International Journal of Mathematical Education in Science and Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmes20

No exam: assessment of third-year engineering students on the basis of self-generated Statistics cases

Nelly Litvak & Fulya Kula

To cite this article: Nelly Litvak & Fulya Kula (2021): No exam: assessment of third-year engineering students on the basis of self-generated Statistics cases, International Journal of Mathematical Education in Science and Technology, DOI: 10.1080/0020739X.2021.1982041

To link to this article: https://doi.org/10.1080/0020739X.2021.1982041

Published online: 14 Oct 2021.

INTERNATIONAL JOURNAL OF MATHEMATICAL EDUCATION IN SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/0020739X.2021.1982041

CLASSROOM NOTE

No exam: assessment of third-year engineering students on the basis of self-generated Statistics cases

Nelly Litvak [a,b] and Fulya Kula [a]

[a]Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, Netherlands; [b]Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands

ABSTRACT
Through the global pandemic, the single greatest challenge at universities has been the move towards digital assessment. In this classroom note, we describe our new no-exam assessment setup in a Statistics course for third-year bachelor Mechanical Engineering students. The main idea is to assess the students based on self-generated problems. Crucially, we supported students by providing a very clear structure of the course material, learning objectives, and requirements. At the same time, we left sufficient space for the students to tune the final assignment to their interests, creativity, and mathematical skills. In our experience, this setup makes assessment meaningful and enjoyable for both the students and the teacher and does not need to demand excessive time investment on the teacher's side. We strongly believe that approaches like ours will have potential lasting effects on the diversity of assessment, quality of learning, and, last but not least, the appeal of Statistics for future engineers.

ARTICLE HISTORY
Received 5 June 2021

KEYWORDS
Statistics; assessment; remote assessment; tertiary education; mechanical engineering

MSC
Statistics; Statistics assessment; Statistics assessment in engineering

1. Introduction

Through the global pandemic, the move towards a digital assessment has become a paramount issue, with heated discussions on campus splashing out to the media. This issue is particularly prominent in undergraduate mathematics, where the assessment is predominantly based on a standard written test (Iannone, 2020). In this classroom note, we describe a new assessment setup of an introductory Statistics course for the class of 162 third-year undergraduate Mechanical Engineering students at the University of Twente. Instead of giving a traditional exam, the students were asked to create their own real-world Statistics cases.

Previous studies show that the assessment methods based on one-time measures, such as a traditional exam, are non-dynamic and fail to explore the developmental nature of learning (Chance & Garfield, 2002). Accordingly, the research has introduced assessment methods such as writing assignments (Woodard et al., 2020), activity-based statistics learning (Gnanadesikan et al., 1997), and projects (Garfield, 1994). One of the main focus

CONTACT Nelly Litvak n.litvak@utwente.nl Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Hallenweg 19, 7522NH Enschede, The Netherlands

points of the pre-pandemic research stands as the assessment of students' reasoning in Statistics (Garfield & Ben-Zvi, 2008) and calls recommended for this purpose implied the move towards alternative assessment techniques (Chance & Garfield, 2002). This invitation became emergent with the rapid and compulsory transition to online instruction throughout the pandemic. We contribute in this direction with our new assessment setup as a viable solution during the post-pandemic. Beyond providing an attractive digital assessment tool, such setup generates a resource of deeper insight into the students' understanding (Garfield & Ben-Zvi, 2008).

This classroom note is organized as follows. First, in Section 2 we describe the workflow of the course. The core of this article in Section 3, where we explain how we organized the assessment. In Section 4, we provide some examples and reflect on the results. Finally, in Section 4 we discuss the strengths and pitfalls of our approach and its possible lasting effects.

2. The course workflow

Presenting the course material. The course covered the basic material from the definition of probability to the single-sample hypothesis testing, with a workload of 2.5 ECTS, equivalent to 70 hours. We used book (Montgomery & Runger, 2018) and the complete series of short video lectures (Litvak, 2020) (openly available on YouTube). To trigger the students' interest, most examples in the course were specifically designed for the engineering profession. We have divided the material into five *Units*; the Units are listed in Table 1 together with examples of learning objectives that will be explained later.

The students studied the book and the videos independently, at a pace of about one Unit per week. For each Unit, we gave an *interactive (online) class*, where students answered multiple-choice questions, using the interactive platform *Wooclap*. We also held several online Q&A's for individual questions.

