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

Noela Müller
Department of Mathematics and Computer Science
Eindhoven University of Technology
n.s.muller@tue.nl

Column Better than blackboard

Competency-based grading

***`Will this be on the exam?'*** **Over the years, I have noticed that my endeavors to make courses more interactive and engaging ultimately came up against the traditional grading system, with students' understandable but counterproductive fixation on the final exam. The main message of today's `Better than blackboard' issue is: It doesn't have to be that way! My co-author for this edition is Noela Müller, an assistant professor at Eindhoven University of Technology. In 2023/24, Noela and I implemented a new grading system in a probability course for 300 Computer Science students. In this column, we will tell you what we did and how it worked out, our story brightened up by cheerful illustrations of the student artist Mara Chelărescu.**

The course *Probability and Statistics for Computer Science* at TU/e enrolls 300 first-year students. It had a long history, but now its position in the curriculum changed, and revision was due.

In 2023, Noela was assigned to teach this course. In discussion with the Computer Science faculty at TU/e, she removed old and added new topics, and made an effort to use more Computer Science-related examples in her slides and exercises. However, she was also not happy with the existing grading system: 4 intermediate tests and one final exam. Students of previous years said that the intermediate tests were demotivating rather than helpful. But then, how to make sure that the students keep up with the course?

Nelly joined this course as an instructor. By a happy coincidence, in the summer of 2023, she had read the book *`Grading for Growth'* by Clark and Talbert [1] and was highly inspired by the concept of alternative grading: It is well developed, well studied, widely-used, comes in many forms, and is suitable for testing of any kind.

One approach described in [1] – standard-based, or, competency-based grading – sounded very suitable for basic mathematics courses, and we decided to give it a try. We believe it worked well, and we are happy to share what we have done, as well as our and our students' experiences.

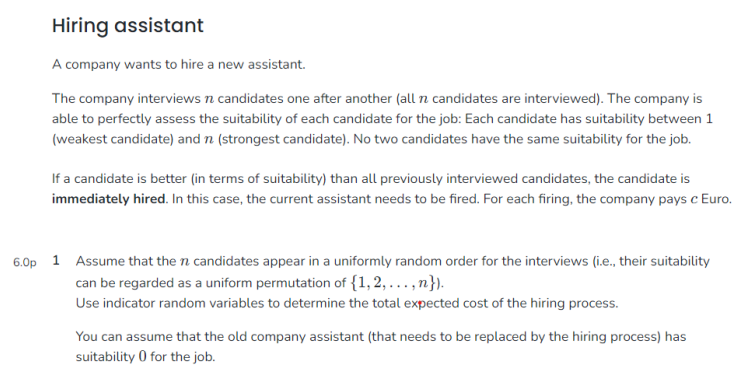
16 competencies

First, from the course content, we identified 16 competencies that we wanted our students to develop. You can see these competencies, called C1-C16, in the first two columns of Table 1.

Following [1], we formulated the competencies in an actionable form: `I can…' . One may say, we could have equivalently formulated them in terms of topics, but we daresay it's not the same. Indeed, take the topic of C14 – `Confidence intervals'. What should a student do to demonstrate understanding of this topic? This is left up to students' interpretation. The actionable `I can…' answers this question, and hence makes the requirements of the course

