Nelly Litvak
Department of Mathematics and Computer Science
Eindhoven University of Technology
n.v.litvak@tue.nl

Lotte Weedage
Faculty of EEMCS
University of Twente
l.weedage@utwente.nl

Column Better than blackboard

Do we teach what we preach?

In this column 'Better than blackboard' Nelly Litvak writes about teaching mathematics at university. She will address problems that many university teachers face. This time she writes the column together with Lotte Weedage, a former student of hers who is now finishing her PhD and teaches mathematics at university herself.

The course Linear Structures 1, in the first quarter of the first year of the BSc Applied Mathematics (University of Twente), covers the basics of abstract linear algebra. In the first week, the students see the axiomatic definition of linear spaces, learn that $1+1=0$ in field Z_2, and have to prove statements like 'zero is unique', startled by the very fact that such proof is needed, let alone knowing where to start. This is students' first encounter with abstract mathematics, their first step in the deep waters of their profession. This course is a thrill to teach!

In 2014, Linear Structures 1 was given for the first time. Nelly was teaching it, and Lotte was following the course as a first-year student. In a couple of years Lotte started helping as a teaching assistant, taking more tasks every year. In the fall 2023, Lotte, now close to completing her PhD, is the responsible lecturer.

Through years, we changed the structure and the grading a lot. Not for the sake of changes. But because we see that even if we do our best to explain, and the students work very hard, they still, somehow, don't get the importance of building a mathematical argument. When doing proofs, they rush to the 'right answer', skip steps, make wrong implications, and do not look critically at their solutions. This is the problem, that echoes Paul Lokhart's beautiful essay *A Mathematician's Lament* [6]:

> "By concentrating on what, and leaving out why, mathematics is reduced to an empty shell. The art is not in the 'truth' but in the explanation, the argument. [...] Mathematics is the art of explanation."

This is what we observe time and again. Students don't grasp the essence of doing mathematics. They are scared of proofs, they skip explanations, they focus on what instead of why. Year after year we have been on the quest of teaching our first-year mathematicians the art of building a mathematical argument. We have gone a long way, and we are not there yet, but we feel now is a good time to share our thoughts and experiences with you.

Illustration: Eline van Hove

Missing skills

Here is a simple example to illustrate our point that students often lack the skill of building a mathematical argument. Look at a multiple choice exam question in Figure 1. Even if someone never heard of a parallelepiped, the word 'volume' already brings the number of reasonable answers down to four. A volume cannot be negative!

This was a digital exam, assisted by a person from technical support, a recent graduate of Computer Science, let's call him Robert. As answers started coming in, Robert's eyes grew wide. What he saw, we show in Figure 2. 25% of the students answered that volume is -6. If you ever taught determinants, you probably already know what happened. The volume is an absolute value of the following determinant:

$$\begin{vmatrix} 2 & -1 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{vmatrix} = -6.$$

The students compute the determinant, and give the answer -6, completely forgetting that the volume is the absolute value of that, 6.

When students see this after the test, they say: "Ah, of course! I forgot the absolute value! This is a stupid mistake…" But, we believe, this is not a stupid mistake. This is a lack of skills.

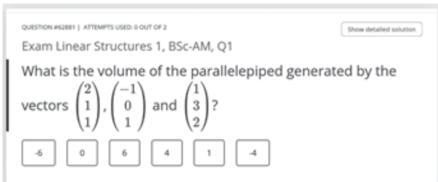

Figure 1 Multiple choice question at the exam Linear Structures 1, 1st quarter, 1st year BSc Applied Mathematics.

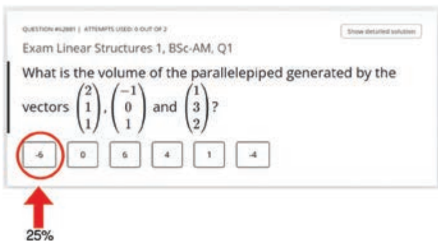

Figure 2 Multiple choice question at the exam Linear Structures 1, 1st quarter, 1st year BSc Applied Mathematics. 25% students answered that the volume equals -6.

