Research Paper

Recommended citation: Immonen, P., Hirvonen, J., Äijälä, M., Ratava, J., Kuosa, M., & Kaarakka, T. (2025). Opportunity Makes a Cheat? - The Effect of Invigilation on the Results of a Mathematics Basic Skills Test. In Kangaslampi, R., Langie, G., Järvinen, H.-M., & Nagy, B. (Eds.), SEFI 53rd Annual Conference. European Society for Engineering Education (SEFI), Tampere, Finland. DOI: 10.5281/zenodo.17631262.

OPPORTUNITY MAKES A CHEAT? -

THE EFFECT OF INVIGILATION ON THE RESULTS OF A MATHEMATICS BASIC SKILLS TEST

P. Immonen [a,1], J. Hirvonen [b], M. Äijälä [c],
J. Ratava [d], M. Kuosa [e], T. Kaarakka [f]

[a] LUT University, Lappeenranta, Finland, 0000-0002-3286-6840
[b] Tampere University, Tampere, Finland, 0000-0003-1667-2760
[c] LUT University, Lappeenranta, Finland, 0000-0002-6626-4207
[d] LUT University, Lappeenranta, Finland, 0000-0002-8816-6165
[e] LUT University, Lappeenranta, Finland, 0000-0002-8055-1691
[f] Tampere University, Tampere, Finland, 0000-0002-3824-8939

***Conference Key Areas**: Teaching mathematics and physics in engineering education, Digital tools and AI in engineering education*
***Keywords**: Mathematics basic skills, Invigilation, Proctoring, Cheating*

ABSTRACT

The main objective of this article is to examine how non-invigilated and invigilated testing conditions affect the results of a basic mathematics skills test of first-year engineering students in the Finnish education system. We also examine which subject areas of invigilated testing particularly affect the results. The data used in the study consists of tests organised at two Finnish universities in 2023 and 2024.

The mean test result is significantly higher in the non-invigilated test than in the invigilated test, both done electronically. The difference in points between the tests was statistically significant in all other questions except for the tasks related to differentiation, and the conceptual understanding of inverse and reciprocal. The most significant decrease in invigilated test points was in the most difficult but easily cheatable tasks related to solving equations, especially trigonometric and exponential equations, and in a task that required calculating a definite integral.

[1] *P. Immonen*
paula.immonen@lut.fi

1 INTRODUCTION

1.1 Reliability of assessment in online exams

Academic dishonesty is likely a phenomenon older than universities. One of its prevalent forms is cheating in academic studies. Therefore, it should come as no surprise that the recent decades of transitioning studies into the electronic format and online environment have brought the age of "e-cheating". There are many drivers and benefits of organising education online, but can electronic assessment, specifically the non-invigilated variety of online assessments, be seen as valid, or should we consider it inherently compromised? Whether to trust such assessments seems to be the question for many higher education teachers and staff in recent years. This short paper will contribute to the available data and discussion by comparing invigilated and non-invigilated electronic skill tests in engineering mathematics education at two Finnish universities.

Previous studies (Chan & Ahn, 2023) have found that traditional, invigilated exams provide the best assessment of student learning. Still, the results can be influenced by test anxiety and students' beliefs about their abilities. An earlier literature review on take-home exams by Bengtsson (2019) additionally suggests that where take-home exams can excel in testing higher-order cognitive skills, preparing for in-class exams can lead to superficial learning. In-class exams do not match well with the pedagogical theory of constructive alignment (Bengtsson 2019). However, similarly to Chan and Ahn (2023), the most often cited drawback of take-home exams was that they are easily compromised by unethical student behaviour.

According to the study by Chan and Ahn (2023), conducted during COVID-19 restrictions, non-invigilated online exams can provide meaningful information about student learning and do not lead to significant cheating based on the scores being consistent with those of invigilated exams. Another review on academic integrity in online assessment by Holden et al. (2021) paints a more inconsistent picture. In 4 out of 8 studies, higher fractions of cheating were reported in online studies; in two studies, the cheating was similar online and on-site; and in two studies the (self-)reported cheating was lower in online activities. The review puts the number of students who engage in online "e-cheating" at 29 - 41% of the participants, whereas for "traditional" classes, the reported figures were slightly lower, at 21 - 32%.