#	Competency	ANS Test	When
1	I can understand and apply the rules of probability.	Num+MC	week2, week3, exam, resit
2	I can understand and apply conditional probabilities.	Num+MC	week2, week3, exam, resit
3	I can understand and apply the basic theory of discrete random variables.	Num+MC	week3, week4, exam, resit
4	I can choose and apply an appropriate discrete distribution in context.	Num+MC	week3, week4, exam, resit
5	I can understand and apply the basic theory of continuous random variables.	Num+MC	week4, week5, exam, resit
6	I can choose and apply an appropriate continuous distribution in context.	Num+MC	week4, week5, exam, resit
7	I can derive the expected running time of simple algorithms using linear functions of random variables.	Open Q	week5, week6, exam, resit
8	I can understand and apply the Central Limit Theorem.	Num+MC	week5, week6, exam, resit
9	I can understand and construct Markov chain models in the context of computer science.	Open Q	week6, week7, exam, resit
10	I can understand and apply the basic theory of Markov chains.	Num+MC	week6, week7, exam, resit
11	I can compute and interpret descriptive statistics.	Num+MC	week7, week8, exam, resit
12	I can understand and carry out the basic analysis of statistical estimators.	Num+MC	week7, week8, exam, resit
13	I can derive parameter estimators using the method of maximum likelihood.	Open Q	week8, exam, resit
14	I can compute and interpret basic confidence intervals.	Num+MC	week8, exam, resit
15	I can understand the basic theory of hypothesis testing and carry out simple tests.	Num+MC	exam, resit
16	I can understand and apply the basic theory of Bayesian estimation and hypothesis testing.	Num+MC	exam, resit

Table 1 An open question on C7.



Hiring assistant

A company wants to hire a new assistant.

The company interviews n candidates one after another (all n candidates are interviewed). The company is able to perfectly assess the suitability of each candidate for the job: Each candidate has suitability between 1 (weakest candidate) and n (strongest candidate). No two candidates have the same suitability for the job.

If a candidate is better (in terms of suitability) than all previously interviewed candidates, the candidate is **immediately hired**. In this case, the current assistant needs to be fired. For each firing, the company pays c Euro.

6.0p 1 Assume that the n candidates appear in a uniformly random order for the interviews (i.e., their suitability can be regarded as a uniform permutation of $\{1, 2, \ldots, n\}$).
Use indicator random variables to determine the total expected cost of the hiring process.

You can assume that the old company assistant (that needs to be replaced by the hiring process) has suitability 0 for the job.

Figure 1 An open question on C7.

clear and explicit.

Competency quizzes

We assessed each competency separately using a small quiz. All quizzes were conducted in ANS software. The competencies that are marked red in Table 1 were tested with an open question (a written test). The remaining competencies were tested digitally, with one multiple choice and one numerical question. Below, we describe the questions in more detail.

Multiple attempts For most competencies (except the last 4), there were 4 opportunities to take a quiz: 2 during the course, 1 at the exam, and 1 at the resit, see the last column of Table 1. Students made quizzes in class, during tutorials/instructions, for 45 minutes. In week n, 4 quizzes were open, on the competencies covered in weeks n-1 and n-2. Usually, 45 minutes were too short to make all 4 quizzes, so the students chose which quizzes to attempt. During the exam and the resit all 16 quizzes were open for 3 hours.

Looking up is allowed In ANS, we have provided a Statistics compendium that summarized most formulas, and the lecture slides. We did this because in a real-life situation, if students look for material online, they will usually look up formulas and summaries similar to those in compendium and slides. The quiz problems were never the ones on the slides, therefore we believe it was adequate to allow the use of these materials.

Open, multiple choice, and numerical questions

As mentioned before, the 'red' competencies in Table 1 were tested using open questions, for which we required a detailed write-up. These written competencies were chosen because they truly require probabilistic reasoning and due to their relevance for computer scientists, for instance:

> C7: I can derive the expected running time of simple algorithms using linear functions of random variables.

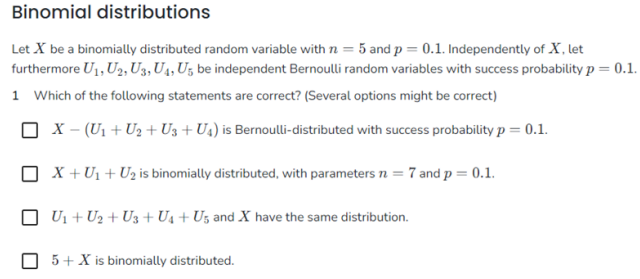
An example of a an open question on C7 is given in Figure 1.