When a mathematician solves a problem, they first start thinking, “What can the answer be? What answer is impossible? What are special cases?” After solving the problem, a mathematician asks themselves: “Does this answer make sense? Is there another way to check this answer?”

If only students asked themselves: “What can the answer be?”, or “Does this answer make sense?”, they would never forget the absolute value! Problem is, they skip these questions. And after they see the test, their reaction is not: “Next time I will think about the answer before starting to solve the problem”, or “Next time I will check my answer critically.” No. Their reaction is: “Next time I won’t forget the absolute value.” They don’t even realize that they are missing the most essential skills! This is very concerning for any teacher who is trying to teach the art of a mathematical argument.

(As a side note, if the question said “$(-1, 0, 1)$, $(2, 1, 1)$ and $(1, 3, 2)$” instead of “$(2, 1, 1)$, $(-1, 0, 1)$ and $(1, 3, 2)$”, then the determinant would be 6, and all these 25% of the students would give a correct answer. This casts serious doubts on the informativeness of the standard tests. We will definitely come back to this in later issues of this column.)

This is of course just a small example. One may argue at length what it really tells us. But we hope we made ourselves clear. Building a mathematical argument is what we value most in doing mathematics, yet it seems that students do not learn this in our courses. We don’t teach what we preach.

Tell me, and I will forget…

One may say: “But I always tell to the students that proofs are important, that each step must be explained.” Yes, we do this, too. At the opening lecture of Linear Structures 1, we tell the students right away, how they should grow to love the abstraction, how the argument is more important than the answer. We come up with most convincing (in our view) arguments, metaphors and cartoons. We keep repeating this during the course. But preaching better doesn’t help. As the saying goes: “Tell me, and I will forget…”

One may say: “But we work out so many problems, students can see how it’s done!” Yes, we do. Students like worked-out examples because this is how they think they can learn. Yet, in our experience, students usually don’t try to produce a similar write-up themselves, for two reasons.

First, they cannot learn how to write by merely looking at it. “Show me and I may remember…” Or, maybe not…

Second, they don’t see the need of writing their solution as carefully as the teacher does, perhaps because they feel that the goal of their writing is to report their knowledge to the teacher, and the teacher doesn’t need a detailed explanation because they already know everything. The question “Should we write it like this at the exam?” is probably familiar to all university math teachers.

Mazur’s system: safe space for mistakes

If we were to write this article in 2016, we wouldn’t talk about missing skills as we do now. Instead, we would write: “Students struggle with proofs.” They don’t know where to start and finish, they don’t explain their steps, they don’t scrutinize their solutions for errors. And we don’t know what a teacher can do about it.

In spring 2016, Eric Mazur, a physics professor from Harvard University, was touring universities in the Netherlands with plenary lectures and workshops about innovative education. Nelly participated in a workshop about his extremely innovative hands-on Applied Physics course. One of his ideas looked promising to teach students how to do proofs.

Mazur’s system works as follows. First, students receive a problem set to do at home. They must solve problems and write solutions using Polya’s four steps [7]:

Step 1. Analyze the problem.
Step 2. Devise a plan.
Step 3. Execute the plan.
Step 4. Reflect on your solution and the answer.

The level of these problem sets is similar or slightly higher compared to a typical exam. In Figure 3 is an example of a problem that we use now.

Important condition is that the students must solve the problems individually, without help of their classmates. Asking this from students may sound naïve, but here is another crucial condition that makes it work: the solutions are allowed to be completely wrong! These solutions are evaluated based on their completeness only. That is, they must show visible and serious effort in all four steps. The goal is to struggle and to learn, not to solve everything correctly. We also tell this to the students over and over again: embrace the struggle! Since there is no benefit in a correct answer, there is no real reason to cheat. (This of course depends on the grading as well, we will explain our approach to grading later.)

Then, in class, students discuss their solutions in small groups and make corrections with red pen. The goal is to identify their own errors, and write a reflection: Was it a computational or conceptual error? What erroneous thinking has led to this error? Et cetera.