Structuring students' practice during the course. We ensured that the students kept practicing throughout the course by introducing *Quizzes* and *Problem Sets*. These activities, described in more detail below, were mandatory but did not affect the grade. The obligation was enabled through self-regulating prerequisites, facilitated automatically through the educational platform *Canvas*, and schematically depicted in Figure 1. For example, Quiz Unit 2 becomes accessible for a student only after Quiz Unit 1 is completed. At the end, the (fully online) Final Assignment becomes accessible only after completing all the prerequisites.

The *Quizzes* consisted of 4–6 standard problems, based on Mechanical Engineering examples, designed by us and not findable on the Internet. The Quizzes were implemented using *formula questions* in Canvas: the students had the same problems with different numerical values, so they could discuss together but the final answers were unique per quiz attempt. We allowed unlimited attempts and required a score of at least 60% to pass the Quiz.

For the *Problem Sets*, we used the design from Mazur (2018) that consists of the *individual phase* at home (the students solve a set of homework problems individually without discussing with others) and the *group phase* in class (the students discuss in small groups, mark their errors, and receive solutions from the teacher). It was mandatory to submit the solutions after the individual phase and to attend the group phase.

Table 1. The units and the examples of learning objectives with corresponding points.

Unit #	Unit title	Learning objective examples	points
1	How to quantify randomness? Probability, mean and variance	1.5 Apply the definition of independent events, and explain your computations	1
		1.12 Compute the variance and the standard deviation using a probability density function, and explain your computations	2
2	What is 6σ in 'lean 6σ'? Probability models	2.10 Reason in which real-life situations the Normal distribution is applied, and explain the meaning of its parameters	2
		2.11 Apply the Normal distribution, and explain your computations	2
3	What an engineer must know about the sample mean? – Linear functions and limit theorems	3.9 Apply the fact that a linear function of normal random variables has a normal distribution	2
		3.10 Recognize the situation where the Central Limit Theorem is applicable, and state when the normal approximation can be used based on this theorem	2
4	Plus-minus how much? Data summaries and confidence intervals	4.2 Compute sample variance and sample standard deviation	1
		4.16 Execute and explain computations for the confidence interval for the population proportion, using the normal distribution	2
5	What does it mean that 'the data is significant'? Hypothesis testing	5.2(b) One-sample t-test for the mean of the normal distribution: Formulate the null-hypothesis and the alternative	2
		5.2(d) One-sample t-test for the mean of the normal distribution: Accept or reject the null-hypothesis based on the appropriate critical region, or apply the notion of p-value	3

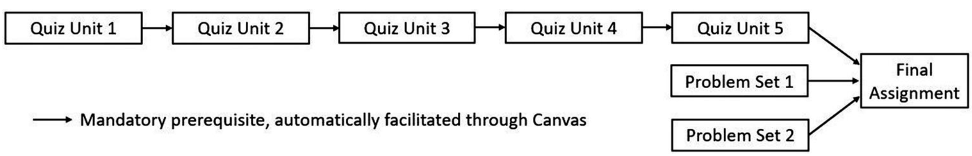

Figure 1. The workflow of the course. The Quizzes are graded automatically. For Problem Sets, only sub-mission and attendance to the group discussion is required.

3. The no-exam assessment setup

The core of this classroom note lies in the Final Assignment replacing the written test. We will first describe the Final Assignment, and then explain the requirements and the grading.

Final Assignment. We asked the students to do the following:

Find own cases from Mechanical Engineering, where the knowledge gained in the Statistics course was useful. The cases could be inspired by other courses, internships, research projects, etc., or other relevant real-world examples.

Formulate clear Statistics problems for the chosen cases. Crucially, we specified the volume and the coverage of the material by asking students to use a given number of the *points of learning objectives*, which we will explain in the next paragraph.

Create complete solutions to these problems by following three steps. We explain these three steps below.

Points of learning objectives. We have formulated very specific learning objectives, several examples are given in Table 1. Each learning objective refers to an exact topic and requires performing either very specific calculations (e.g. 2.11, 5.2(d)) or reasoning (e.g. 2.10, 3.10). One Unit contained about fifteen learning objectives.