We restricted the open questions to only three competencies because grading open questions takes a lot of time, and is not always useful. For instance, competency C14 - *I can compute and interpret confidence intervals*, is largely procedural, and can be tested very adequately with numerical and multiple choice questions, with the great advantage of automated grading and immediate feedback.

All other competencies were tested with a digital test of two questions: a multiple choice question on understanding the theory, and a numerical question on executing procedures.

We believe that multiple choice questions work very well, see also our previous column [2]. When students are confronted with a yes/no question: `Is this statement right or wrong?', they cannot talk their way around it, as sometimes they can in an open question. See the example in Figure 2 for competency C4 - *I can choose and apply an appropriate discrete distribution in context*.

Our numerical questions were standard, akin to a usual exam, and we often used randomized numbers, so that different students had different correct answers. See an example in Figure 3.



Binomial distributions

Let X be a binomially distributed random variable with $n = 5$ and $p = 0.1$. Independently of X, let furthermore U_1, U_2, U_3, U_4, U_5 be independent Bernoulli random variables with success probability $p = 0.1$.

3.0p 1 Which of the following statements are correct? (Several options might be correct)

- ☐ $X - (U_1 + U_2 + U_3 + U_4)$ is Bernoulli-distributed with success probability $p = 0.1$.
- ☐ $X + U_1 + U_2$ is binomially distributed, with parameters $n = 7$ and $p = 0.1$.
- ☐ $U_1 + U_2 + U_3 + U_4 + U_5$ and X have the same distribution.
- ☐ $5 + X$ is binomially distributed.

Figure 2 Example of a multiple choice question on competency C4 - *I can choose and apply an appropriate discrete distribution in context*.

Marks for the quizzes

Following [1], instead of numerical grades for quizzes, we gave only three marks to `indicate progress':

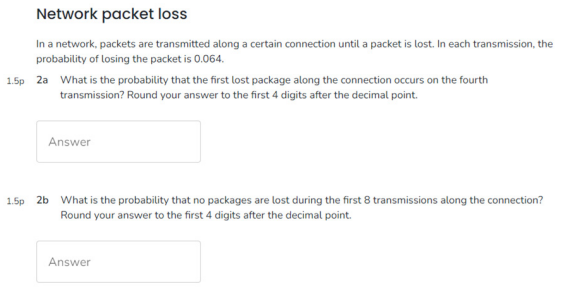
S = Success
G = Getting there
N = Not yet.

In a digital quiz (multiple choice + numerical question) we gave S if both answers were answered 100% correctly, G if one of the answers (numerical or multiple choice) was answered 100% correctly, and N otherwise.

In multiple choice questions, we didn't give partial scores to minimize the chances of success by a random guess.

In numerical questions, we counted only correct answers, with a margin to account for rounding errors. We made it clear to the students that by `I can execute a procedure' we mean: bringing calculations to the correct answer. As a consequence, there were no partial points, but this was amply compensated by the enormous advantage of immediate grading and feedback, and revision opportunities.

Open questions were graded manually. Since it was only one question per competency, and not all students chose to make the quiz in that week, we always could grade within a week.


Network packet loss

In a network, packets are transmitted along a certain connection until a packet is lost. In each transmission, the probability of losing the packet is 0.064.

1.5p **2a** What is the probability that the first lost package along the connection occurs on the fourth transmission? Round your answer to the first 4 digits after the decimal point.

Answer

1.5p **2b** What is the probability that no packages are lost during the first 8 transmissions along the connection? Round your answer to the first 4 digits after the decimal point.

Answer

Figure 3 Example of a numerical question on competency C4 - *I can choose and apply an appropriate discrete distribution in context*.