After the group discussion, students receive full correct solutions, and the teacher can explain important points at the board, address common difficulties, and answer questions.

By design, students learn the missing skills when they write and discuss detailed solutions. And maybe even more importantly, this method gives the students a safe space to make mistakes. All literature, from education [1, 2] to neuroscience [3], agrees on this: for learning, it is important to make mistakes, and equally important to have a safe space where you can make mistakes without being punished. Usually, graded assignments don’t give such safe space because every error makes the

Proof of the Week 2.
Linear Structures 1. Structures and Models

IMPORTANT:
– Write your solution in full sentences.
– Follow the four steps in the solution, see information on Canvas.

Consider the set S that consists of all double-differentiable functions $f \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ such that:

$$f'' + f = 0.$$

Is S a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$? Why or why not?

Figure 3 One of the problems that we use in the Mazur’s system.

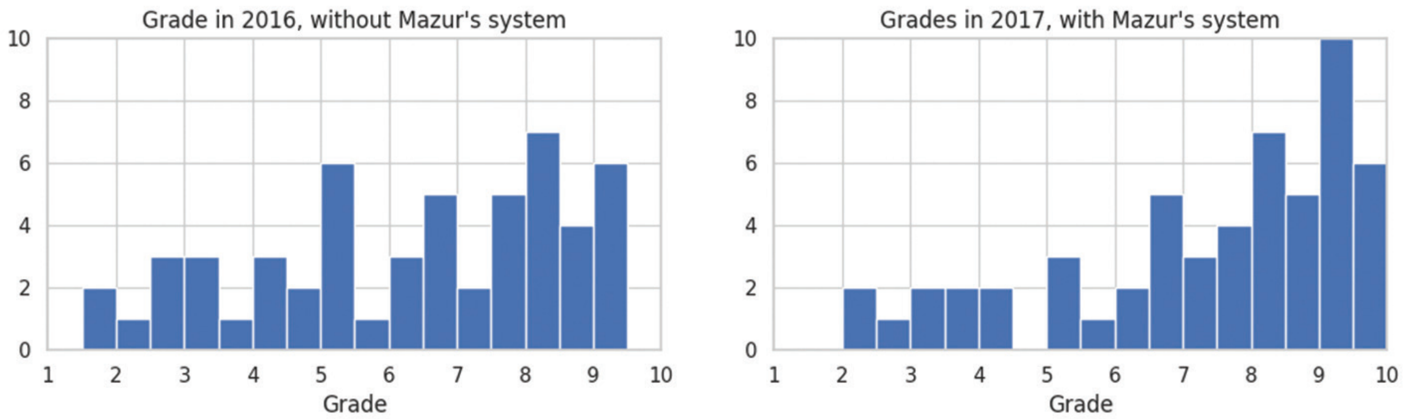

Figure 4 Exam grade distribution without and with Mazur's system. On the horizontal axis, the intervals of grades from 1 (lowest) to 10 (highest). On the vertical axis, the number of students that received that grade.

grade lower. In contrast, in Mazur's system, students can receive a high grade for a completely wrong solution, if they rigorously went through the four steps, and fully reflected on their errors. This gives the students the very much needed safe space to make mistakes!

Mazur's system works

Nelly introduced Mazur's system in the course right away, and was very happy with the results, see Figure 4. Many students find the system very helpful. For example, one student, Lavinia Lanting, is now doing research on this system for her second MSc degree in educational sciences. She could choose any topic, but she picked this one because, she says, at that time, this system helped her a lot to understand the subject. It is not obvious for a first-year student that explaining the problem in four steps helps to understand an abstract concept. Now, when she is doing MSc in mathematics, Lavinia knows that it works, and wants to discover more about it.

From problem sets to 'Proof of the week'

At the beginning we used sets of three to four problems three times in a quarter. However, this took students a lot of time, because the problems were difficult, and the students were not used to write in such detail, especially when they weren't sure that their solution was correct. Some students said they spent twenty hours on one problem set! Also, there was not enough time in the class to discuss all difficult points, and there was no time left to discuss the write-up.

