Inside the "Sandbox"

The Effects of Unlimited Practice for Summative Online-Tests

https://doi.org/10.3991/ijet.v17i23.35939

Miriam Clincy(✉), Karin Melzer, Gunther Schaaf, Achim Eichhorn, Nathalie Verné
Hochschule Esslingen, Esslingen am Neckar, Germany
miriam.clincy@hs-esslingen.de

Abstract—Online tests with automated response systems are often used as pre-assessments or to give students formative feedback on their progress during a course. As an incentive, bonus points are commonly awarded. The pass threshold plays a crucial role: A low pass threshold encourages participation whereas a high pass threshold signals clear expectations for the course, yet is sometimes mitigated by allowing for repeated tries. In two first-year maths courses in different degree programs online tests with STACK questions were introduced. These tests are offered independent from the course format. They have a high pass threshold, allow only a limited number of trials and count towards the final mark, i.e. are considered a summative assessment for the course. However, to allow for practice and formative assessment, "sandbox" environments were implemented allowing for unlimited practice. This article will give an overview over the choice of parameters for different implementations and discuss the findings on the use of the "sandbox".

Keywords—STACK, online test, mathematics, assessment

1 Introduction

Universities of Applied Sciences attract students from a variety of educational backgrounds: While some come straight out of secondary school others have completed vocational training, decide to study after an entry-level job or go for a mid-career job change. For STEM related subjects, these diverse backgrounds are particularly pronounced in mathematics proficiency which is a key success factor as shown in [1], [2] with a more detailed analyses of the situation in Germany in [3].

Universities address this heterogeneity with support measures on different levels: Tutorials, self-study courses and in particular, online quizzes (or tests) can address a range of issues besides content knowledge such as organizational or intra-personal skills.

One possibility of categorizing the test purposes can be captured in terms of the assessment: On the one hand, online tests can be used for summative feedback such as in digital exams. On the other hand, automating support with online tests in mathematics allows for highly individualized feedback as well as time- and space-independent study opportunities and is therefore often linked to formative assessment. Both applications were boosted by the pandemic lock downs and distance teaching.

Depending on the aims of individual lecturers, the scope of the course programs or the study culture at institutions online tests can serve a multitude of purposes which is reflected in the choice of settings.

In this article, we first try to give an overview – albeit surely incomplete – of common settings. In our own institution, we introduced online tests in different courses of study in first year mathematics. We will compare two of these implementations which both try to combine formative and summative feedback by allowing for unlimited practice in a sandbox setting. We conclude by outlining some of our findings.

2 Choices of settings in online tests

2.1 Dimensions of settings

Learning management systems (LMS) such as Moodle or ILIAS offer many different choices for online test settings, the actual selection of questions being only one dimension. When speaking to colleagues, no two settings seem to be alike. The choices depend on the expected or desired student behavior, but also on external factors such as the embedding of the tests in the exam regulations and the role of the tests within the learning arrangement. In short, assessment depends on the aim and the timing [4].

Here, an overview over common characteristics of online test settings is given to structure the following discussion:

1. Assessment: This can be classified in a continuum ranging from formative to summative following the definition by Scriven [5]: In a formative assessment, feedback assessing the individual and subjective progress is essential to allow students to improve, the focus is on learning; in a summative assessment, the student output is measured against an objective scale evaluating the outcome at the end of a learning process.
2. Participation mode: The tests could be compulsory or optional. An optional test might be incentivized, e.g. by earning bonus points for an exam. This is usually determined by the exam regulations. A detailed overview over incentives in digital learning can be found e.g. in [6].
3. Feedback:
   - Detail: Implementations in LMS such as Moodle or ILIAS allow for different levels of feedback: It can range from no feedback at all to indicating only correctness up to very detailed levels of automated response taking into account the student's answer and categorizing according to the most common mistakes. A precise control over the level of feedback depending on the student answer is in particular given by systems involving a Computer Algebra System (CAS) to evaluate student answers such as STACK (System or Teaching and Assessment using a Computer algebra Kernel) [7].
   - Timing: Feedback could be given immediately after an attempt or with a time delay. The latter has been found to have a profound impact on the long-term learning effect, e.g. [8].