The availability of math solver applications exacerbates the tendency to cheat on homework and tests taken at home. Sloan-Lynch et al. (2022) have noted that these tools facilitate cheating in mathematics by allowing students to bypass the need for conceptual understanding. Students who do not understand the course content might be more likely to cheat in the course's online exams. Noorbehbahani et al. (2022) suggest that one way of reducing motivation to cheat is by taking the students' learning styles better into account.

Many students encounter significant difficulties with basic concepts, such as rational functions. In a systematic review, Lima et al. (2019) identified fractions, functions, and linear equations as the most frequently mentioned problematic areas in higher education mathematics studies. They also noted issues with rational expressions, logarithms, trigonometry, geometry, propositions, decimal numbers, and mathematical notations (Lima et al., 2019). According to Metsämuuronen and Tuohilampi (2017), at the end of Finnish upper secondary school, the students' proficiency is strongest in algebra and functions, and weaker in geometry and statistics. In addition, Laatikainen (2022) found that proficiency in integral tasks is generally weaker than in other tasks in the Finnish matriculation exam. Lukin (2023) noted that, for example, calculating definite integrals without electronic devices is considered challenging for Finnish upper secondary school students. Mathematics Basic Skills Test results from 2006-2017 were analysed by Myllykoski et al. (2018). The weakest competencies were found in trigonometry, differentiation, and logarithms test questions. Challenges in these topics, which are crucial in university-level mathematics, may contribute to the propensity to cheat in take-home tests.

There are many aspects to consider when interpreting individual results of both non-invigilated and invigilated tests, including math anxiety. Numerous studies have indicated that math anxiety might cause exam underperformance (e.g., Ashcraft & Moore, 2009). According to the old research theory of the *fraud triangle*, three components are needed for fraud or cheating in an educational setting. These components include the opportunity, the incentive, the pressure or need to cheat, and the rationalisation or attitude that justifies cheating to oneself (e.g., Holden et al., 2021). While the opportunity is typically available in online tests, the pressure to perform may conversely be lower than in on-site exams (Holden et al., 2021). As online exams are often perceived as easy to cheat in, a student can understandably rationalise that cheating is prevalent among peers, regardless of whether it is true.

1.2 Mathematics Basic Skills Test

Tampere University of Technology began analysing the mathematical skills of incoming engineering students with the Mathematics Basic Skills Test (BST) as early as 2002 (Joutsenlahti et al., 2016). The original BST has been modernised and is still used at Tampere University. Since 2023, the BST has also been used at LUT University. The test consists of 16 basic mathematics tasks and is implemented using STACK (System for Teaching and Assessment using a Computer Algebra Kernel) (Sangwin, 2013), automatically evaluating students' responses. Kangaslampi et al. (2024) have used the BST to compare results between men and women and between Finnish and international students.

The data used in this article consists of the responses to the BST taken by Finnish students participating in the first first-year mathematics course at the beginning of the academic year in 2023 and 2024. The BST consisted of 16 tasks, each worth one point. The numbers used in the questions were randomised. The questions of the BST with example numbers are presented in Table 1. The points that students received on the BST did not affect their course grade. However, as an incentive, students who took the test were awarded a small number of points on the associated course, irrespective of their results on this test.

Students had 50 minutes to complete the test. They were instructed to take the test independently, without calculators, mathematical software, or other materials. In 2023, LUT University and Tampere University students took the BST in a non-invigilated environment; therefore, they cannot be ensured that they followed the given instructions. In 2024, students in LUT took the electronic test in an invigilated classroom, where an invigilator ensured that the students followed the instructions. In

Tampere, the test was also invigilated; students took the test in a video-invigilated exam class.

Table 1. The questions of the BST.