Last three years we give only one problem per week. We call it 'Proof of the week'. The problem is about the material of the week before. In the class, the students discuss and correct their solutions, but since we have more time, they also write down their joint 'group' solution. Usually we ask them to write down one or two steps, for instance, Steps 1 and 2, or Steps 3 and 4. After that, they can check solutions of other groups and/or the teachers come to every group and give feedback on the group solution.

We were worried that devoting a complete class to only one problem was an overkill. But students like it, and we rarely finish early. Since the students go into the problem and the material around it in great detail, one problem, surprisingly, is enough.

Struggles with Polya's Steps 1,2 and 4

With Mazur's system, it becomes painfully clear that in students' mind, the 'solution' is Step 3 only. It doesn't help to repeat to them that in mathematics papers, Step 3 often goes to the Appendix!

Students truly struggle with Step 1, 2 and 4. In Step 1 (analyze the problem), they often simply copy the problem. They don't understand what Step 2 (devise a plan) is about. In the best case, they write Step 2 after Step 3 is done (which is ok, it just demonstrates the difficulty of gaining the skill of mathematical argument).

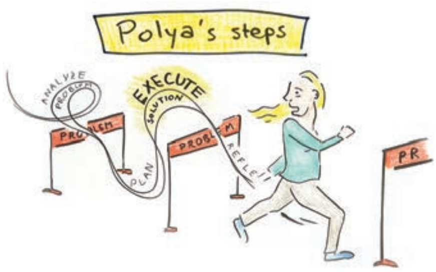

Illustration: Eline van Hove

In Step 4 (reflect on your solution and the answer), they often write "The answer makes sense", without explaining why, and even if it doesn't, or without actually checking their answer. For example, a problem was to find matrix B that is an inverse for matrix A. A group of students found a wrong inverse matrix. Yet, in Step 4, they write: "Our solution is correct, since A times B equals the identity matrix." Except it doesn't! So they knew what they had to check but didn't know how to check it. As another saying goes: "They heard the bell ring but didn't know where the clapper was ..."

We provide very detailed instructions for each step, and keep revising them, but honestly, this remains a point for improvement. Maybe you have a brilliant idea how to explain the meaning of Steps 1,2, and 4 to the students, then please share it with us.

To grade or not to grade in Mazur's systems

At the beginning, participation in Mazur's system was optional, for a bonus point. Many students did participate, either for the bonus point, or because they hoped to understand the difficult material better.

In the original approach by Eric Mazur, he assigns points for completeness of the original solutions (visible effort on all four steps in each problem) and for correctness and completeness of the error correction and reflection on the errors. In his system, 50% or less, means zero points.

We started out this way as well, but soon figured out that we do have to give zeros, and students find it very demotivating. Students look at their work very differently than we do. We see a solution with scarce scribbles, some steps missing, and we give 0, insufficient effort. However, it might be that the student worked very hard on the problem. Ironically, we believe that part of the problem is that our students are used to being very good at math in high school, used to math being easy for them. Now suddenly they struggle and spend lots of time on a problem. They don't know how to solve it, let alone how to write these confusing four steps! In their perception, they work very hard, even when they just stare at a blank paper (which we keep telling them not to do, but they do it anyway). Of course, after hours of frustration, they are very disappointed to get a zero for effort.

At some point, we stopped grading these problems. The quality of work didn't suffer because students who chose to do it, did it for the right motivation to learn. The downside was that students felt that they worked hard but received no feedback.

With 'Proof of the week' we have a solution that works for now. We discuss only one problem per class, and we have two to four teachers who give feedback to each group on their group solution. We have made a rubric, and give it to the students as an indication of the quality of their work, but always with the strong disclaimer that this is not a grade.

Instead of giving a grade, we made the submission of all proofs of the week mandatory for the admission to the exam. Recently we learned that this resembles one of alternative grading systems, called 'Ungrading' [2]. This solution is not ideal, the best would be to give individual feedback to every student. We believe however that giving no grade is the right approach because it removes the stress from the students and contributes to a safe space for making mistakes.