4. Allowing repeats: The number of tries could be limited or unlimited. In addition, there might even be a penalty imposed on repeats, especially if solution hints are given, for a suggestion for an implementation in STACK compare e.g. [9].
5. Pass thresholds: The pass thresholds might be high to communicate the high level of expectations [10], [11] or low to incentivize participation [12].
6. Bloom's taxonomy: Another dimension is added by considering Bloom's taxonomy of learning objectives [13], restricted here to the cognitive scale with its six levels: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Online tests are often implemented for the lower levels of this scale. With higher levels student answers become more individualized and corresponding tasks more text based calling for more advanced techniques of automated assessment. An evaluation of text based answers in STACK has been suggested by [14], also in this issue.

2.2 Metric suggestion

Trying to find a metric, the above features have been grouped in Table 1 assigning values between 0 and 3 for each characteristic. The highest numbers roughly correspond to an exam situation: strictly summative and compulsory, feedback only after some time, high pass thresholds, etc. On the other end of the spectrum, 0 would correspond to a purely formative and optional test, such as free online tests to self-evaluate math proficiency before deciding on a course of study. Optional tests as part of a course or incentivized tests would be assigned a higher value for the participation mode.

Table 1. Suggested metric for online test settings

Value	Assessment	Participation	Repeats Allowed	Feedback Detail	Feedback Timing	Pass Threshold	Bloom's Taxonomy
0	Strictly formative	Optional, freely available	unrestricted	Full feedback including path of solution/ feedback specific to student input	Immediate feedback per answer	0%	Knowledge (0.5)
1		Optional, as part of course	Several repeats allowed	Solution hints, general feedback on problem	Feedback after test in completed		Comprehension (1.0) Application (1.5)
2		Incentivized with bonus points	One repeat allowed	Correct answer	Feedback several hours later		Analysis (2.0) Synthesis (2.5)
3	Strictly summative (exam)	compulsory	once	Right/ wrong	Several weeks delay	100%	Evaluation (3.0)

In a two-dimensional overview focusing on the assessment type and the participation mode some models could be displayed as in Figure 1 with additional features grouped around the main columns.

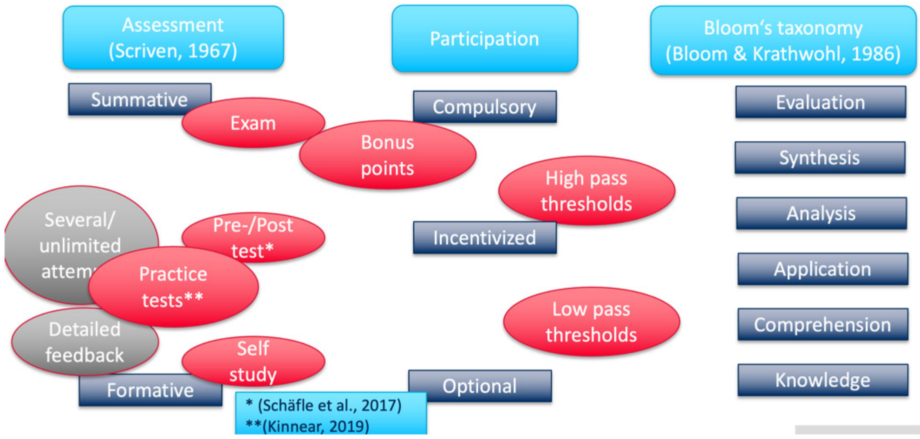

Fig. 1. Examples of settings in online tests

2.3 Interdependencies

There are obviously interdependencies between the settings: for example,

- a summative test such as an exam would usually contain compulsory elements or be at least strongly incentivized;
- allowing for repeated tests and detailed feedback would correspond to a setting in which a learner receives formative feedback – summative assessment with repeated attempts and detailed feedback would lead to a situation in which the student runs the test initially to only receive the feedback for the second try;
- for optional tests a low threshold might have a positive effect on the participation rate which can then be combined with a formative assessment ([15] have coined the term *threshold formative assessment* for such a setting).

Different modes of tests have been combined in "Just in time teaching (JiTT)" arrangements [16] with pre-tests after working through the course material and post tests after the lecture or further study as in [11], [12]. These arrangements are therefore closely linked to the course format.

3 Online tests and the sandbox: implementation

3.1 Overview

In this article, we retrospectively compare the implementation of online tests in two different courses as also reported in [17].