To define the volume and the coverage of the Final Assignment, inspired by Faber and Dresscher (n.d.), we assigned 1–3 (usually, 2) *points* to each learning objective, see examples in Table 1. The points are approximately proportional to the corresponding weights we usually give in a standard written test. We required to use a minimum of 25 and a maximum of 30 points of learning objectives in the Final Assignment. In addition, we set minimal required points from each Unit, summing up to 21, and leaving the remaining 4–9 points for students' choice.

The students learned to relate problems to the learning objectives in Quizzes and Problem Sets. In Quizzes, after submission, the list of learning objectives for each problem appeared as a comment, and we explained to the students how to use this as a hint for the correct solution. In Problem Sets, during the discussions in the group phase, the students received a list of learning objectives for each problem and were encouraged to use this information in the discussion. They followed this advice because they wanted to practice for the Final Assignment.

Three steps of the solution. We asked the students to present their solutions in three steps, given in Table 2 below. These steps are inspired by Polya's four steps approach for solving mathematical problems (see e.g. Iannone, 2020). We adapted this approach to Engineering students by combining Polya's steps 1 and 2 and formulating the steps in a less abstract and more actionable way.

The students practiced the three steps in the Problem Sets. As in Mazur (2018), in the individual phase, the evidence of effort was that all three steps were worked out in all problems. As an example, the teacher presented the solutions using the three steps as well.

Grading using rubrics. The criteria that we used for grading are provided in Table 3. Importantly, we required that *each* criterion is met with *at least* 50% quality. For example, if Step 3 is missing in half of the problems, then the student gets a *zero* for this criterion. Moreover, the rule was that *failing one criterion resulted in failing the assignment*. This choice is motivated by our previous experience that the students in this course often aimed for a pass rather than a high grade. We wanted to encourage completeness of the work and to avoid that the students strategically skipped some parts to score just enough for a pass.

Table 2. The three steps of the solution.

Step 1	Explain in words how you would approach the problem: what needs to be computed? Which probability models are relevant and how? Which method will you apply (e.g. which statistical test) or which result will you apply (e.g. Central Limit Theorem)? Which steps will bring you to the answer?
Step 2	Solve the problem: Perform the computations as you described in Step 1.
Step 3	Look back at your solution and the answer: Is the answer reasonable/expected? Why or why not? If the answer does not seem correct, what could have gone wrong? What difficulties did you experience solving the problem? Which learning objectives are relevant for solving this problem?

Table 3. Grading criteria and their weights.

Criterion	%
The originality and the quality of the problem formulations	15%
The match between the problem formulations and the stated learning objectives	20%
The completeness and correctness of Step 1	20%
The completeness and correctness of Step 2	20%
The completeness and correctness of Step 3	10%
Video[a]	15%

[a]We have asked the students to upload a 4-minutes video where they explain their assignment. This is an unusual and somewhat controversial part of the assignment, and we will refer to this in Section 4.

To be consistent in the grading, the two teachers first graded two randomly chosen assignments together. Then, each student was graded by one of the teachers, and we have discussed questionable cases when needed.

4. Results

Quality of submissions. The cases were very interesting and varied, from material science and renewable energy to artificial intelligence and digital twins. Many assignments referred to recent engineering literature, see example in Figure 2 by student T. Edelijn. As customary in the Netherlands, we offered two attempts. After the first attempt, taken by 144 out of the 162 students, only 9 students failed to meet the passing grade of 55%; the average score was 75.57%, with a standard deviation 15.35%.

Importantly, the students did not make typical errors that we observed in the previous years when we used standard formative written tests! For example, the greatest majority of the students formulated the null – and the alternative hypothesis correctly and applied the correct test. In the previous years, these were sources of very common errors, even when the students used a formula sheet.

Experience with the learning objectives. The volume and the coverage of the students' assignments were as we intended it to be, that is, comparable to the standard written test in previous years. Since the points of learning objectives (see Table 1) were the only method we used to regulate the volume and the coverage, we conclude that this method was effective for this purpose.

> Fused deposition modelling (FDM) is a form of 3D printing which is getting more and more attention. Shape memory materials highly rely on the FDM process to achieve an accurate structure. The role of operational parameters as printing velocity and layer height is investigated by Barletta et al.[a]. They found that the layer height greatly influences the compression load the structure could endure. Therefore, accuracy is of the highest importance. The manufacturer of the machine claims that the accuracy of each layer is within 20 microns in 95% of the cases. Some students checked all individual layers of their 3D printed product and found: sample mean 212, sample variance 31, and n = 200. Is there enough statistical evidence to substantiate this claim?