We used 3 criteria:

1. The solution is correct.
2. All variables are defined.
3. All steps are explained.

We gave S if all three criteria were met; G if 2 out of there criteria were met; N otherwise. We always gave a detailed feedback that students could use when retrying the quiz. We gave S even if there were slight flaws in a write-up, as long as the student confidently demonstrated understanding and ability to explain. Looking back, the three criteria and grading rules were not very convenient because often a student met all three criteria partially. In these cases, we struggled to communicate which criteria were met and why we gave G or N.

Altogether, the quizzes had high requirements: 100% correct answers, correct procedures, good write-up. But we also offered a high support with immediate feedback and multiple attempts. We were happy to learn that this is exactly the right principle in mentoring young people: *high standards - high support* [3].

Final grade

Number of S's and G's The final grade was defined by the number of S's and G's that the student has obtained on the competency quizzes. Here, we had the conversion rule that S in 1 competence = G in 2 competencies.

Open question requirement For passing the course, we required at least one S in an open question, and the number was higher for a higher grade (see below). This gave the competencies with open question a higher weight. We found this fair because these competencies are important for Computer Scientists, and because writing a probabilistic argument is an important skill.

Quizzes at home don't count Technically, any student could access weekly quizzes on ANS, also at home. However, we registered attendance, and quizzes made at home didn't count for the grade.

Exam requirement At the exam, we used STEP sticks that locked student's browsers. During weekly tests, we didn't have this facility. After a discussion with the examination board, we imposed an additional requirement: S in at least 4 competencies during one exam session (either at the exam or at the resit). Again, S in one competency could be replaced by G in two competencies. Importantly, for the exam requirement, students could redo the competencies where they already had S or G before. If they did so, they did not improve their grade, but this gave them a choice to meet the exam requirement more securely. It was up to the students how many new S's and G's they tried to score during the exam.

Computing final grade To summarize, here are our requirements for each grade, across weekly quizzes, the exam, and the resit:

- Grade 6: S in 8 competencies in total; S in 1 competency with an open question; S in 4 competencies at the exam or resit.
- Grade 7: S in 10 competencies in total; S in 1 competency with an open question; S in 4 competencies at the exam or resit.
- Grade 8: S in 12 competencies in total; S in 1 competency and G in 1 competency with an open question; S in 4 competencies at the exam or resit.
- Grade 9: S in 13 competencies and G in 1 competency in total; S in 2 competencies with an open question; S in 4 competencies at the exam or resit.
- Grade 10: S in 15 competencies in total; S in 2 competencies and G in 1 competency with an open question; S in 4 competencies at the exam or resit.

In Figure 4, we made a hypothetical example of a student with final grade 7. In the last column, in yellow, are S's and G's scored at the exam. This student did very well on the new competencies C14, 15, but they also retook familiar competencies C1, 3, 4. Quizzes made at home in week 7 don't count for the grade.

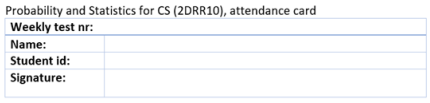
Creating quizzes

For each competency, we made 6 quiz-

Example: Grade=7

Competency #	week2	week3	week4	week5	week6	week7	week8	Exam	Resit	Total
1	S	no attempt						S		S
2	G									G
3			S					S		S
4			S					S		S
5			G	N						G
6			N	G						G
7				N	S					S
8					S					S
9						at home				
10										
11							S			S
12							S			S
13							N			
14								S		S
15								G		G
16										

Figure 4 Results of a hypothetical student, final grade 7. Attempts at home don't count. At the exam, the student may attempt quizzes for new competencies, or repeat the ones they passed before.


Probability and Statistics for CS (2DRR10), attendance card

Weekly test nr:	
Name:	
Student id:	
Signature:	

Figure 5 Attendance slip.

zes: 2 practice quizzes, 2 weekly quizzes, 1 exam, and 1 resit. Altogether, 16×6=96 quizzes. This is a lot of work, even if we had some question banks already available. Since the course had 300 students, we had 4 instructors. We have divided the competencies, four per instructor: C1-4, C5-8, C9-12, C13-16. Each instructor made 24 quizzes, 6 per assigned competency. It is still a lot, but definitely doable.