Grading for the mathematical argument

For many years, we used the four steps in Mazur's system, but not in the exam. But two years ago we changed that, and now we ask students to write their proofs in Polya's four steps at the exam as well. It might be an overkill for advanced students, but for the first-years, we find it highly appropriate. Our motivation is, if we want to teach what we preach, we must grade what we preach, too. If we believe that a correct and complete argument is important, we must thoroughly assess the quality of the argument.

To prove: if A is a 3x1 matrix and B is a 1x3 matrix, then rank$(AB) \le 1$.
Proof: by Theorem 3.7, rank$(AB) \le$ rank(B) so this proves that rank$(AB) \le 1$.

Figure 5 Student's solution to an exercise.

One may say: "But we do this in a standard exam as well! We give points, and this depends on how good the argument was." This is right, but insufficient in our opinion. Here is a very typical solution of a good student, in Figure 5.

The student made correct implications, but completely skipped the explanations: why does $\mathrm{rank}(AB) \le \mathrm{rank}(B)$ imply that $\mathrm{rank}(AB) \le 1$? The argument that $\mathrm{rank}(B) = 1$ because matrix B has only one row, is very essential, and is completely missing. We cannot read the students' thoughts on the written exam: should we just believe they wrote this down with the correct reasoning in mind? No! Every proof has to be a story, which any reader should be able to understand. This story is inherently incomplete. Yet, in a standard test, this (almost correct!) solution will be worth lots of points, at least 60%. Indeed, the line of argument is correct, and the crucial theorem is used correctly as well. At the exam

1. Given is that $T: V \to W$ is a linear transformation.
$\{v_1, v_2, \ldots, v_k\}$ - basis for $N(T)$, $\beta = \{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ - basis for V.
To prove is that $\{T(v_{k+1}), T(v_{k+2}), \ldots, T(v_n)\}$ is a basis for $R(T)$.

We know that $N(T)$ is a subspace of V, so it's logical that its basis can be extended to a basis for V. Also, we know what the dimension of V equals to $\dim(N(T)) - \dim(R(T))$. Since we know that the dim of V is n (number of vectors in the basis of V), and the dim of $N(T)$ is k (number of vectors in basis $N(T)$), we know that to satisfy the equality, the dim of $R(T) = n-k$. Also, we know that any linearly independent subset with dimension equal to the dimension of the given space is in fact a basis.

2. I will use the Dimension Theorem to prove that the given set is a basis for $R(T)$.
Then I will use the fact that any generating set for vector space V with size equal to $\dim(V)$ is a basis. not executed 3

3. The Dimension Theorem states: $\dim(V) = \dim(N(T)) + \dim(R(T))$.
We know that there are n elements in the basis for V, and k elements in the basis of $N(T)$.
Then, $\dim(V) = n$, $\dim(N(T)) = k$. We get: $n = k + \dim(R(T))$.
To satisfy the equation, we see that $\dim(R(T)) = n - k$. 2
Since $\dim(R(T)) = n - k$, and the ~~dim~~ size of $\{T(v_{k+1}), T(v_{k+2}), \ldots, T(v_n)\}$ is also $n - k$, and since we know that any subset of a linearly independent set is linearly independent, we see that this set is linearly independent ? and has the same dimension as $R(T)$.
Therefore, it is proven that $\{T(v_{k+1}), T(v_{k+2}), \ldots, T(v_n)\}$ is a basis for $R(T)$.
this is not proven

4. Looking back, I can't think of a way to solve this without the Dimension Theorem. All given info was definitely needed, and I don't think I could prove this without any piece of information that was given. I was glad I could use some connection of dimension and basis vectors, that was definitely handy. 1

Figure 6 Example of an exam solution written in Polyá's four steps (handwriting changed to computer font for privacy reasons). Maximal number of points are: 4 points for Steps 1+2, 4 points for Step 3, 2 points for Step 4.

review, we will have difficulty defending score 50% or lower.

