
Nelly Litvak


Michel van Etten


John van der Schaaf


Georgios Skantzaris


Lucas van Kreveld

*Department of Mathematics and Computer Science*
*Eindhoven University of Technology (TU/e)*
*n.v.litvak@tue.nl*

*Department of Chemical Engineering & Chemistry*

**Column** Better than blackboard

# Digital testing: The chemistry experiment

**In this column 'Better than blackboard' Nelly Litvak writes about teaching mathematics at university. She will address problems that many university teachers face.**

For some time now I've been imagining this exam set-up. Take one large room, assign invigilators from the morning till the evening, and let students book slots for exams. A student takes a seat, logs in, and then they may take any exam assigned to them. If they fail, they can simply book another slot and try again. Of course, there are many logistics and administrative problems in arranging exams this way, so the idea remained on my close-to-impossible wish list. Until in summer 2023, I was asked to teach Statistics for the first-year Chemical Engineering & Chemistry (CE&C) students at the TU/e. I was told that the CE&C program took a very unusual approach to exams, and since I was into teaching innovations, I would probably like it. I agreed, and very soon I had a call with Michel van Etten from the Chemical Engineering department. I was listening to him and couldn't believe my ears: it was exactly my dream exam set-up! They are actually doing it!

Welcome to the new issue of `Better than Blackboard'! I am honored to introduce my co-authors. John van der Schaaf and Michel van Etten are chemical engineers responsible for the new assessment system at the CE&C program. Georgios Skantzaris and Lucas van Kreveld are mathematicians; Georgios teaches Calculus at CE&C; Lucas teaches Statistics at CE&C, and last year we taught it together.

Today we will tell you the story about the new approach to exams at the CE&C program at the TU/e, with colorful illustrations by student artist Mara Chelărescu; check out more of her art at https://cara.app/vinyllaroll.

**It started with one difficult course**

In 2014, John took over the course Chemical Reactor Engineering. This is a famously challenging course, it integrates the knowledge of advanced calculus, chemical kinetics, catalysis, physical transport phenomena, thermodynamics, and reactor engineering.

The course had the conventional set-up: classical lectures, guided self-study with exercises, one intermediate written test and a final written exam. As often happens in this set-up, the attendance to guided self-study was low. Students were not prepared and were copying solutions from the teacher or peers, to study later. This learning strategy is disastrously ineffective. Even if a student understands a solution of others, it doesn't mean they can solve the problem; the only way to really learn is to try to solve a problem yourself!

No surprise that the passing rate was historically around a dismal 30%. Students were approaching John at the start of the course on the verge of tears, stressed out that they were not (ever) going to pass this course. This was heartbreaking. John had to do something about it.

After initial attempts with digitized testing, John replaced written exams with digital tests in 2017. For that, he used the Moodle-based OnCourse platform that was developed by Hans Cuypers at the department of Mathematics and Computer Science for Calculus, in combination with a BYOD (Bring Your Own Device) authentic testing system.

In John's new exam set-up, students had two timeslots per week to assess themselves on a module by making a digital test in a protected exam environment. During the test, students could get feedback whether their numerical answer was right or wrong, and could correct the answer without penalty; the time spent on finding the error in their work-out is already enough penalty on its own.

The course consisted of five modules. Access to the next module was only given if the preceding module test was passed with a score of at least 50%. Students could retake a test to reach a passing score