[a] M. Barletta et al., 4D printing of shape memory polylactic acid (PLA) components: Investigating the role of the operational parameters in fused deposition modelling (FDM) , 2021.

Figure 2. An example of a problem generated by a student. The solution uses the χ^2 test for the variance of the normal distribution.

Another effect of the learning objectives was shifting the emphasis from computations to methodology. In written tests, in our experience, the students were very focused on calculations and did not pay attention to which technique they used and why. Now they were required to state the technique explicitly. This introduced an extra reflection step for the students and gave us the tool to assess the depth of their understanding. For example, a common error in the assignment was claiming learning objective 3.10 instead of 3.9 (see Table 1). This is a subtle error because calculations are the same in both cases. We used this to distinguish between excellent and average work, and could easily justify to the students that correct calculations did not necessarily result in a high grade.

Experience with the three steps of the solution (see Table 2). In most of the students' submissions, the solutions presented in three steps were clear, complete, and pleasant to read. We allowed some freedom of interpretation and did not reduce the score if, for example, information that belongs in Step 1 was included in Step 2.

We noticed that Step 3, the reflection on the answer, was hard for the students. Very commonly they wrote 'the answer makes sense' but their explanation did not reflect on the methodology. Also, there were some typical errors in Step 3, for example, interpretation of a confidence interval as if it was a prediction interval (we note that prediction intervals were not included in the course).

Overall, this format of the solution gave us one more tool to differentiate between excellent and average work by the quality of their arguments.

Experience with the rubrics. We found grading with rubrics as in Table 3 very suitable for this assignment because the resulting grades matched our general impression of the students' work. No student rendered their grade unfairly low and the work that received high grades was obviously of excellent quality.

In the criterion '*The originality and the quality of the problem formulation*' we considered 'originality' to be high when the content of the problems was based on own experience, and had little overlap with Quizzes and Problem Sets, while 'problem formulation' was required to be unambiguous and formally correct. We noticed that many students scored high on 'originality' and low on 'problem formulation', or the other way around. In the future, we will consider splitting these into two different criteria.

As discussed above, the criterion 'The match between the problem formulations and the stated learning objectives' was very informative for assessing a deeper understanding of statistical techniques. However, grading for this criterion was very time-consuming because the vast majority of the students identified at least some learning objectives incorrectly, which resulted in complicated calculations of their scores. In the future, we will keep this criterion but with a simplified version for scoring.

The *Video*, we believe, is potentially a powerful assessment tool, but it was of limited added value in this setup because most of the students simply summarized what already was written in the assignment and did not demonstrate any additional understanding in the video. Besides, some students strongly disliked the video. We will reconsider how to use the video, if at all, in the future.

Students' feedback. Based on the panel evaluation report, open comments in the standard course evaluations, and informal feedback from the students, they appreciated the assignment-based assessment but reported that it was too much work. For this reason, some of them said to have preferred a written test. We agree about reducing the workload in the future and can do so without damaging the comprehensiveness of the assessment.

We believe, however, that one reason for the perceived high workload was that the students learned better, which is evidenced, for example, by mostly correct application and explanation of methodology. Clearly, such deeper learning takes more time! We are very pleased that the assignment motivated the students to spend time on the course. Several students said that they enjoyed working on the assignment. We quote one of them: *I think that this form of examination is a neat way to make a student understand the content of the course without leading to high-stress levels and therefore found it a nice project to do*.

The students were very positive about the online activities of the course: the video lectures, the Quizzes, and the interactive classes.

In this classroom note, we presented qualitative results of our no-exam assessment setup. We close by discussing the strengths, weaknesses, and modifications of this approach, as well as its potential use in post-pandemic education.

Strengths, Weaknesses, Modifications. The strengths of this no-exam assessment setup are multidimensional, below we list those that we found most prominent.

- The emphasis was less on the recall of statistical procedures and more on exhibiting competence due to the clear focus on methodology rather than calculations, also accepted by the students.
- We strongly witnessed students' gains in terms of learning demonstrated by avoiding typical errors that occur in a written test.
- We observed students' positive attitudes towards Statistics in that many of them enjoyed creating problems close to their interests.
- Students' gradual work on the assignment, as well Quizzes and Problem Sets, created a spacing effect that is shown to produce better learning outcomes (Kornell, 2009).
- The students could plan their work and did not have the stress of a time-constrained test.
- The specification of learning objectives and the three steps of the solution were effective tools for assessing the depth of the students' understanding.
- Finally, the student-generated problems are a rich source for teachers to link Statistics with the engineering profession, invaluable for future courses.