Grading

Only the competencies with an open question had to be graded, which was done by the four instructors. Each instructor had a group of about 75 students, and they graded their own group.

Altogether, it was not a lot of grading. Indeed, there were only three competencies with exactly one open question. In the weekly quizzes, not all students attempted these open questions, and the quizzes were spread over the weeks, so we could always grade within a week. Also, students chose to attempt the competency themselves, so even wrong and badly written solutions were still meaningful, and we were happy to engage with students' work and give feedback. Plus, we also knew that our feedback was useful because the students could retry. This made grading easier and more satisfying.

It was very helpful that we gave only three marks. Usually it was very clear which mark to give. This saved us a lot of exhausting decision making of traditional grading: to give or not to give those 0.5 points? As we mentioned before, possibly the criteria for open questions can be improved to make the grading even easier.

Organization and Logistics

We hired five teaching assistants (TA's): one per instructor, and one for administration.

At the beginning of each quiz, we distributed attendance slips on paper (see Figure 5). Then a TA collected the slips and marked the attendance in a shared Excel file.

During the quiz, the instructors and the TA's were walking around to make sure that the students had only ANS on their screens.

After the quiz, one of the instructors (Nelly) downloaded the results and saved them in a folder shared with the TA responsible for the grade administration. This TA combined the quiz results with attendance, and posted the results (S,G,N) on Canvas, together with a preliminary grade (provided the student will meet the exam requirement). Already around week 5-6, the students who consistently did well on quizzes starting seeing their preliminary grade turn into a 6 or higher.

We were lucky with an especially talented TA for grade administration who has even created a manual, which we can now use in the next editions of the course.

The four pillars of alternative grading

The book [1] describes the four pillars of alternative grading: clear standards, helpful feedback, marks indicating progress, and revisions without penalty. Altogether, these pillars uphold the feedback loop, vehicle of any learning process. Here is how we experienced the four pillars.

Clear standards These were our 16 competencies. We believe that this worked very well. Here are some quotes from the students:

> "[...] I appreciate this system, because it is completely transparent what is asked from us [...]"

> "The competency system is very useful in helping the student understand their shortcomings and strengths"

This is exactly what we wanted to achieve!

Besides, we believe it is very valuable that the students know exactly which competencies they mastered and which they skipped. This structures their prior knowledge of probability in subsequent courses (see our previous column on the prior knowledge problem [4]). For instance, suppose a later course on randomized algorithms uses Markov chains. Naturally, the teacher will expect this prior knowledge from our course. But if a student consciously skipped competencies C9 and C10, they have no illusion of their prior knowledge. They know they must revise that material.

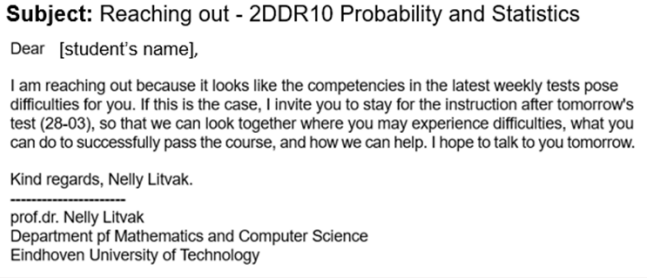
Helpful feedback Digital quizzing software gave immediate feedback on numerical and multiple choice questions. Instructors gave timely feedback on open questions. We believe that the most helpful part was that the students could use the feedback to redo the quizzes in the same competency, on which they received the feedback.

One the other hand, we also observed that even in this new grading system, students often saw feedback mostly as criticism. For instance, this was a quote from evaluations:

> "[..] grading [of open questions] is really particular about `notation' and showing all work"

Clearly, this student didn't find our feedback very helpful. We are now looking into new insights and ideas on how to give helpful and motivating feedback to young people [3]. We hope to report more on this in future issues of the column.