1. Compute the value of the statement $x^4 + x^3$ when $x = 4$.
2. What is the inverse reciprocal of -5.
3. Simplify the following statement $\lvert 6 - \sqrt{5}\rvert + \lvert 1 - \sqrt{5}\rvert$.
4. Compute the value of function $f(x) = \sqrt{7} + \frac{5}{x} - 4$ when the variable x gets assigned a value $x = 25$.
5. Simplify the statement $\frac{a^2-4}{a-2}$ (Assuming the denominator is not zero.).
6. Let $f(x) = 3x + 1$. Calculate the value of the composite function $(f \circ f)(x)$ at point $x = 2$.
7. Solve the first-degree equation $5x - 8 = 0$.
8. Solve the third-degree equation $4x^3 - 80x = 0$.
9. Solve the fractional equation $\frac{x^2}{x-6} + 3 = 0$.
10. Solve the first-degree inequality $2x + 6 < 0$ Input your solution as "$x < a$" or "$x > a$", where a is the accurate value of the root.
11. Compute the values of the following statements, when $a = 5$ and $b = 5$, $(-a)^b$, $-a^b$, a^{-b}, $-a^{-b}$.
12. Find a solution of the equation $\sin x = \frac{\sqrt{3}}{2}$ located in the first quadrant $0 \leq x \leq \frac{\pi}{2}$.
13. Find the smallest solution to $\cos(5x) = 0$ within the first or second quadrants $0 \leq x \leq \pi$.
14. Differentiate $f(x) = 7x^4 + 5$
15. Calculate the definite integral $\int_{-1}^{3} (4x^4 - 3x^2 + 1)\, \mathrm{d}x$.
16. Solve the exponential equation $3^x = 10$.

1.3 Research questions

This article investigates how non-invigilated and invigilated testing situations affect the results of the basic skills test for first-year engineering students in the Finnish education system. It also analyses which subject areas invigilated testing particularly impact students' results.

The objectives of the study are summarised in the following research questions:

1. How do non-invigilated and invigilated testing situations affect students' scores on the basic skills test?
2. In which subject areas does non-invigilated testing particularly affect students' competence?

2 METHODOLOGY

The study involved first-year engineering students from two Finnish universities, Tampere University and LUT University. In 2023, 121 students from LUT University participated in the study, and in 2024, 96 students participated. The test included students from Finnish-language engineering programs. Consent to participate in the study was obtained from the participants. At LUT University, the data was collected from students of the LUT School of Energy Systems (LES) departments of Energy Technology, Mechanical Engineering, Electrical Engineering and Sustainability Science. In 2023, 674 students from Tampere University participated in the study, and in 2024, 633 students participated. At Tampere University, the data was collected from first year students of all Bachelor of Science programmes in engineering.

The participants completed the electronic BST at the beginning of their first mathematics course. Students' answers were collected, and the differences between the means and standard deviations of the final score and the means of question-specific scores were compared. All statistical analyses were performed using Matlab.

3 RESULTS

In the BST, a correct answer in each of the 16 questions gave the student one point, except question 11, which gave 0.25 points for each of its four separate parts. Due to technical difficulties with question 10 in 2023, it was left out of the analysis, making the maximum score 15. In the non-invigilated test taken in 2023 by students of both universities (n=795), the mean score was 11.05 points with a standard deviation (SD) of 2.70 points. When BST was an invigilated test in 2024 (n=729), the mean score was 8.51 points with an SD of 2.60 points. The difference of means is statistically highly significant, as indicated by a two-sample t-test, $t(1522)=18.69$, $p< .001$. The participating universities' individual means are found in Table 2.

Table 2. Basic Skills Test mean points and standard deviations by university.

	Non-invigilated (2023)			Invigilated (2024)		
University	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
LUT (LES)	121	10.81	2.91	96	7.84	2.82
Tampere	674	11.10	2.66	633	8.61	2.55
Both	795	11.05	2.70	729	8.51	2.60

The histograms of the test results are presented in Figure 1. They show an evident change in the distribution of points. Regarding top scores, 7.92 % of students scored more than 14 points in the non-invigilated test, whereas only 0.69 % reached this level in the invigilated test. On the other hand, the percentage of students scoring 8 points or less was 14.84 % in the non-invigilated test and 46.64 % in the invigilated test.