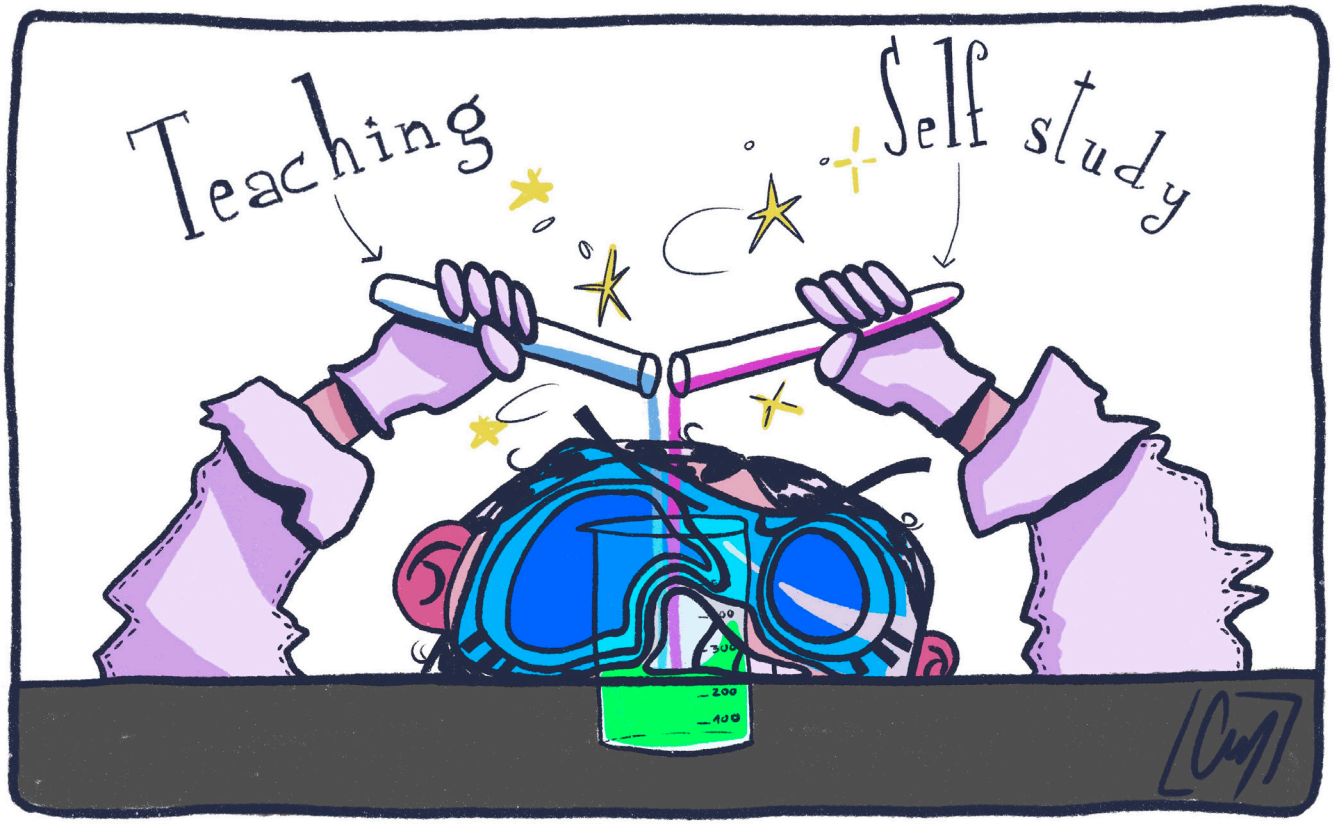


Illustration: Mara Chelărescu

The Chemistry of education.

but also to improve their grade.

John also stopped giving live lectures. He recorded all theoretical content in the studio, and added Lightboard recordings of blackboard exercises. Students could progress at their own pace through the course, and test their knowledge digitally with representative exercises. If students had questions on the theory or on the systematic approach to a problem, they could discuss it with John during the scheduled guided self-study hours.

The transformation was extraordinary. Students could test themselves from the first week on and were engaged from the start. Meaningful and advanced questions on the theory and systematic problem solving were the topics of the guided self-study.

John also saw how students gained much more agency and control over their study. Some students needed more time in the first weeks of the course to grasp the concepts and accelerated later to finish in the designated 10 weeks. Some students were able to finish the course already in week 5. Peer groups formed naturally to study the content, delegates from these groups came to John with specific questions they could not solve themselves, and shared the feedback later.