Marks indicating progress It helped us to have only three marks, the grading was easy. However, having letters or phrases instead of grades didn't make much difference to the students. Mentally, the students still perceived S, G and N as a grade, not as an indication of progress. Students are very used to the idea that high grade is good and low grade is bad. We noticed


Subject: Reaching out - 2DDR10 Probability and Statistics

Dear [student's name],

I am reaching out because it looks like the competencies in the latest weekly tests pose difficulties for you. If this is the case, I invite you to stay for the instruction after tomorrow's test (28-03), so that we can look together where you may experience difficulties, what you can do to successfully pass the course, and how we can help. I hope to talk to you tomorrow.

Kind regards, Nelly Litvak.

---

prof.dr. Nelly Litvak
Department pf Mathematics and Computer Science
Eindhoven University of Technology

Figure 6 Reaching out to struggling students.

Illustration: Mara Chelărescu

Figure 7 The four pillars of alternative grading.

that renaming grades did not help much. However, according to [3], another approach could be helpful: constantly repeat to the students what the grade means. An N for a competency doesn't mean that you are not-smart-enough or failing. It is just a current status in that particular competency. We will try to communicate this better next year.

Revision without penalty Along with *Clear standards*, this was a huge advantage of alternative grading. Students liked this a lot, here are some quotes:

> "[.. .] we do not have to stress about lacking in one or two competencies as we can easily make up for it [...]"

> "I really liked the new examination system [...]. It made it a lot less stressful as we had multiple tries per competency."

> "I like that not everything comes down to the final exam [...] ."

Multiple attempts felt very natural to us. We believe that revisions without penalty are simply fair!

Better learning

In our previous column [5] we wrote what is needed for learning: engagement, inconsequential error feedback, and spacing.

Engagement means actively working with the material. *Spacing* means that students work on the course regularly. We have achieved engagement and spacing because 100% of the students were present every week for the quizzes. Here are some testimonies of engagement:

> "The weekly tests are nice and make you actively learn the material"

> "[...] the weekly tests forced us to actively learn the course material, which I think will help the knowledge stick longer.''

About 10% of the students came to the resit to take some more quizzes and improve their grade, although they already passed the course. We did not mind that they chose to work on the course some more!

Inconsequential error feedback is exactly what we did: error feedback and revision without penalty.

In addition, students found the new system less stressful, which is also good for learning.

At the end, almost 80% students passed the course, and the most common grade was 7. While this is higher than with a standard exam, we cannot conclude much on the basis of this result, because grades commonly are higher with alternative grading [1]. We believe, however, that our grades adequately match the abilities and the interests of this group.

Students in control

One crucial advantage of alternative grading is that the students are in control of their results. They know exactly what to do to get a desired grade. We believe the students experienced this exactly as intended. For example, at the tests, we often saw students strategically choosing which quizzes to try. Many students liked having their results under control:

> "I love the grading scheme! It removes so much randomness in the grade - grades don't depend on a single random exam."

Some students even expressed doubts that they had too much control:

> "It's great, but for first year students, I believe, there is too much freedom."

> "[..] learning harder competencies properly is sort of discouraged as succeeding at easier ones is much much easier.''

Answering these doubts, we trust that students are mature enough to choose what grade they want. As for focusing on simpler material, students commonly do this in the traditional grading as well, maybe unconsciously. We believe that no system can eliminate such slacking completely, but the advantage of our system is that the slacking is a transparent and conscious choice.

Reaching out to struggling students

While the course was running, David Clark, a co-author of [1], published blogpost [6] about how alternative grading helps to support struggling students. It was just on time, because we were approaching the last week of the course!