When we ask the students to write solutions in Polya's four steps, we grade them 20%-20%-40%-20% for Steps 1-2-3-4. Steps 1 and 2 are often hard to take apart, therefore we essentially give 40% for Steps 1 and 2 together. When the student merely copies the problem formulation, we give 0 points for Step 1. Step 3 (execute the plan) is at most 40% to begin with, and we give only 20% for Step 3 if the implications are missing or not explained. We give 0 points for Step 4 if it says only "the answer makes sense" without explaining why, and we only give 20% for Step 4 if the student says something extra about the problem, for instance, offers a generalization, or explains relation of the problem to other material. We believe that we can expect this for a perfect grade. We continuously say to students that Steps 1, 2 and 4 are important, and they believe us, because this is consistent with our assessment. Using a proverb again: "We put our money where our mouth is."

The quality of solutions at the exam improved dramatically, here is an example of a solution, in Figure 6. Such solutions are very informative, and nice to read. Consequently, the exam grades became generally higher because in four steps students must explain their thoughts, and by doing so they often demonstrate their understanding even if their solution is not entirely complete or correct.

Teach and grade what we preach

Putting it all together, Figure 7 shows the current workflow of the Linear Structures 1 course. There are two lines of activities in weeks 1–9, each corresponding to one part of the exam.

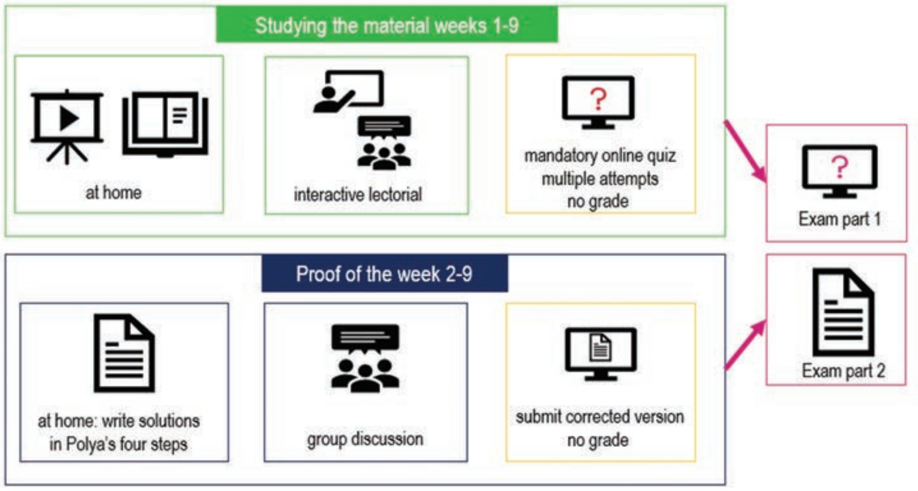

Figure 7 The workflow of the course Linear Structures 1, first year BSc Applied Mathematics, quarter 1, University of Twente.

Learning concepts, definitions, statements, and studying the material

The green rectangle above is where students learn definitions, concepts, statements, and calculations (computing a determinant is an example of the latter).

This part is given in a flipped classroom format. At home the students read the book and watch interactive videos. The videos are pencasts that Nelly made in 2020, also available on YouTube [8]. 'Interactive' means that after a short explanation the video pauses, and students get a question (multiple choice, numerical answer, or drag-and-drop) to check their understanding. The questions are inserted in the videos using H5P plug-in on Canvas. We were very lucky that a very talented teaching assistant Anete Valnere volunteered to make these questions. She did it with lots of rigor and creativity. For example, our favourite question — "Reconstruct the proof" — was Anete's idea. An example is in Figure 8. The proof is chopped in pieces, and the students have to drag-and-drop them in the right order.

The use of videos triggers mixed reactions among mathematicians and computer scientists, but our students appreciate them a lot. (In fact, Nelly also occasionally receives grateful messages for her YouTube pencasts from all over the world.) We like to use the videos because we can refer students back to them, and mainly because we have time for interaction in the class. We find this a much better approach than standard classroom lecture, see also the previous issue of the column, 'We shouldn't give classroom lectures anymore' [5].