The passing rate increased to 70%. Out of those who failed, around 80% were in fact (auto-)enrolled in the course, but never started studying for the course. This makes the success rate for the committed students even higher, over 90%. As a teacher, it was rewarding to see a struggling student connecting the dots and passing a test, and it was equally rewarding to help a brilliant student with a specific tricky question that they missed at the test; the student retook the test and improved the score from 95% to 100%.

Now when students were stressed out at the start of the course, John could confidently say that if they engage with the course, study, practice, ask feedback and test themselves, they will succeed.

Inspired by this convincing success and rewarding experience, John, in his role of program director, decided to implement the new exam set-up, wherever suitable, at the entire Chemical Engineering & Chemistry (CE&E) program. This has triggered the pilot that we will tell you about.

**DIAMS tests and how they work**

In September 2023, all first-year courses in the CE&C program started using the new digital assessment system, called DIAMS tests (DIAMS = Digital Interim Assessment Moment with STEP (Secure Test Environment Protocol)).

Here is how it works. Each course is divided into modules. A module usually consists of a collection of coherent topics that build up to 1-2 ECTS. A databank of test questions is available for each module in a digital assessment system (CE&C program uses the system called SOWISO).

When a student is ready to take a test, they make a reservation for a one-hour DIAMS session in a calendar within the Learning Management System (TU/e uses Canvas). It does not matter which test a student is going to make, the reservation is just for a seat in the DIAMS venue. The students can make reservations during the period of 8 weeks. There are three DIAMS sessions per day (in the afternoon)

in weeks 1-7, and six DIAMS sessions per day (morning and afternoon) in week 8, which is the week where no new material is taught in any course.

At the DIAMS venue, an invigilator is always available. A student brings their own laptop and receives a STEP USB stick. They boot their laptop with the STEP stick, which gives them unique access to the server where they can make a test of their choice. The STEP stick blocks communication outside of the server, and provides a calculator to use during the tests.

When a student selects a test to start, the timer of one hour starts running. The test is then generated from a question bank; the questions are selected from the question bank at random.

The questions are always different because the question banks are large and because the degree of randomization in SOWISO is sky high. Each problem has at least three to four randomized variables, which can be a number, but also a formula or a phrase. For instance, there can be a variable that takes values `greater than', `smaller than', `equal to', etc.

The answers mostly are either a number or a formula. After inserting the answer, a student can check whether the answer is correct, and make a correction. This is very important, because it prevents the all-or-nothing situation when a student may fail due to a small computational or rounding error. Other types of questions (multiple choice, drag and drop, ranking, dropdown) are sometimes used as well, but for these types of questions students cannot check their answer before submitting the test.

All DIAMS tests are automatically graded. At the end of one hour, the test closes automatically, and the result is immediately known. The students hand in their STEP stick and their scrap paper. If a student didn't pass the test, they can make a new reservation.

**Figure 1** DIAMS venue

The next attempt of the same test can only be done after 36 hours. At the beginning, the students could take a new attempt after 15 minutes, but we noticed that the students were sometimes trying the same test multiple times in one afternoon. Usually, when students did not pass the first attempts in that afternoon, they also did not pass the second and/or third attempts. The mandatory 36 hours prevent such prize shooting. Instead, students get the time to reflect, ask for feedback, and prepare for the next attempt.

Since students do different tests at different times, there is no teacher or other content expert at the DIAMS venue. This is not a problem, because the students can retake the tests almost unlimitedly, and because the results can be discussed with the teacher and eventual mistakes can be fixed afterwards.

Especially when a question bank is used for the first time, there is a small chance that there is a mistake in the test. But later on, in almost all cases when a student suspects an error in the test, it is actually the students' own mistake.

**Creating question banks**

Any digital assessment scheme stands and falls with large, highly randomized, and high quality question banks. Creating such question banks is a major initial investment and therefore a major bottleneck of converting to digital tests.

The CE&C program hired student assistants to create DIAMS question banks for each module.

Statistics had only two modules. Most questions were initially inspired by exercises in a textbook [1]. We were lucky with a talented student assistant, who is doing a double-degree in chemistry and mathematics. He created many interesting questions, mostly based on chemistry-related examples.