The online tests were implemented in two first-semester mathematics courses in information technology (IT) and automotive engineering (AE). In IT, STACK tests were introduced in the summer semester 2019 to automate manually marked problem sheets. In automotive engineering, the first tests replaced a long-standing 60-minute midterm exam for the first time during the lock down in the summer semester 2021.

In both courses of study, the design of the tests allowed them to be employed irrespective of the lecturer and the course format (as long as a minimal agreement on the sequence and time frame of the content is observed). Because of this ubiquitous design, they have been in use ever since undergoing only minor changes over time (such as the pass thresholds).

To allow for a common framework for comparison, we therefore focus on the initial implementation under a comparably high pass threshold which is the winter semester 2019/20 for the tests in IT (second run) and the summer semester 2021 in automotive engineering (first run).

In the representation chosen in Figure 1, these tests can be located graphically as in Figure 2. They have a summative component by being incentivized with bonus points for the final summative exam; a formative element is introduced by practice quizzes closely resembling the tests. These can be attempted an unlimited number of times to prepare for the test proper. With reference to development systems in software engineering operating independently from the live system, these practice quizzes have been coined "sandbox".

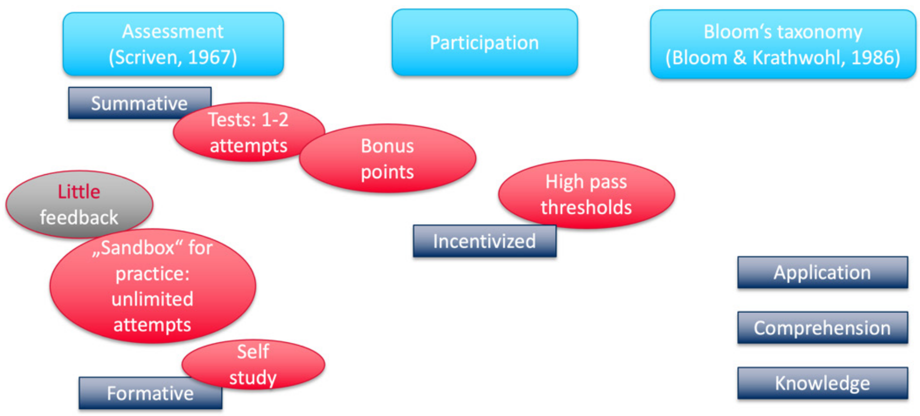

Fig. 2. Choices of settings

3.2 Comparison

In IT, the five online problem sheets to be handed in covered the course material in linear algebra over one semester. These online tests were preceded by a similar, slightly longer practice test, i.e. the sandbox, on the same topic (questions being randomized). The real test could be accessed only when the sandbox had been successfully completed with a minimum score of 75%. The tests had to be completed within a pre-set two-week time frame. No time limit was set on either the practice quiz or the test itself.

A bonus of 10% was granted in the linear algebra exam if 75% of the *accumulated* points in all 5 tests were achieved.

In automotive engineering, the former midterm exam was implemented as six 10-minute online tests over the course of one semester covering both topics from linear algebra and analysis. The questions themselves were randomized. For each test the sandbox allowed practice for an unlimited number of times under the same conditions (10 minutes each, similar test questions). Practice was strongly encouraged e.g. by explaining the probability that sufficient practice would have covered the test variation. For students not reaching the pass threshold in the test itself, a single repeat was allowed. A bonus of 10% on the mathematics exam was granted if 5 out of 6 tests were successfully completed with a 75% pass threshold *each*.

As described above, the initial aim of the tests was different (problem sheets versus midterm exam) and therefore the settings differed as shown in the next sections and in an overview Table 2.

