that assessed the provided examples and asked for another one with different properties. The analysis of a single task about matrices showed that the students who worked with the interactive e-task were more successful than those working with the static e-task, in terms of the number and the range of generated examples. These preliminary findings open the doors to further study of the design and use of interactive example-generation tasks in university mathematics education.

Keywords: Assessment, Concepts, Feedback, Learner-generated examples.

Introduction

Example-generation tasks have been suggested as an effective way to promote students' learning about concepts (Watson & Mason, 2005), and as a way for researchers (and teachers) to gain insight into students' understanding of concepts (Zazkis & Leikin, 2007). E-assessment offers the potential to provide large undergraduate classes with formative example-generation tasks, giving automated feedback on a scale that would not be feasible for teachers to do manually (Sangwin, 2003). Such formative tasks can serve a dual purpose: on the one hand, students' responses provide the teacher with assessment information about their students' knowledge of concepts; on the other hand, the tasks can also prompt students to consider examples that they might not otherwise think about, thereby promoting further learning about the concepts. The design of e-assessment tasks that expand students' understanding of concepts and the capability to generate concept examples has been identified as an open question that is particularly of interest at the undergraduate level (Kinnear et al., 2022).

Fahlgren and Brunström (2023) note the potential of e-assessment tasks that give feedback in the form of prompts for further examples, with the prompts depending on the examples given so far by the student. Such an approach would automate the recommended approach in clinical interview settings, of "asking for 'another and another' [example] and for 'something different'" (Zazkis & Leikin, 2007, p. 19). Prompting in this way may both stimulate the learners to consider examples beyond the most immediately obvious ones and provide richer data about the learners' knowledge.

Here we describe the design and evaluation of interactive example-generation tasks, that prompt students for further examples based on the examples they have given so far, implemented as prototypes in an e-assessment system. The tasks address topics in linear algebra, and were devised and piloted in collaboration with a group of fourth-year students as part of their undergraduate research project. Our overarching aim was to develop e-assessment tasks that prompt students to generate a rich range of examples. We were particularly interested in the effect of interactivity on students’ example-generation activity, so we developed static versions of the tasks to serve as a comparison. The research question guiding this study was: *how do the examples produced by students compare between static and interactive e-assessment tasks?*

Theoretical background

A central theoretical notion of this study is an *example space* (Watson & Mason, 2005), which is a collection of possible examples of a mathematical concept. Watson and Mason define a conventional example space for a given concept as that “generally understood by mathematicians” (p. 62). Each individual has their own personal example space, based on their past experience, and will access it in different ways depending on the situation (e.g., in response to particular cues in a task). The structure of example spaces can be described by the *dimensions of possible variation* (DofPV), which are the features of examples that can vary, and their associated *range of permissible change* (RofPCh). For instance, if asked for a quadratic polynomial, one DofPV is the coefficient of x^2, for which the RofPCh is any non-zero real number. Fahlgren and Brunström (2023) used these notions to analyse three example-generation tasks, and we similarly used them to guide the design and analysis of our tasks.

Zazkis and Leikin (2007) proposed a framework for characterizing students’ example spaces, that includes a focus on *accessibility* of the examples: what are the most obvious concept examples, and how readily can students generate examples beyond those. Watson and Mason (2005) offer the metaphor of “example space as larder”, where finding an example can be thought of as “either immediately picking out something familiar or having to look for it for a while” (p. 162). According to the *principle of intellectual parsimony*, “when solving a problem, one intends not to make more intellectual effort than the minimum needed” (Koichu, 2010, p. 217), so it may be that students will not consider less accessible examples without explicit prompting.

Watson and Mason (2005) give advice about designing example-generation tasks that prompt students for a range of examples. They suggest asking for a sequence of examples satisfying additional constraints, as the “increasing constraints extend awareness of what is possible” (p. 132). In this way, example-generation tasks can help to draw learners’ attention to DofPV (or to the extent of the associated RofPCh) that they were not previously aware of. Arzarello et al. (2011) further argue that helping students to see structure in example spaces through this sort of prompting is a key role for teachers. Our study begins to explore how to prompt students in this way using e-assessment.