The Calculus course is larger, it has four modules. For the question banks, Georgios provided past exam exercises (start-test, mid-terms and finals) of the previous Calculus course, and used exercises from classic Calculus books [1,2].

Besides the enormous effort of student assistants, teachers had to invest a lot of time in checking, debugging and correcting errors, too. This was much work, but we found it more rewarding and future-proof than composing and grading a traditional written exam.

**Assessment structure with DIAMS tests**

For most courses, the exercises were categorized per topic and assigned points proportional to their difficulty. Lastly, the four DIAMS exams of one hour were created over the span of all the topics of the course and included easy, medium and difficult exercises.

DIAMS tests of a module usually contain at least five different questions. The students must score at least 50%. For DIAMS tests that assess learning objectives at their final level of a course, there is usually a mix of about 40% easy questions, and about 60% more difficult questions. As the students have to score at least 50%, they must answer at least one difficult question correctly.

In some courses, the DIAMS tests do not assess the learning objectives at their final level, and are mainly used for encouraging the students to remain on track with the material, and to assess the basics of the course. Sometimes, a minimum grade is required, but not necessarily. This is up to the responsible teacher of the course.

**DIAMS tests assess what we want to assess**

One may argue that DIAMS tests are easier for the students, but we would say no.

For instance, for Calculus, the students must pass all four DIAMS tests with at least a 5.0, so they cannot skip specific modules, which would be possible in a traditional system with a final written exam.

The spectrum of different exercises in four DIAMS tests is wider than in one traditional exam. Since students often make several attempts, eventually they are prepared to solve diverse questions.

Digital assessment is quite strict. A student gets points only if their answer is fully correct. There are no partial points for correct intermediate steps. Thus, students get trained in not making small mistakes and in finding these mistakes.

Moreover, numerical answers in DIAMS tests often require high precision. For example, the conversion of a reaction needs to be calculated with a precision of 0.001. Even if a student can guess approximately what the answer should be and check correctness, it is very unlikely to get all decimals right without actually calculating the answer.

Due to the high randomization, the formulas and numbers are different every time, so only the systematic problem approach can be memorized/learned, which is exactly what the students should learn. Scoring 50% for the module is almost impossible without understanding the material.

In some rare cases a student could be lucky with a set of random questions. But no student can be so lucky that they pass a course due to sheer chance. Rather, they must genuinely study and develop their skills before they can demonstrate proficiency in DIAMS tests.

One may fairly note that DIAMS tests don't allow to assess the students on how they present solutions. Yes. However, in the traditional basic math exam, such as first year Calculus, we haven't truly been grading them for the quality of writing, nor training them for that. So in this sense, the traditional exam doesn't differ much from DIAMS. For the sake of evaluation, Georgios looked at the students' Calculus scrap papers, and did not observe anything unusual. Some students had excellent write-up, even on a scrap paper; some wrote messy solutions that still made sense; and others were totally off and didn't pass the digital test either.

As it should be in a thorough pilot, everyone involved is very alert and reflective about possible shortcomings and pitfalls. That said, based on our experience and pure logic, we do believe that the DIAMS way of testing is much more transparent and accurate than a traditional written exam.

**Revisions without penalty**

The high grades and passing rates with DIAMS tests are not surprising, because students can redo the tests as many times as they want.

Revisions without penalty are only fair, and are one of the pillars of alternative grading [6]. We discussed this in our previous column [5] and a recent blogpost [7]. As book [6] explains in depth, revisions without penalty mirror the natural learning process: *get new knowledge - try it out - fail - get feedback - try again*.

Revisions without penalty remove the unnecessary stress typical for the `all-or-nothing' exam. This is a great relief for the students! Maybe it is a little bit early, but we have heard from students that they do feel stress because of DIAMS, but that it is a positive stress, as they have time to repair their actions throughout the term. Students also feel that they are in charge of what they want to do when, and do it when they are ready, within the set timeframe of eight weeks. Both positive stress and agency are key to successfully mentoring and educating young people [8].