In the class (which we call 'interactive lectorial'), the teacher poses questions, and the students answer anonymously with an online tool. (University of Twente uses Wooclap, and we like it, but another software will do as well.) The questions in the class are all about conceptual understanding and common errors. We mostly use multiple choice questions, they are surprisingly suitable for checking conceptual understanding [4]. Often we list four to five statements and ask students to select the correct ones. The students often give wrong answers, which is good, because otherwise, why should they even come to the class? Then the teacher explains the solutions and answers further questions.

One may ask, how do we make sure that the students watched the videos or read

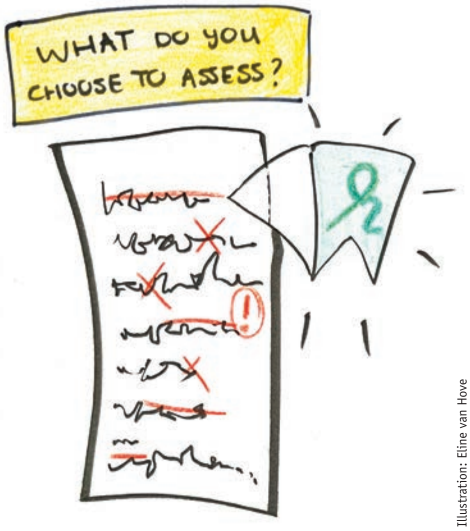

Illustration: Eline van Hove

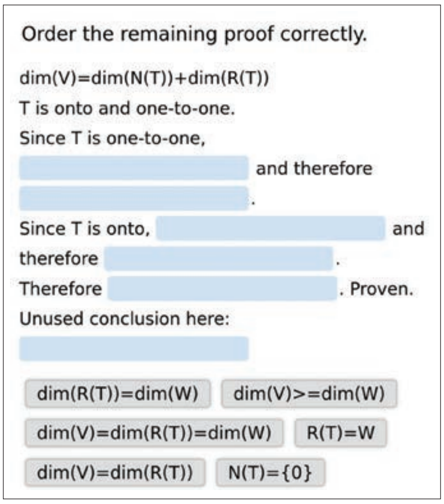

Figure 8 "Reconstruct the proof" question in a video on Canvas, made using H5P plug-in, by Anete Valnere.

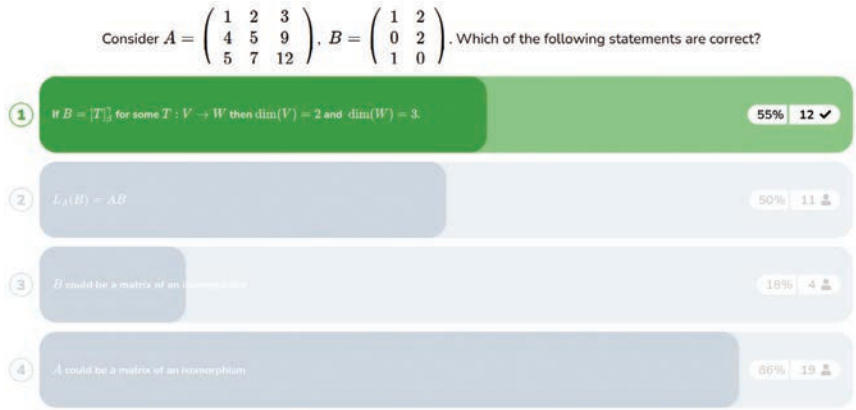

Figure 9 A multiple choice question at an interactive lectorial. Notice that 86% students chose incorrect answer 4. Making these errors helps students to learn.

the book before the class? We don't. In the class, we just assume that they did, and they soon figure out that it makes no sense to come to the class unprepared.

Finally, each week, there is an online quiz, with multiple attempts, no deadline, and no grade. Students can make quizzes together. There are hints and feedback when the answer is wrong, and often a worked-out solution when the student answered correctly. We require a high score of 70–80% to pass a quiz, students must pass quizzes 1–9 to be admitted to the exam.