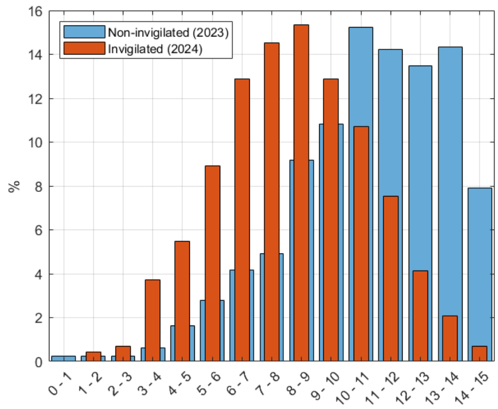

Fig. 1. Basic Skills Test score distributions.

The major difference in the fraction of students scoring less than 8, along with the absence of other explanations – such as differences in student skill distributions across different years – provides a rough estimate of the percentage of students e-cheating: 31.80%.

This figure should be considered a high or maximum estimate as alternative explanations are not explored. The percentage, however, agrees with Holden et al. (2023) figures of 29 - 41% cheating for online exams.

The mean scores for individual test questions (Figure 2) show a point drop for each question when comparing the results of the non-invigilated BST to the invigilated BST. Performing a two-sample Welch's t-test for each question individually indicates that the score difference between the non-invigilated and invigilated tests is statistically significant for all questions except Questions 2 and 14. Table 3 lists the means, their differences and p-values of Welch's t-test for each question.

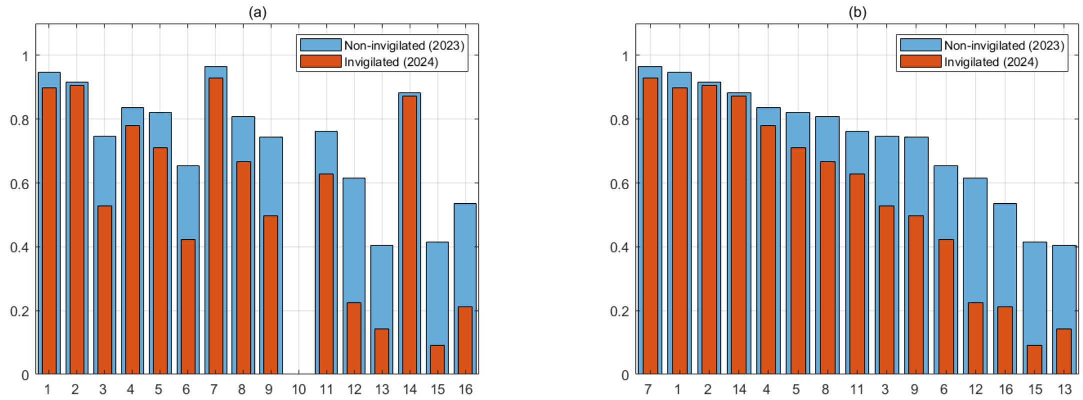

Fig. 2. (a) Mean points per question in the Basic Skills Test. (b) Same data, ordered by mean score in the non-invigilated test.

Table 3. Basic skills test mean points by question, standards deviations, differences of means, and Welch's t-test statistic values and p-values.

Question	2023		2024		Difference of means	Welch's t-test	
	Mean	SD	Mean	SD		t	p
1	0.95	0.23	0.90	0.30	0.05	3.53	< 0.001
2	0.92	0.28	0.91	0.29	0.01	0.70	0.481
3	0.75	0.44	0.53	0.50	0.22	9.04	< 0.001
4	0.84	0.37	0.78	0.41	0.06	2.90	0.004
5	0.82	0.38	0.71	0.45	0.11	5.13	< 0.001
6	0.65	0.48	0.42	0.49	0.23	9.24	< 0.001
7	0.96	0.18	0.93	0.26	0.03	3.02	0.003
8	0.81	0.39	0.67	0.47	0.14	6.36	< 0.001
9	0.74	0.43	0.50	0.50	0.25	10.22	< 0.001
10	-	-	-	-	-	-	-
11	0.76	0.34	0.63	0.37	0.13	7.37	< 0.001
12	0.62	0.49	0.23	0.41	0.39	16.88	< 0.001
13	0.41	0.49	0.14	0.35	0.26	12.08	< 0.001
14	0.88	0.32	0.87	0.33	0.01	0.55	0.580
15	0.42	0.49	0.09	0.29	0.32	15.76	< 0.001
16	0.53	0.50	0.21	0.41	0.32	13.81	< 0.001