The traditional manual assessment of basic skills deprives our students from multiple attempts for no good reason except the limited grading capacity. This is a very strong argument to use digital testing and automatic grading whenever possible.

**Calculus**

The passing rate for Calculus variant B (same as CE&C has) was around 45% over the last four years, with average grade of 5.2. For the first iteration of the DIAMS tests, the passing rate increased to 77%, with an average grade of 5.5.

Between the first and the second iteration of DIAMS, there were two changes. First, we now require 36 hours waiting time between two consecutive attempts for the same test, as mentioned above. Second, we have provided practice tests, which are drawn from the same question banks, but are never the same as DIAMS questions due to the high degree of randomization and the large size of the question bank. Now students can practice more than enough at the right level.

After the second iteration of the course, the passing rate has increased even a little bit more to 80%, and the average final grade increased significantly to 6.4.

Keep in mind that the DIAMS tests are not easier than a traditional written exam (students must pass four DIAMS with at least 5.0; only fully correct answers count; larger variety of problems). It is even safe to say that four times ≥ 5.0 for DIAMS tests with a weighted average of 5.5 is a much stronger result than a usual 5.5 on a final exam covering all topics. Most likely, the average 5.5 at the first iteration of DIAMS tests is already a significant improvement compared to the average 5.2 over the last four years in the traditional system.

One may say, we must be careful comparing grades received in two completely different ways. This question is of course on the radar of CE&C. Full evaluation/research on the pilot is currently ongoing, and carried out by a postdoc from the Eindhoven School of Education.

Besides the higher grades, students appear engaged from day one. They are eager to take the test again and again, to pass, but also to improve their grade. This already makes them practice more and become more skillful in Calculus. Moreover, students study the upcoming material, and exhaust the study resources in order to succeed earlier. Especially mathematically strong students and those who already studied calculus at high school can pass the course in the early weeks, and may improve their grade later if they wish.

One may worry that the course will turn into fully self-study, but Georgios observed no change in attendance of lectures and guided self-studies, compared to the previous years.

The initial investment in question banks was demanding, but other than that, Georgios does not see any disadvantages of DIAMS compared to the traditional testing methods that he used before.

**Statistics**

The Statistics course had only two modules: three weeks in quarter 1 and four weeks in quarter 4. We had one (interactive) lecture and one self-study per week. The Statistics modules are officially a part of larger courses, and require only a pass. In quarter 1 of 2024/25, some more than half of attempts were successful, and only 11 out of 175 students failed.

The interactive lectures were attended very well. The attendance of self-study was as usual, moderate to low; those students who were present, worked seriously and asked good questions.

As we told above, most problems were formulated in the context of chemistry. This was motivating for the students, but, on the other hand, they could recognize familiar problems by the context. Therefore, Nelly and Lucas divided the problems in two sets: one for self-study, and a very similar one for the DIAMS tests.

A challenge, typical for Statistics, was to prevent that the students blindly execute standard procedures, without understanding the reasoning behind them. Therefore, in DIAMS we included theoretical questions, usually multiple choice with multiple correct answers (for more detail on such questions, see previous columns [4,5]).

Illustration: Mara Chelărescu

Collaborating with technology.

Altogether, Nelly and Lucas share Georgios's positive experience with DIAMS. Even in the first iteration, our work on the question banks already paid off as we did not have to grade this very small module in the middle of term.

**What it takes to change the rules**

In order to start this pilot, many stakeholders needed to be consulted and convinced. This included:

- *Educational board of the TU/e*: Does the pilot fit in the educational vision of the TU/e? Do they approve the Innovation Call proposal to finance the pilot for the first three years?
- *Managers of the Education and Student Affairs*: Is the pilot feasible from the organizational and administration point of view?
- *Digital Assessment Support team of Library and Information Services (LIS)*: Is the pilot feasible from the ICT point of view? Can the pilot be supported by the LIS team?
- *Faculty Board, Faculty Council, Program Committee CE&C, Examination Committee CE&C*: Do they approve and financially support the pilot within the faculty?