As in every new initiative, there are points of improvement, some listed below.

- The students report that the workload of the Final Assignment was too high. A balance must be found between the workload and the comprehensiveness of the assessment.
- The grading of the students' Final Assignment took more time than the grading of a written test. The scoring systems must be simplified where possible.
- As usual with an essay/project type of assessment, plagiarism represents a potential limitation. We did not encounter plagiarism in our course, but this is a point of attention in the future.

Applications in post-pandemic education. We strongly believe that our no-exam setup is applicable in the future due to many benefits outlined above. Moreover, this setup is credible because it is been developed through years on the synergy of ideas and materials (e.g. (Mazur, 2018; Faber & Dresscher, n.d.)). The pandemic was merely catalysis for this work.

For the teachers who want to apply our no-exam assessment setup, we stress that the change in the assessment method cannot be considered separately from the ways of teaching. The assessment and the teaching described in this article are strongly aligned and connected. For example, even if the students cover only a fraction of the material in the Final Assignment, we ensure that they have practiced structurally throughout the course with all topics in the mandatory Quizzes and Problem Sets. Therefore, we would like to offer this entire setup as a road map. Some choices are specific to the teacher and the education programme, for example, the definition of learning objectives and the points assigned to them, as well as criteria and weights in the rubrics. Naturally, defining all the specifics for this type of assessment, together with a well-thought and designed course structure, requires a lot of work. However, once this setup is prepared, it can be used for many years.

Finally, in 2021 the course was given exclusively online. In the post-pandemic education, we will certainly hold the discussions in small groups (the group phase of the Problem Sets) on campus and would prefer to have interactive classes and Q&A's on campus as well. Nevertheless, the online tools such as video lectures, Quizzes, and digital platforms proved very useful, enabling, exciting, and greatly appreciated by the students. We will definitely continue to rely on these digital means in the future.

Acknowledgments

We thank Nico Rutten and Chantal Scholten for their great help in setting up the course. We are grateful to the Mechanical Engineering program, in particular Lisa Gommer and Genie Stoffels, for their support of our new assessment methods. We thank Erik Faber for sharing his ideas of assessment with us. Finally, we thank the students for their deliberate work on the course and their cooperation in exploring new assessment methods.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Nelly Litvak https://orcid.org/0000-0002-6750-3484
Fulya Kula http://orcid.org/0000-0003-0367-1099

References

Chance, B. L., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal* [Online], *1*(2), 38–41.

Faber, E. J., & Dresscher, D. (n.d.). *Introduction to physical systems*. https://osiris.utwente.nl/student/OnderwijsCatalogusSelect.do?selectie = cursus&cursus=202000968&collegejaar=2020&taal = en.

Garfield, J. B. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education* [Online], *2*(1). https://doi.org/10.1080/10691898.1994.11910462

Garfield, J. B., & Ben-Zvi, D. (2008). Assessment in statistics education. In J. B. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting re-search and teaching practice* (pp. 65–90). Springer.

Gnanadesikan, M., Scheaffer, R., Watkins, A., & Witmer, J. (1997). An activity-based statistics course. *Journal of Statistics Education*, *5*(2).

Iannone, P. (2020). Assessing Mathematics at University: Covid-and beyond. *Research Gate preprint*. https://www.researchgate.net/publication/342765751.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(9), 1297–1317. https://doi.org/10.1002/acp.1537

Litvak, N. (2020). Statistics: Why these formulas make sense? *YouTube* playlist. https://www.youtube.com/playlist?list=PLMc0PsvITOku19afGWLhtA63gWPZpWBuK.

Mazur, E. (2018). *Applied Physics 50A Syllabus*. Retrieved July 30, 2021, from https://canvas.harvard.edu/courses/29611.

Montgomery, D. C., & Runger, G. C. (2018). *Applied statistics and probability for engineers* (7th ed.). Wiley. ISBN 9781119585596.

Woodard, V., Lee, H., & Woodard, R. (2020). Writing assignments to assess statistical thinking. *Journal of Statistics Education*, *28*(1), 32–44. https://doi.org/10.1080/10691898.2019.1696257