Table 2. Settings for online tests in first year mathematics for IT and automotive engineering

	Computer Science (IT)	**Automotive Engineering (AE)**
Since	Winter term 2019/20	Summer term 2021
Replacing	Manually marked problem sheets	Midterm exam, 60 minute duration
Course	Linear algebra	Analysis and linear algebra
Test content	1. Sets of linear equations 2. Vector algebra 3. Lines and planes 4. Matrix algebra 5. Complex numbers	1. Sets of linear equations & Functions I 2. Functions 2 3. Vector algebra 4. Limits and Differentiation 5. Matrices and Integration 1 6. Integration 2
Integration in curriculum	University-wide exam regulation: ≤ 25% of score may be obtained during the semester.	*60 min* «Midterm-exam» in exam regulation (course): split into 6 tests à 10 min each
Pass threshold for bonus points	75% of total score (cumulative)	5 out of 6 tests have to be passed individually (75% each)
Test settings	One attempt, no time limit within two-week time frame	2 attempts in 2 weeks, 10-minute time limit
Sandbox settings	Questions partly identical to test, no time limit Test made available at 75% pass rate in sandbox	Identical tests in sandbox (randomization), 10-minute time limit

3.3 Common ground

The above differences between the tests might seem like 'comparing apples to oranges' therefore a brief overview of the common denominator seems in order:

Both implementations allow for a 10% bonus in the exam having similar pass thresholds of 75% (however, different modes of reaching these). The tests were spaced out over the semester in two-week intervals with practice tests available for an unlimited number of trials in advance.

The problem types were in the first levels of the Bloom taxonomy, i.e. the aim of the test was mainly focusing on providing practice in applying calculation techniques in first year engineering mathematics: No detailed problem solving or proofs were required (nor part of the course). The settings were such that immediate short feedback was given with the correct answer, but no detailed solution. This was due to the nature of these basic problems which were either standard textbook questions or had been presented in the lectures. Hence, the focus was on training basic calculus techniques leaving the solution of more complex problems to the lectures and exam preparation.

Taking into account the parameters and metrics outlined in Table 1, the multidimensional settings can be compared in web diagrams as shown in Figure 3: Even though the purpose of the tests was different, the settings are fairly similar when represented in terms of parameter choices. The implementation in information technology (IT) is characterized by a slightly higher level in Bloom's taxonomy and some more detail in the feedback.

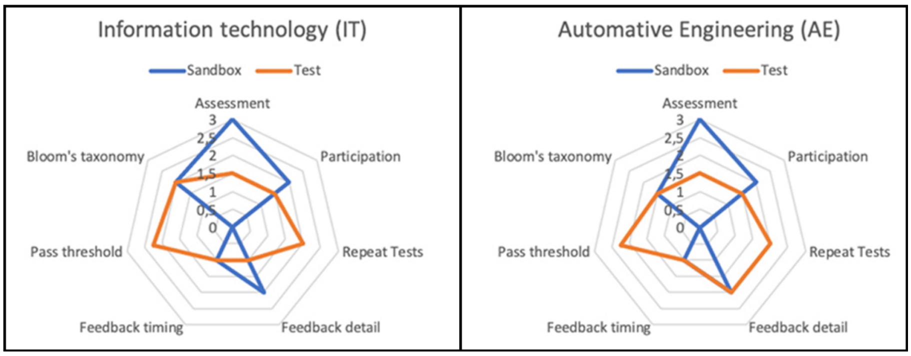

Fig. 3. Multidimensional comparison of settings according to the metric suggested in Table 1

3.4 Question types and technical realization

The questions for the IT implementation were almost exclusively STACK questions with occasional inherent Moodle formats. For the implementation in AE, the questions were a mixture of STACK questions and randomized questions from an in-house development from a MATLAB-Moodle-Interface [18].

The questions aimed for the lower levels of Bloom's taxonomy of learning objectives [13]: remembering, understanding and straightforward applications. The implementation made use of the CAS to include open questions to support student understanding. An example is shown in Figure 4 where there is an infinite number of possible solutions.

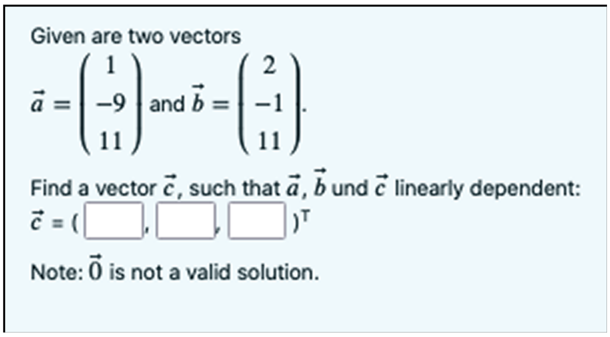

Fig. 4. Example of open question

4 Results

The following analysis does not contain a comparison of exam results, because of the different syllabi and different times (only one of the implementations having been conducted during distance teaching).