Method

To investigate the potential for interactive example-generation tasks to probe students’ example spaces, we developed three prototype tasks for use in clinical interviews with undergraduate students.

Participants and protocol

We invited students taking a first-year linear algebra course to participate in the interviews, near the end of the semester while they were revising for the final assessment. The course is compulsory for students on mathematics and computer science degree programmes, and an option for students on many other degree programmes. The course operates using a flipped classroom design, with preparatory activities that include weekly e-assessment quizzes (for further details about the course, see Docherty, 2023). Out of 581 students on the course, 10 students volunteered to participate in the interviews. During the interview, each student was asked to complete three tasks, with each task in a different format: on paper, as a static e-assessment task, and as an interactive e-assessment task. The task-format combinations were permuted across the 10 interviews, with each task-format combination occurring 3 or 4 times in total. After the students had completed each task, the interviewer asked them to explain how they produced their answers, and whether they could give any other types of examples that had not been covered so far. The interviews were audio-recorded and transcripts were produced for the relevant episodes. However, our main analysis is based on the students' concept examples as recorded by the e-assessment system.

Materials: design of the static and interactive tasks

Three tasks on linear algebra were created for the study, addressing concepts from the course: Eigenvalues, Span, and Reduced row-echelon form. There were two e-assessment versions of each task: static and interactive. The static e-assessment version asked students to provide three examples, and to "try to make each example as different as you can." The interactive e-assessment version asked for a single example; depending on the answers given, students were then prompted for further examples with different properties. Full details for all the tasks can be found at https://osf.io/7wrgz. In this paper, we focus on the Eigenvalues task, which asked students for examples of matrices with eigenvalues 1 and 5 (as shown in Figure 1).