Quizzes in class and at home offer plenty of safe space for mistakes. There is no grade, multiple attempts, and a lot of feedback and explanations. These activities are very clear to students, they like them and learn a lot from them.

Learning the art of mathematical argument
The blue rectangle below is the 'Proof of the week' that we explained in detail above. Here the students learn to build and write down mathematical arguments.

Exam
At the end of the course, we have an exam that consists of two parts. Part 1 (1,5 hours in the morning) is a digital test on Chromebooks, with the same software and the same type of questions as in the practice quizzes. Part 1 has multiple choice questions on definitions, concepts and statements, and final answer questions on calculations. The quiz is graded automatically. The students must score at least 50% to proceed to Part 2.

Part 2 (1,5 hours in the afternoon) is a written test. It has two proof problems, and the solution must be written in Polya's four steps.

Compared to the standard written test, we have much better experiences with this exam form. First, students are only admitted to the exam after they submit nine quizzes and seven proofs of the week. This sends a clear message that the students must work during the course. It is impossible to submit everything in the last couple of days, so the students don't postpone all the learning to the very last moment, and we can generally count on a reasonable preparation level.

Second, and maybe even more importantly, we grade manually only Part 2, so we grade only two problems per student, and only if they pass Part 1, thus these students are generally well prepared. Besides that this grading is interesting, it also saves lots of time! Back in 2019 we used to hire four teaching assistants, and the six of us were grading for a full eight hours. Now with just two teachers, in four hours we are done with the grading. We now can give the results to the students on the day of the test, and still be at home for dinner.

But most of all, we like that our teaching activities and the exam are consistent with what we want our students to learn. We do feel that we teach what we preach. At least, we try.

Are we there yet?
No, we are not yet where we want to be. Most importantly, our innovative teaching methods often clash with the traditional grade system. For example, in evaluations students complain that during 'Proof of the week', different teachers give different feedback. This is of course only natural in writing mathematics, but it makes students anxious: "What is the 'right' feedback and what should I write at the exam?" After the exam, students also complain, for instance, about getting no partial points for numerical answers in Part 1, and they try to negotiate points in Part 2. All these are irrelevant for doing mathematics. But we cannot blame the students. Their main guideline is the assessment, and they will necessarily focus on this. This is how things are at the moment, even if this is not good for students' learning. In fact, Eric Mazur's plenary talk in 2016 was called 'Assessment: a silent killer of learning.' Luckily, as we recently learned, there are other ways [2], and we will tell you more about it once we try them out.

Colleagues often say: "This is nice what you do. But should we all do it this way?" No, of course not. Polya's four steps is not the only way to learn how to build a mathematical argument. And there are many effective course designs, we just attempted to create one of them.

But we do invite you to join us in honestly asking yourself: "Do I explicitly teach and explicitly grade what I value most in doing mathematics?" We preach what is right. And we owe this to our students: teaching what we preach.

References
1 J. Boaler, *Mathematical Mindsets: Unleashing Students' Potential Through Creative Math, Inspiring Messages and Innovative Teaching*, Wiley, 2015.

2 D. Clark and R. Talbert, *Grading for Growth: A Guide to Alternative Grading Practices that Promote Authentic Learning and Student Engagement in Higher Education*, Taylor & Francis, 2023.

3 S. Dehaene, *How We Learn: Why Brains Learn Better than any Machine... for Now*, Penguin, 2021.

4 B. Huntley, J. Engelbrecht and A. Harding, Can multiple choice questions be successfully used as an assessment format in undergraduate mathematics?, *Pythagoras* 69 (2009), 3–16.

5 N. Litvak, We shouldn't give classroom lectures anymore, *Nieuwe archief voor Wiskunde* 5/24(3) (2023), 145–149.

6 P. Lockhart, *A Mathematician's Lament: How School Cheats us Out of our Most Fascinating and Imaginative Art Form*, Bellevue Literary Press, 2009. Available at http://worrydream.com/refs.

7 G. Polya, *How to Solve It*, Uchpedgiz, 1961.

8 https://youtu.be/ZDocEkktP7w.