The pilot required additional 0.1 fte for administration/organization: scheduling the DIAMS sessions, enrolling students via the Canvas registration page to DIAMS sessions via Canvas Calendar, enrolling students to the test software (SOWISO) and arranging extra time in SOWISO for students with extra facilities.

The room was prepared for DIAMS complete with power facilities, new chairs, Wi-Fi check and amplifier. For the DIAMS sessions themselves, STEP sticks and a dedicated STEP server were bought and made ready, plus loan laptops for in case that STEP would not work with a laptop of some students.

Specific regulations for DIAMS were made, based on the regulations for central final exams, so that the Examination Com-

mittee could act accordingly. These regulations had to be approved by the Program Committee, Faculty Council, and Faculty Board. In this way, the students know their rights, what is expected from them, and what should be arranged by the department.

The DIAMS venue requires about 1 fte of invigilators at the moment. About €120.000 was spent on student assistants, and about €200.000 on PostDocs for development and management of the SOWISO question banks and implementation of DIAMS, excluding the PostDoc at the Eindhoven School of Education for researching the pilot.

The tremendous effort is paying off: the pilot will continue at CE&C for sure. Moreover, CE&C is now in conversation with most other departments at the TU/e that would like to join the pilot (which will not be a pilot anymore then) in academic year 2026/2027. Potentially, one or two departments could join already in 2025/2026. Of course, upscaling will bring new challenges, from room availability to funds. But it looks quite realistic that a DIAMS-like system will be implemented TU/e-wide.

**Collaborating with technology**

When asked what the main barriers for DIAMS tests were, John answered: "Well, there are still barriers. It is particularly challenging to convince critics of the quality of this way of testing; in their minds, nothing beats a written exam".

It is puzzling why the written exam, with its superfluous all-or-nothing stress for the students, and excruciating grading for the teachers, is still so popular. And why automated grading, with revisions without penalty for the students, and no (!) grading for the teachers, is often received with skepticism.

Here is our humble attempt to discern three reasons.

First, possibly, digital testing has a firm association with `multiple choice study'. However, this is a very limited view on automatically graded tests. There is by now a huge variety of very creative automatically graded questions: formula questions, drag-and-drop, matching, finding errors, etc. We can now assess understanding on quite high levels using purely digital tests.

Second, one might think of digital testing in a very binary way: a fully written test or a fully digital test. But it can be both! DIAMS-type tests are especially suitable for learning and assessing basic knowledge and skills, and can be easily combined with assessment of higher level skills by the teacher, such as writing mathematics. We gave two examples of such hybrid set-ups in our previous columns [2,3].

Third, when talking about digital testing, one may think merely of replacing a traditional exam with a digital exam. But this is akin to comparing a simple phone to a smartphone only by convenience of calling. Automatic assessment offers opportunities far beyond grading! More practice, more feedback, revisions without penalty, students in control of their study, are only some of the enormous and immediate advantages.

Meaningful use of technology has been the answer to questions like `can we move faster?' or `can we live longer?'. Similarly, we can teach and assess our students better if we collaborate with technology in a meaningful way. We believe that DIAMS tests are an example of such a meaningful collaboration.

**References**


1 Berendsen, H. J. (2011) *A student's guide to data and error analysis*. Cambridge University Press.

2 Adams, R. A. and Essex, C. (2021) *Calculus: A Complete Course*. 10th ed. Toronto: Pearson (ISBN: 9780135732588)

3 Hass, J. R. , Heil, C. E. , Weir M. D. (2020) *Thomas's Calculus*. 14th Edition in SI Units. Pearson.

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

5 Litvak, N., & Müller, N. (2024) Competency-based grading. *Nieuw Archief voor Wiskunde*, December 2023, pp. 227-233.

6 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.

7 Litvak, N., & Müller, N. (2024) Alternative grading in probability and statistics: How we used competency-based grading for 300 computer science students at a Dutch university. *Blogpost in `Grading for Growth'*. December, 2024.

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