4.1 Overview

In the winter of 2019/20, 172 IT students in three different parallel cohorts completed the first test – meaning that the practice in the sandbox had yielded at least 75% to unlock it. Averaged over all five tests, the number of average attempts in the sandbox was 2.2 ± 0.2, i.e. after roughly the second attempt the 75% threshold was reached. The maximum number of tries in the sandbox ranged from 7 to 10. The overall pass rate (accumulating at least 75% of the total number of points) was 79.4% meaning that 4 out of 5 students who attempted the first test were able to secure the bonus points for the exam.

A comparison of these figures with the first implementation of the time restricted tests in automotive engineering shown in Table 3: 83 students in two parallel cohorts attempted the first test. No threshold was required and the concept of unlimited practice of similar problems in the sandbox was heavily advertised. The average number of attempts in the sandbox was about 10, i.e. five times higher than in the IT implementation. Individual students completed the sandbox tests up to 30 times or more. Yet the pass rate (counted as 5 out of 6 tests successfully completed with a score of at least 75% each) is considerably lower with slightly less than two-third of the students who attempted the first test.

Table 3. Participation and results

	No. of Participants (first test)	No. Average Attempts	Max. Number of Attempts per Student	Pass Rate
IT (unlimited time)	172	2.2 ± 0.2	7–10	79.4%
AE (10 minutes)	83	9.9 ± 3.0	> 30	63.2%

There are several possible explanations for the lower pass rates, one of them being that the AE tests were first conducted during the pandemic with students having a harder time motivating themselves to complete the course. One consideration beforehand had been whether the sandbox might have an effect of 'training to the test' thereby trivializing it. With the observed pass rate, this concern was retrospectively dismissed as unfounded. Another explanation could be that the time limit was considered an additional hurdle; however, this was partly compensated by more practice attempts.

4.2 Retention rates

These observations lead to the question whether there are observable differences in the retention rates, i.e. the percentage of students who participate in the tests throughout the semester. The percentage of participants scaled to the first test is shown in Figure 5 for both implementations for the sandboxes as well as the tests. As expected, the last tests show a drop in attendance, as for many students the pass had already been achieved before the last attempt. In terms of percentage, the implementations throughout the course appear similar – none of the models displays a clear preference in retention rates.

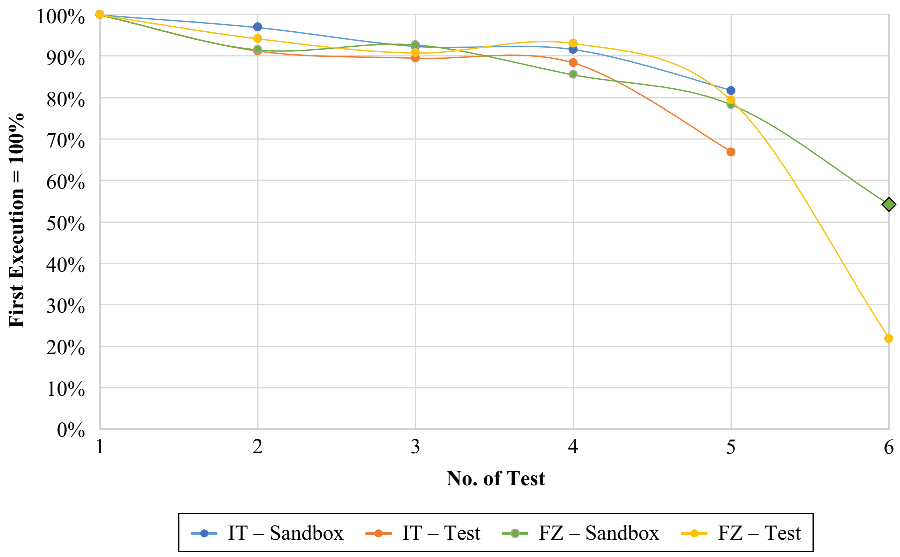

Fig. 5. Ratio of test participants scaled to initial participation

4.3 "Learning curve"

The AE implementation had a strict time limit on the tests as well as on the attempts in the sandbox. The 10-minute time frame seemed to induce a high number of tries as shown in Table 3. Hence the question arose whether a gradual improvement in the results would be observable over time. In Figure 6, the average number of points is shown as a function of the number of the attempt in the sandbox. With the low number of participants, the sandbox for test 6 shows inconclusive results (black triangles). However, for the first five tests, a gradual increase of points up to the pass rate can be observed. It seems that the majority of students take between 5 and roughly 7 attempts before the pass threshold can be reached confidently leading to what could be termed the *learning curve*.