Questions 1, 2, 7, and 14 had the highest mean scores in both years. The difference in points between the two years was the smallest for these questions, at most 0.05

points in each case. Question 1 involved finding the value of a function given a specific variable value, Question 2 tested conceptual understanding of the terms inverse and reciprocal, Question 7 required solving a first-degree equation, and Question 14 was a differentiation problem.

Questions 12, 13, 15, and 16 were the lowest-scoring questions in both years. They also had the greatest points difference between the two years, dropping more than 0.25 points each. Questions 12, 13, and 16 involved solving equations – specifically, the first two were trigonometric equations, while the third was an exponential equation. In Question 15, students had to calculate a definite integral.

4 DISCUSSION AND CONCLUSIONS

In our evaluation, all questions in the BST are at Bloom's taxonomy levels 1-2 and are not aimed at testing higher-order cognitive skills. In both the non-invigilated test and in the invigilated test, the results of BST questions 1, 2, 7, and 14 did not differ very much. These questions involved functions, understanding the concepts of an inverse and a reciprocal, linear equations, and differentiation. This aligns with the findings of Metsämuuronen and Tuohilampi (2017).

Questions 12, 13, 15, and 16 proved to be the most difficult ones in both test settings. Additionally, the differences of the results in the invigilated and non-invigilated tests were the largest for these questions. The biggest drops in BST results mirror some of the findings of Lima et al. (2019) and Myllykoski et al. (2018), in which common difficult mathematical areas such as trigonometry and logarithms (or exponential functions) were identified. The question involving a definite integral also proved to be difficult. Evaluation a definite integral involves first finding the indefinite integral and subtracting the integral's value at the lower limit from its value at the upper limit. Subtraction by hand may be difficult for a student that is used to using a symbolic calculator. According to Laatikainen (2022), Lukin (2023), and Perälä (2023), the students' competence in integrals has been lower than in other fields of mathematics in the mathematics test of the Finnish matriculation exam. This is also in line with our observations.

The difficult questions of the BST are easily solved with symbolic calculators, AI tools, or other math solver applications. While alternative explanations for the difference could not be accounted for, e-cheating is a possible explanation for differences of the magnitude that was observed in the results. Based on our observations, we suggest that students are more likely to cheat, when cheating is easy and the task at hand requires a little more complexity or a longer solving technique than others. This corresponds to what is suggested by Sloan-Lynch et al. (2022). Bengtsson (2019) recommends against using take-home exams for testing the lowest taxonomic level skills due to the ample opportunity, readily available tools, and the ease of e-cheating in the tests. Based on our results, we echo this recommendation.

Regarding the fraud triangle theory (e.g., Holden, 2021), we conclude that the absence of pressure or incentive to cheat (i.e., no course points awarded based on test performance) did not eliminate the cheating phenomenon in case of the BST. This observation may call into question the validity of the theory in our context – namely, that pressure, incentive, or need are pre-requisite factors for cheating. On the other hand, it is also possible that students experienced a different type of pressure, such as a social need to avoid revealing their lack of basic skills to the teacher, for fear of losing face. Alternatively, students may have been uncertain about whether poor test performance could have undisclosed repercussions.

Perhaps the most significant piece of advice we can offer based on these results is that if you genuinely want to assess the students' level of comprehension at the beginning of their university studies using cognitively less challenging tasks, the test should be invigilated regardless of whether it affects the course grade.