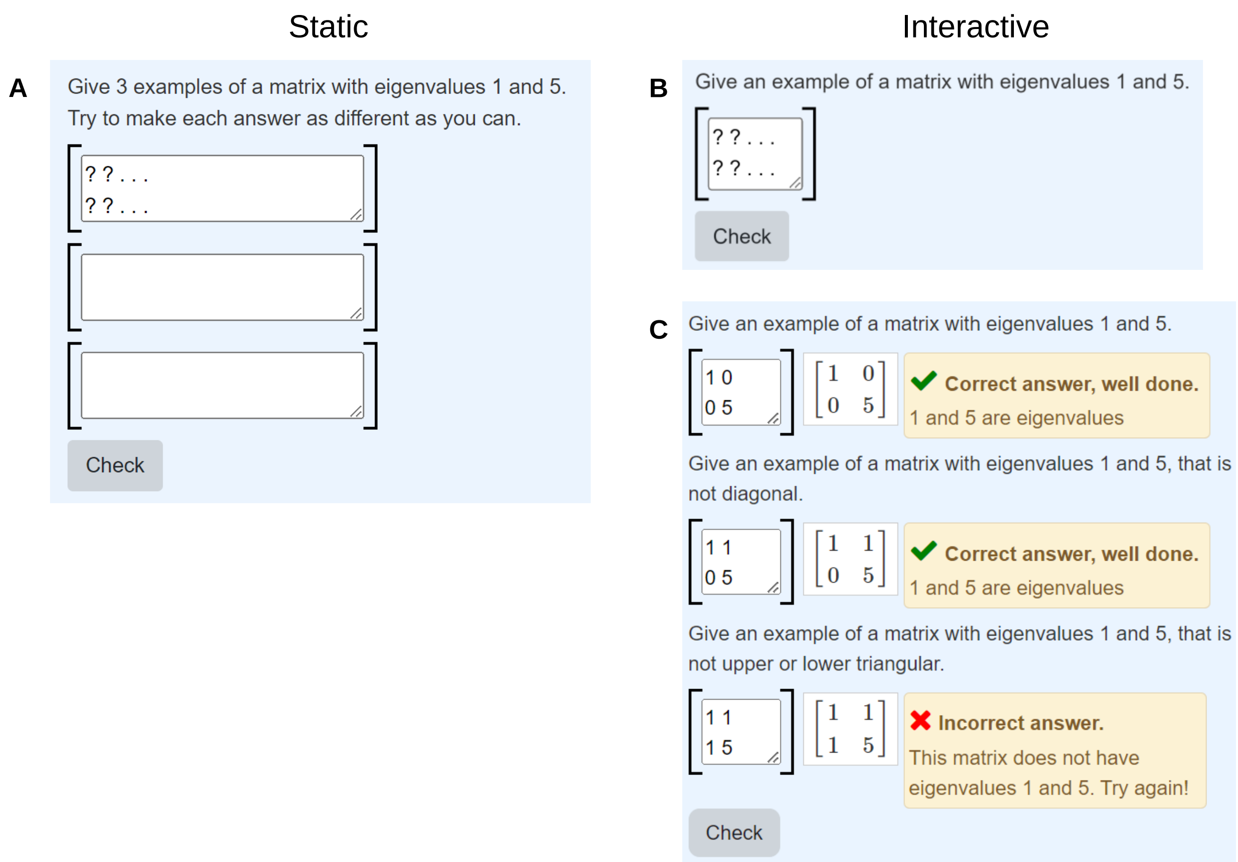

Figure 1: The Eigenvalues task: the static version (A) asked for three different examples, while the interactive version (B) asked for one initial example before prompting for further examples (C).

For the Eigenvalues task, we anticipated that the most accessible example for students would be the 2×2 diagonal matrix with entries 1 and 5 on the diagonal. We designed the interactive version to prompt for further examples by exploring the DofPV that we identified while devising the task. One DofPV is the type of matrix: in addition to diagonal matrices, upper/lower triangular and non-triangular matrices are possible. According to the principle of intellectual parsimony (Koichu, 2010), we expected that students would first try a triangular example (since that would avoid the need for any calculation of eigenvalues) and would only resort to constructing a non-triangular example if specifically prompted to do so. Another DofPV is the size of the matrix, which can be $n \times n$ for any $n \geq 2$ (i.e., the range of permissible change is $n \geq 2$). This dimension may not be immediately obvious to students (despite the course dealing with matrices of different sizes), since the two eigenvalues provided in the task may cue students to think of 2×2 examples. Moreover, the task is deliberately vague in not specifying that 1 and 5 are the *only* eigenvalues, and in not specifying their multiplicity.

We implemented the interactive tasks in the STACK e-assessment system, using the feedback messages displayed after a student submitted an answer. If the answer was incorrect, a feedback message was displayed (e.g., "This matrix does not have eigenvalues 1 and 5. Try again!"); the student could then change their answer and re-submit. If the answer was correct, the feedback message included hidden JavaScript code that revealed the next prompt and input box (the code is available at https://osf.io/7wrgz). We opted to use this relatively simple design, even though it imposed a constraint on the decisions about which prompt to show next: these decisions could only be based on the first and last submitted answers, rather than the full sequence of examples provided so far. In line with this constraint, we developed a flowchart for deciding which prompt to present next. The flowchart for the interactive Eigenvalues task is shown in Figure 2.