Inspired by [6], each instructor looked at their group and identified students who started out well, but were getting all N's

Illustration: Mara Chelărescu

Figure 8 Students in control: catching a 9.

lately. In Nelly's group there were 7 such students out of 75. We wrote an e-mail to them, see Figure 6. Importantly, the subject said `Reaching out' so that our intention to help was immediately clear. Not all students replied, but some did, and some came to the instruction for help. To those who responded, we suggested to make practice quizzes and gave them feedback. Students appreciated this. Here is what one student wrote:

> "[...] thanks a million for the feedback, if I do pass the quiz tomorrow, and the course in general it is not lost on me that this correspondence has been instrumental for it."

This student was just one G short of passing after the exam, and has passed the course after the resit.

What we will improve next year

Better explanation of the system. A common complaint of the students was that the new system was hard to understand:

> "I think the structure with the competencies and weekly tests S and G system is very confusing for 1st year students to understand.''

> "[...] this system with competencies was more than confusing for the vast majority of students, I saw questions being asked even on week 8."

At some point, even a study advisor wrote to us asking for explanations! In week 8, right before exam, the most common question was: `For the exam requirement, can I retake quizzes where I already have an S?' (The answer is, yes!)

We now have a lot of feedback from the study advisor and the students, and we will pay special attention to explain the system better next year.

Revising competencies Students experienced that C1-4 were too easy, and C7 was far too hard. Also, students often felt that the emphasis on open questions was too strong:

> "[...] the written tests take disproportionally more time, which makes taking them extra stressful.''

Students also didn't like that there was no weekly quiz on C15, 16, as this material was offered in the last week.

Next year we will combine some competencies and split others to make them more equal in difficulty. We will also reduce the number of competencies from 16 to 14 so that we can give quizzes on the last 2 competencies in week 8.

Larger question bank At the moment, we have exactly the minimal number of questions to fill all quizzes. We want to extend the question bank, and possibly choose questions at random from the question bank.

Humane alternative grading

Traditional exams bring upon us, teachers, the uninspired `will-it-be-on-the-exam' questions and dreadful grading.

But students suffer even more! They feel that their entire future depends on a `random exam'. Only two weeks ago, in another course, Nelly observed a student who was explaining each problem, bright and clear, to their friends. But at the intermediate test, speeding through 8 problems in 45 minutes, they read two questions incorrectly, and scored only 6/8. It was soul crashing for the young person. There is no good reason why this student couldn't try again.

The new system felt as a great relief because we could set high standards (good write-up; 100% correct answers), and yet be more relaxed and sympathetic to our 300 anxious first-year students.

We were especially happy to read this very rewarding testimony:

> "Thanks for being human. I wish that other courses would treat us in the same way as we were treated in this course.''

Acknowledgement

Last, but not least, we want to thank all the wonderful people that have made the course a success. First, we thank the remaining three instructors Vinay Kumar Bindiganavile Ramadas, Fabian Burghart and Florian Henning, who have created all digital tests with Nelly. We are also very grateful for the enthusiastic and reliable student assistants of the course, Pascal Otjens, Andrei Pătularu, Diego Rivera Garrido, Floortje van Santen, Georgi Tsonkov. Finally, we would like to thank our colleagues from the computer science domain as well as the course support at TU/e, who have provided invaluable help and support throughout the course.

References

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

2 Litvak, N., & Weedage, L. (2023). Do we teach what we preach? *Nieuw Archief voor Wiskunde*, December 2023, pp. 233-238.

3 Yeager, D. (2024). 10 to 25: The Science of Motivating Young People. *Random House*.

4 Litvak, N., & Kula, F. (2023). The prior knowledge problem. *Nieuw Archief voor Wiskunde*, September 2024, pp. 161-166.

5 Litvak, N (2024). How students learn. *Nieuw Archief voor Wiskunde*, June 2024, pp. 112-116 .

6 Clark, D. (2024) How alternative grading helps me support struggling students (*gradingforgrowth.com*).

7 Haidt, J. (2024). *The anxious generation: How the great rewiring of childhood is causing an epidemic of mental illness*. Random House.