In future research, more effort should be devoted to producing more robust data on how many and what kind of students resort to e-cheating, under what circumstances they do it, and what exactly drives them. This data would help improve our evaluation methods and more conclusively evaluate the validity of non-invigilated online exams.

5 ACKNOWLEDGEMENTS

An artificial intelligence application, SCOPUS AI, was used to find the reference literature. The AI application was used also for English language refinement of the final version of the manuscript.

REFERENCES

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, 27(3), 197-205. https://doi.org/10.1177/0734282908330580

Bengtsson, L. (2019). Take-Home Exams in Higher Education: A Systematic Review. *Education Sciences*, *9*(4), 267. https://doi.org/10.3390/educsci9040267

Chan J.C.K., & Ahn D. (2023). Unproctored online exams provide meaningful assessment of student learning. In *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2302020120

Holden, O. L., Norris, M. E., & Kuhlmeier, V. A. (2021). Academic integrity in online assessment: A research review. *Frontiers in Education*, 6, 639-814. https://doi.org/10.3389/feduc.2021.639814

Joutsenlahti, J., Ali-Löytty, S., & Pohjolainen, S. (2016). Developing Learning and Teaching in Engineering Mathematics with and without Technology. In *SEFI 2016 Annual Conference Proceedings European Society for Engineering Education SEFI*.

Kangaslampi, R., Hirvonen, J., Kaarakka, T., Immonen, P., Äijälä, M., Bhayo, B., Naukkarinen, J., & Kuosa, M. (2024). First-Year Engineering Students' Mathematical Skills And Perceptions Of Studying Mathematics. In *Proceedings of the 52nd Annual Conference of the European Society for Engineering Education* (pp. 578-587). https://doi.org/10.5281/zenodo.14254744

Laatikainen, V. (2022). *Integraalilaskenta ja sen osaaminen pitkän matematiikan ylioppilaskirjoituksissa*. Master's thesis. University of Helsinki. http://hdl.handle.net/10138/351956

Lima, P. D. S. M., Silva, L. D. A., Felix, I. M. & Brandao, L. D. O. (2019) Difficulties in Basic Concepts of Mathematics in Higher Education: A Systematic Review. In *49th*

IEEE Frontiers in Education Conference (pp. 1-7). https://doi.org/10.1109/FIE43999.2019.9028658

Lukin, V. (2023) *Lukion opiskelijoiden ja opettajien näkemyksiä haasteista ja opiskelumotivaatiosta integraalilaskennan opintojaksolla*. Master's thesis, University of Eastern Finland. http://urn.fi/urn:nbn:fi:uef-20230932

Metsämuuronen, J., Tuohilampi, L. (2017). *Matemaattisen osaamisen piirteitä lukiokoulutuksen lopussa 2015*. Kansallinen arviointikeskus.

Myllykoski, T. J., Mattila, P., Ali-Löytty, S., Kaarakka, T., & Viro, E. (2018). Yliopistomatematiikan Sähköisten Tehtävien ja Matemaattisen Ajattelun Kehittäminen. *FMSERA Journal*, *2*(1), 46–55. https://journal.fi/fmsera/article/view/69887

Noorbehbahani, F., Mohammadi, A. & Aminazadehm M. (2022). A systematic review of research on cheating in online exams from 2010 to 2021. *Education and Information Technologies*, *27*(6), 8413–8460. https://doi.org/10.1007/s10639-022-10927-7

Perälä, H., (2023). *Integraalitehtävien osaaminen ennen ja jälkeen sähköisen ylioppilaskoeuudistuksen (2016-2020)*. Master's thesis. University of Helsinki

Sloan-Lynch, J., Gay, N., & Watkins, R.D. (2022). Too Fast for Their Own Good: Analyzing a Decade of Student Exercise Responses to Explore the Impact of Math Solving Photo Apps. In *12th International Learning Analytics and Knowledge Conference.* https://dl.acm.org/doi/10.1145/3506860.3506868

Sangwin, C. (2013). *Computer Aided Assessment of Mathematics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199660353.001.0001