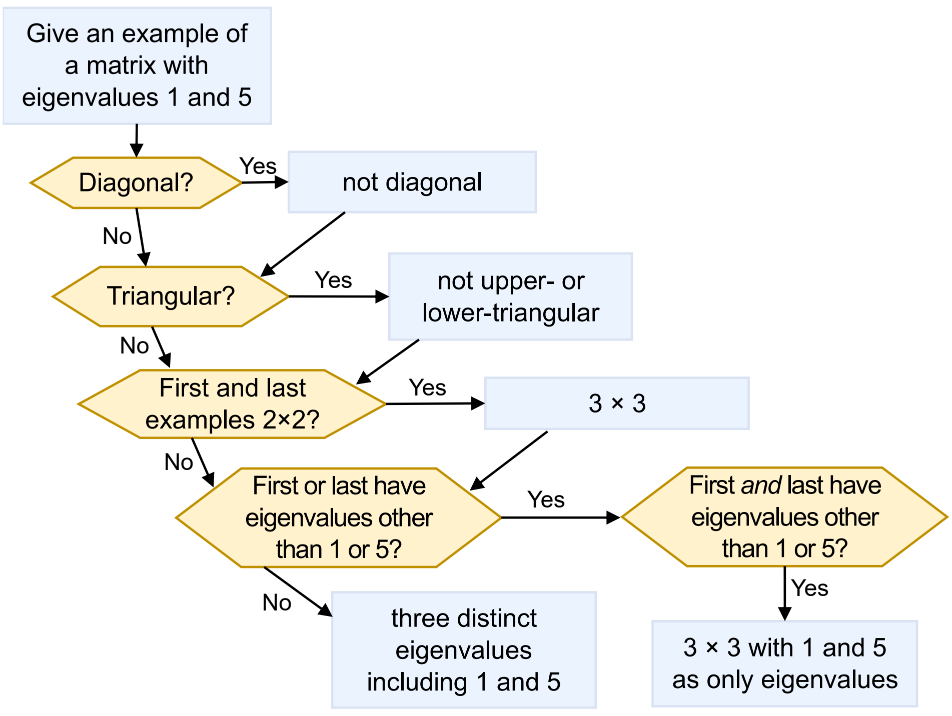

Figure 2: Flowchart showing the properties that were checked to determine the next prompt in the interactive Eigenvalues task

Results

The student responses to the Eigenvalues task are shown in Table 1 (interactive version; 3 students) and Table 2 (static version; 4 students). All students gave the diagonal matrix with entries 1 and 5 as their first example, confirming our expectation that this would be the most accessible example.

For the interactive version of the Eigenvalues task, after the initial diagonal example, only one of the three students (S1) proceeded to produce a triangular example as expected (i.e., the triangular examples were less accessible than we had anticipated). The other two students (S2 and S10) moved immediately to considering the characteristic polynomial of a general 2×2 matrix and produced non-triangular examples. S2's example was incorrect; the student was stuck at this point so the interviewer intervened with a correct example so they could proceed to the next prompt. S1 similarly produced an incorrect 2×2 example when the task prompted them for a non-triangular matrix, however they switched (unprompted) at that point to consider 3×3 examples. When asked to explain their method after completing the task, S1 said that "there probably is a way" to make a 2×2 non-triangular example. All three students were able to produce 3×3 examples with the required properties.

	Student 1	**Student 2**	**Student 10**
	$\begin{bmatrix}1&0\\0&5\end{bmatrix}$	$\begin{bmatrix}1&0\\0&5\end{bmatrix}$	$\begin{bmatrix}1&0\\0&5\end{bmatrix}$
Prompt 1	*Not diagonal?* $\begin{bmatrix}1&1\\0&5\end{bmatrix}$	*Not diagonal?* $\begin{bmatrix}\sqrt{2}&\sqrt{2}\\-5\sqrt{2}&5\sqrt{2}\end{bmatrix}$ (✗) *[interviewer intervened]*	*Not diagonal?* $\begin{bmatrix}3&2\\2&3\end{bmatrix}$
Prompt 2	*Not triangular?* [incorrect 2×2] $\begin{bmatrix}1&0&0\\1&1&1\\0&0&5\end{bmatrix}$	*3×3?* $\begin{bmatrix}1&0&0\\0&5&0\\0&0&1\end{bmatrix}$	*3×3?* $\begin{bmatrix}1&0&0\\0&1&0\\0&0&5\end{bmatrix}$
Prompt 3	*Three distinct eigenvalues?* $\begin{bmatrix}1&0&0\\0&5&0\\0&0&2\end{bmatrix}$	*Three distinct eigenvalues?* $\begin{bmatrix}1&0&0\\0&5&0\\0&0&3\end{bmatrix}$	*Three distinct eigenvalues?* $\begin{bmatrix}1&0&0\\0&5&0\\0&0&2\end{bmatrix}$

Table 1: Student responses to the interactive Eigenvalues task.