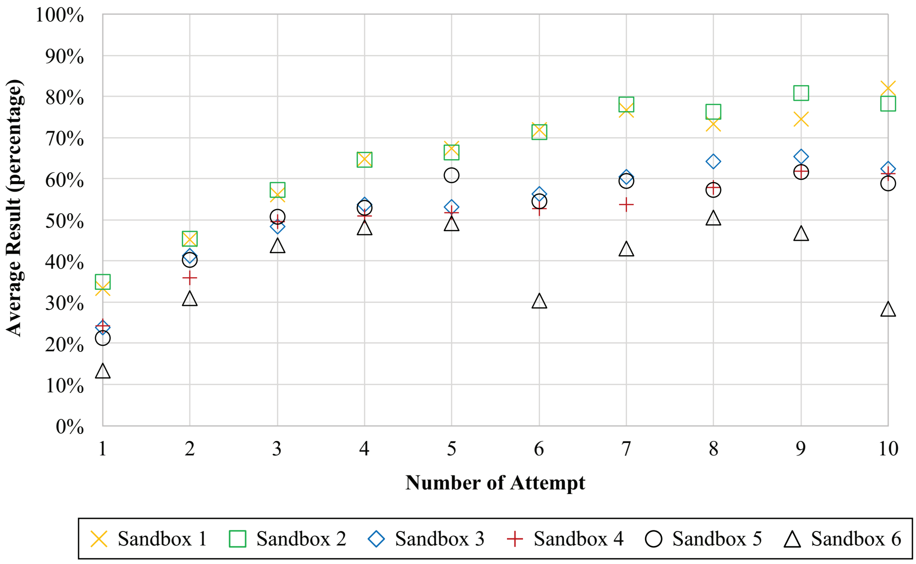

Fig. 6. Average score in sandbox as a function of attempt (for automotive engineering)

4.4 Collaboration

For the implementation in IT, no special precautions had been taken to prevent collaboration. As with manually marked problem sheets, working in groups to discuss solutions was considered acceptable. In particular, because the questions were randomized, any meaningful discussion between students would therefore necessarily have to focus on the path of solution.

The six online tests in automotive engineering replaced a summative midterm exam, hence a prevention of collaboration would have been desirable. However, the implementation was overshadowed by the discussions on proctoring, data security as well as attempts of deception in online exams taken at home. It was therefore decided to circumvent these discussions by allowing the students to take the tests entirely unsupervised from a computer at home.

Nonetheless, there was an interest in analyzing the data from the first cohort in AE by searching for patterns. Assuming that collaboration would manifest in students finishing the tests exceptionally fast with a pass.

Five out of six tests were mandatory, hence for Test 5 (labelled E) the incentive to seek cooperation was assumed to be maximal. In Figure 7, the results are shown (for technical reasons, the time taken for the tests was rounded to full minutes). It can be seen that some students passed the test in as little as 4 minutes. To investigate further whether these time scales are realistic, we analyzed the same data for the sandbox of test 5 where there was not pressure to perform and therefore no need to collaborate. The data shown is for all attempts. As can be seen, even in the sandbox, some students pass the test in the same time as the test, i.e. in as little as 4 minutes.

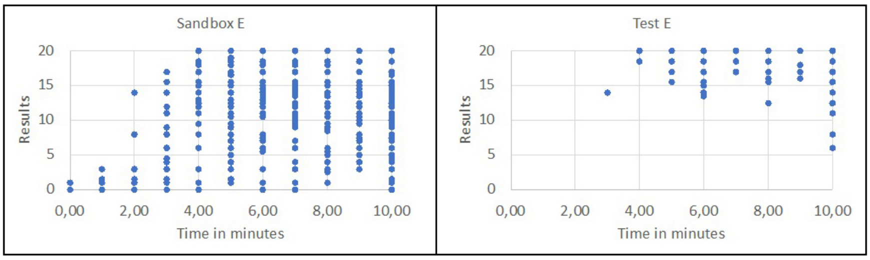

Fig. 7. Comparison of scores as a function of time spent on test for sandbox 5 (E) and test 5 (E) (automotive engineering)

We therefore conclude that despite not having full insight into student behaviour and not being able to rule out any attempts at deception categorically, there is no indication for a wide scale collaboration.