Four other students attempted the static Eigenvalues task (see Table 2). Two of the students (S5, S9) only managed to provide the diagonal example; while both students spent several minutes working on paper with the characteristic polynomial of a general 2×2 matrix, neither was able to generate a further non-diagonal example, and neither considered larger matrices (although, when prompted at the end of the interview, both students were able to produce further examples). The other two students (S6, S7) generated further examples after the diagonal matrix. S7 noted that the task "didn't say anything about the multiplicity" when explaining their use of a 3×3 matrix for the second example. They also demonstrated awareness of the generality represented by this example (an upper-triangular matrix) when explaining their answer: "If it is a triangular matrix then the determinant is the product of the leading diagonal. If you take the characteristic polynomial it ends up $(x-1)(x-1)(x-5)$". They were not confident in their final example ("it's possibly wrong") but reasoned using the cofactor expansion for computing determinants, where "I think the zeros would cancel out." S6 also reasoned using the cofactor expansion, but gave an incorrect 3×3 example. When asked how they produced this example, they explained that when choosing the matrix entries, "as long as I have one row or one column of zeros, I can fill in any random values" (i.e., they appeared to overlook the need to consider the determinant of the bottom-right 2×2 sub-matrix). S6 was able to produce a correct non-diagonal example: an upper-triangular 2×2 matrix that they described as "a variation of" their first example.

Student 5	**Student 6**	**Student 7**	**Student 9**
$\begin{bmatrix}1 & 0\\ 0 & 5\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 5\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 5\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 5\end{bmatrix}$
	$\begin{bmatrix}1 & 0 & 1\\ 0 & 5 & 2\\ 0 & 3 & 1\end{bmatrix}$ (✗)	$\begin{bmatrix}1 & 3 & 6\\ 0 & 1 & 4\\ 0 & 0 & 5\end{bmatrix}$	
	$\begin{bmatrix}1 & 2\\ 0 & 5\end{bmatrix}$	$\begin{bmatrix}5 & 0 & 0\\ 2 & 1 & 9\\ 5 & 0 & 5\end{bmatrix}$	

Table 2: Student responses to the static Eigenvalues task.

A summary of the results is shown in Table 3, where each student's examples are classified according to the two DofPV that we identified for this task: the type of matrix and size of matrix. This enables comparison between students (and between the two task modalities), in terms of which DofPV they showed awareness of through their examples (e.g., providing both 2×2 and 3×3 examples shows awareness of the "size of matrix" DofPV). For the static version, S5 and S9 did not show awareness of either DofPV since they provided only one example, while S6 and S7 showed awareness of both DofPV. For the interactive version, all three students showed awareness of both DofPV, although only S1 provided a triangular example within the "type of matrix" DofPV.

		Type of matrix			Size of matrix	
		Diagonal	Triangular	General	2×2	3×3
Static	S5	●			●	
	S9	●			●	
	S6	●	●	○	●	○
	S7	●	●	●	●	●
Interactive	S2	●		○	●	●
	S10	●		●	●	●
	S1	●	●	●	●	●

Table 3: Summary of student responses to the Eigenvalues task showing awareness of DofPV as evidenced by correct (filled disc) or incorrect (hollow disc) examples.

Discussion

We developed prototypes of interactive example-generation tasks and tested them with students in a laboratory setting, alongside static versions of the same tasks. The analysis presented here focused on one task, about matrices with given eigenvalues. Overall, the interactive version of the task appears to have been successful in prompting students to consider both DofPV that we had targeted in the design of the task. This stands in contrast to the static version of the task, where two of the four students provided only a single example and therefore did not demonstrate awareness of any DofPV (though they were later able to do this with explicit prompting from the interviewer).

Interactive example-generation tasks have the advantage of providing a controlled approach for prompting learners to generate examples, which is important when seeking to "make inferences about participants' knowledge from the examples they generate" (Zazkis & Leikin, 2007, p. 19). Since all the prompts in the e-assessment task need to be decided in advance, careful thought should be given to the sequence of prompts; and where the e-assessment system has powerful mathematical capabilities, properties can be checked more quickly and reliably than an interviewer or a teacher could manage to do on the spot. However, the pre-designed prompts may turn out to be sub-optimal when learners respond in unanticipated ways. For instance, with the Eigenvalues task, S2 and S10 were not prompted to produce a triangular example since they skipped over that step in the anticipated sequence of "diagonal → triangular → non-triangular" (i.e., the triangular examples were less accessible than was anticipated). This demonstrates the value of pilot studies in the development of interactive example-generation tasks, so that the design can be refined in light of students' responses.