One reason might lie in the sandbox approach as such: There is enough time and opportunity to practice without pressure and the scope of the test is limited. Whether this is the case and whether the time limit is adequate for all tests is currently being investigated.

Another explanation might lie in the social dynamics during the lock-downs: Many of the first-year students had little or no contact to other students outside virtual lectures and felt isolated. Finding a study group and working on the problems together might have been an additional hurdle many could not take. As the tests are continued with students physically present at lectures again, collaboration or the lack of it in passing the online tests could be an issue to investigate further.

5 Outlook

In this paper, an overview over different settings and features of online tests was attempted. For two different courses of study, the implementation of a series of tests during the semester incentivized by bonus points for the exam was given with the details of the settings. The tests were preceded by practice tests in a "sandbox" environment adding a formative component to the assessment. Participation rates and the development of pass rates with attempts as well as indication of collaboration were analyzed. Especially an implementation with a time restriction of ten minutes corresponded to high number of attempts with an observable increase in scores over time. No strong indication of collaboration for these tests were found.

The tests were implemented irrespective of the course format and have therefore been adopted as part of the curriculum in both courses of study. They cover the lower levels of Bloom's taxonomy of learning objectives. Therefore, these tests could contribute to a transparent separation of learning outcomes for a course by outsourcing basic calculation skills to STACK tests with automated response systems during the semester and focusing exam questions more clearly on higher learning objectives.

6 Acknowledgment

Special thanks to our colleagues from TH Rosenheim for critically questioning our approach to student collaboration and prompting further analysis.

7 References

[1] A. Sithole, E. T. Chiyaka, P. McCarthy, D. M. Mupinga, B. K. Bucklein, and J. Kibirige, "Student Attraction, Persistence and Retention in STEM Programs: Successes and Continuing Challenges", *HES*, vol. 7, no. 1, p. 46, Jan. 2017, https://doi.org/10.5539/hes.v7n1p46

[2] P. Juric, M. B. Bakaric, and M. Matetic, "Cognitive Predispositions of Students for STEM Success and Differences in Solving Problems in the Computer Game for Learning Mathematics", *International Journal of Engineering Pedagogy (iJEP)*, vol. 11, no. 4, Art. no. 4, Jul. 2021, https://doi.org/10.3991/ijep.v11i4.20587

[3] R. Biehler, R. Hochmuth, and T. Wassong, "Transition von Schule zu Hochschule in der Mathematik: Probleme und Lösungsansätze", in *Beiträge zum Mathematikunterricht 2011*, 2011, p. 4. [Online]. Available: http://hdl.handle.net/2003/32085

[4] W. Harlen and M. James, "Creating a Positive Impact of Assessment on Learning", presented at the Annual Meeting of the American Educational Reseach Association, New York, Apr. 1996. [Online]. Available: https://files.eric.ed.gov/fulltext/ED397137.pdf

[5] M. Scriven, "The methodology of evaluation", in *Perspectives of curriculum evaluation*, R. W. Tyler, Ed. Chicago: Rand McNally education series, 1967.

[6] A. A. Alsulaimani, "Developing a Model for e-Learning Activities Based on Digital Incentives Within the Context of COVID-19 and its Effectiveness in Developing Meta-Cognitive Thinking Among Higher Education Students", *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 17, Art. no. 17, Sep. 2022, https://doi.org/10.3991/ijet.v17i17.29403

[7] C. Sangwin, "Computer Aided Assessment of Mathematics Using STACK", in *Selected Regular Lectures from the 12th International Congress on Mathematical Education*, S. J. Cho, Ed. Cham: Springer International Publishing, 2015, pp. 695–713. https://doi.org/10.1007/978-3-319-17187-6_39

[8] N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler, "Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention", *Psychol Sci*, vol. 19, no. 11, pp. 1095–1102, Nov. 2008, https://doi.org/10.1111/j.1467-9280.2008.02209.x

[9] M. Kallweit and J. Lache, "Playing with Interactions in STACK Exercises", presented at the Playing with interactions in STACK exercises, https://www.stack21.edu.ee/schedule, Tallinn/online, 2021.

[10] A. W. Chickering and Z. F. Gamson, "Seven Principles for Good Practice in Undergraduate Education", *AAHE Bulletin*, no. March, pp. 3–7, 1987, [Online]. Available: https://files.eric.ed.gov/fulltext/ED282491.pdf

[11] G. Kinnear, "Delivering an Online Course Using STACK", Feb. 2019, https://doi.org/10.5281/ZENODO.2565969

[12] C. Schäfle, S. Stanzel, E. Junker, and M. Zimmermann, "Teil 1: Aktivierende und konzeptorientierte Lehrmethoden", *Didaktik-Nachrichten*, pp. 3–12, Jun. 2017. [Online]. Available: https://diz-bayern.de/DiNa/06_2017
[13] B. S. Bloom and D. R. Krathwohl, *Taxonomy of educational objectives. 1: Cognitive domain*, 29. print. London: Longman, 1986.
[14] A. Eichhorn and A. Helfrich-Schkarbanenko, "Question Answering in STACK Applying String Similarity", presented at the International Meeting of the STACK Community 2022, Leoben, online, Apr. 24, 2022.
[15] S. Jordan, "Thresholded assessment: Does it work?", *eSTEeM Final report*, p. 56, 2014, [Online]. Available: https://www.open.ac.uk/about/teaching-and-learning/esteem/sites/www.open.ac.uk.about.teaching-and-learning.esteem/files/files/ecms/web-content/2014-09-Sally-Jordan-Final-Report-Thresholded-assessment.pdf
[16] G. M. Novak, Ed., *Just-in-time teaching: blending active learning with web technology*. Upper Saddle River, NJ: Prentice Hall, 1999.
[17] M. Clincy, K. Melzer, G. Schaaf, A. Eichhorn, and N. Verné, "Online midterm tests between summative and formative assessment – Online-Midterms zwischen formativer und summativer Bewertung", *ZFHE*, pp. 141–156, 2022, https://doi.org/10.3217/ZFHE-17-01/09
[18] G. Schaaf, A. Eichhorn, and A. Helfrich-Schkarbanenko, "Digitaler Rückenwind für Mathematikveranstaltungen – Aufgabengeneratoren", presented at the Poster für die 'MATLAB EXPO 2019', Jul. 2019. [Online]. Available: https://www2.hs-esslingen.de/~aeich/digitalerrueckenwind/aufgabengeneratoren/

8 Authors

Prof. Dr. Miriam Clincy, Professor for STEM Teaching in Secondary Education and Mathematics at Hochschule Esslingen, Germany, is currently Representative for Didactics in Higher Education at her institution. She has implemented online tests in mathematics and physics and is interested in e-learning and blended teaching formats in secondary and higher education (email: miriam.clincy@hs-esslingen.de).

Prof. Dr. Karin Melzer, Professor for Mathematics at Hochschule Esslingen, has implemented and used online-tests with automatic assessment at the department Computer Science and Engineering (email: karin.melzer@hs-esslingen.de)

Prof. Dr. Gunther Schaaf, Professor for Mathematics at Hochschule Esslingen has been responsible for the design of e-learning tests for mathematics courses at various faculties at his institution (email: gunther.schaaf@hs-esslingen.de).

Achim Eichhorn is employed as a graduate computer scientist at the Esslingen University of Applied Sciences, Germany. He is a very experienced software developer, specialized in web technologies and interactive, multi-media e-learning content (email: achim.eichhorn@hs-esslingen.de).

Nathalie Verné is a research assistant at the Department of Teaching and Further Education, Esslingen University of Applied Sciences, Germany. Her tasks include working as an instructional designer and developing e-learning modules according to subject-specific requirements and media-didactic guidelines (email: Nathalie.Verne@hochschule-esslingen.de).

Article submitted 2022-09-10. Resubmitted 2022-10-10. Final acceptance 2022-10-11. Final version published as submitted by the authors.