The results of this pilot study of the Eigenvalues task suggest that the flowchart that we developed (Figure 2) could be refined to ensure that students are prompted to consider the full range of anticipated DofPV. We have already begun to work on a more sophisticated implementation, that could make decisions about which prompt to give based on details of the entire sequence of examples generated so far: for instance, "both of your examples are 2×2, can you give one that is 3×3?", or "could you think of a simpler example, like a triangular matrix?". This approach could be combined with asking for two or three different examples at the outset (similar to Fahlgren & Brunström, 2023), to get some indication of the most accessible examples and the DofPV that the student is aware of, and thus determine which DofPV would be fruitful ones to explore next.

Two further improvements could be considered for this task, which may also apply to other tasks. First, it could be worthwhile to offer students hints about how to proceed if they struggle to generate an example, even for types of example that are expected to be the most accessible. For instance, S2 was unsure how to proceed after their second example was incorrect; the interviewer intervened to move them on, and without this intervention they may not have had the chance to consider other DofPV. This sort of intervention could perhaps be formalized, by giving the student the option of asking for a hint. Second, the task could include prompts for examples that are not possible, to probe the students' understanding of the RofPCh for particular DofPV. For instance, students could be asked to "give an example of a 2×2 matrix with eigenvalues 1, 2 and 3 (or enter none if this is not possible)".

This was a small-scale pilot study of these tasks, so we do not seek to make any strong claims about the comparison between the interactive and static task formats. However, our findings do suggest that the interactive example-generation task was able to stimulate learners to consider a broad range of examples, and that this range was broader than the one demonstrated by the static task group. Indeed, students completing the static versions of the tasks admitted to giving up on the instruction to produce examples that were "as different as possible". We believe that the interactive example-generation task format therefore warrants further development and study.

Acknowledgment

We would like to thank Molly Berrick, Abbie Whitehouse, Eilidh Millin and Álvaro Beattie Eizaguirre who collected data for this study as part of their fourth-year project at the University of Edinburgh.

References

Arzarello, F., Ascari, M., & Sabena, C. (2011). A model for developing students' example space: The key role of the teacher. *ZDM*, *43*(2), 295–306. https://doi.org/10.1007/s11858-011-0312-y

Docherty, P. (2023). Case study 3: An introductory linear algebra course. In *Effective Teaching in Large STEM Classes*. IOP Publishing. https://doi.org/10.1088/978-0-7503-5231-4ch9

Fahlgren, M., & Brunström, M. (2023). Designing example-generating tasks for a technology-rich mathematical environment. *International Journal of Mathematical Education in Science and Technology*, 1–17. https://doi.org/10.1080/0020739X.2023.2255188

Kinnear, G., Jones, I., Sangwin, C., Alarfaj, M., Davies, B., Fearn, S., Foster, C., Heck, A., Henderson, K., Hunt, T., Iannone, P., Kontorovich, I., Larson, N., Lowe, T., Meyer, J. C., O'Shea, A., Rowlett, P., Sikurajapathi, I., & Wong, T. (2022). A collaboratively-derived research agenda for E-assessment in undergraduate mathematics. *International Journal of Research in Undergraduate Mathematics Education*. https://doi.org/10.1007/s40753-022-00189-6

Koichu, B. (2010). On the relationships between (relatively) advanced mathematical knowledge and (relatively) advanced problem-solving behaviours. *International Journal of Mathematical Education in Science and Technology*, *41*(2), 257–275. https://doi.org/10.1080/00207390903399653

Sangwin, C. J. (2003). New opportunities for encouraging higher level mathematical learning by creative use of emerging computer aided assessment. *International Journal of Mathematical Education in Science and Technology*, *34*(6), 813–829. https://doi.org/10.1080/00207390310001595474

Watson, A., & Mason, J. (2005). *Mathematics as a Constructive Activity*. Routledge. https://doi.org/10.4324/9781410613714

Zazkis, R., & Leikin, R. (2007). Generating examples: From pedagogical tool to a research tool. *For the Learning of Mathematics*, *27*(2), 15–21.