Suhonen, S. (2023) Automatically Scored, Multiple-Attempt, Recurring Weekly Exams In A Physics Course: Can They Improve Student Wellbeing And Learning Outcomes? Proceeding of the 51st Annual Conference of the European Society for Engineering Education (SEFI). TU Dublin.

URL: https://doi.org/10.21427/0N63-HP91

Automatically Scored, Multiple-Attempt, Recurring Weekly Exams In A Physics Course: Can They Improve Student Wellbeing And Learning Outcomes?

Sami SUHONEN
Tampere University of Applied Sciences, Finland, sami.suhonen@tuni.fi

AUTOMATICALLY SCORED, MULTIPLE -ATTEMPT, RECURRING WEEKLY EXAMS IN A PHYSICS COURSE: CAN THEY IMPROVE STUDENT WELLBEING AND LEARNING OUTCOMES?

S. J. Suhonen [1]
Tampere University of Applied Sciences
Tampere, Finland
0000-0002-3279-3813

Conference Key Areas: *Innovative Teaching and Learning Methods, Fundamentals of Engineering: Mathematics and the Sciences*
Keywords: *Engineering physics, Automatic scoring, Online exams, Student wellbeing*

ABSTRACT

This paper describes the arrangements and assessment methods employed in the engineering physics online courses offered by Tampere University of Applied Sciences. The grading process involves the use of automatically assessed weekly online exams in Moodle, which comprise mostly numerical questions, as well as conceptual questions, force diagrams, and multiple-choice questions. Students are allowed to attempt each exam three times, and their best score is recorded. The questions and initial values were randomized for each try to reduce possibilities for trial-and-error method and copying from peers. By completing the week exams with enough points, the students were able to pass the course with low grades. The main idea was to make the course completion more flexible and time and place independent and reduce exam stress. It also reduced teachers' workload in relation to assessment and retakes. Most students took more than one attempt in the exams, and the majority of students who initially scored low points showed improvement in subsequent attempts. According to student feedback vast majority of students agreed that this exam arrangement worked well and that retakes reduced stress, was flexible and improve their learning experience and outcomes. Almost no one would like to change back to one-attempt exam checked by the instructor.

---

[1] Corresponding Author
S. J. Suhonen
sami.suhonen@tuni.fi

1 INTRODUCTION

The assessment methods are chosen to align with the curriculum and the intended learning outcomes of a course. In engineering physics at the bachelor's level, the topics, and problems to be solved require both conceptual understanding and algebra/calculus-based problem-solving. Some typical elements used in the assessment are formative assessments during the course with the help of quizzes, forms, or polling surveys, laboratory, and project work with written or oral reports, homework assignments, and summative assessments containing mid-term exams and/or final exams. However, this study concentrates only on automatically scored, recurring, online weekly exams.

Taking an exam typically has two opposing effects on students' behaviour: the exam situation is often perceived as stressful, but on the other hand, students' effort and intensity of working are at a high level. For example, the review by Richardson et al. (2012) suggests that stress and anxiety are important factors to consider in understanding student academic performance, including performance on exams. The levels of stress and anxiety are negatively associated with academic performance, including exam scores. Also, the study by Pascoe et al. (2020) highlights the negative impact of stress on academic outcomes, including grades, attendance, and scores. Students who experience high levels of stress are more likely to drop out of school or fail courses. Stress levels typically increase during mid-term and final exams (Zunhammer 2013).

The question arises: how to reduce the stress level but keep the working intensity high? How to better harness the exams to work as a learning tool rather than only a grading tool without increasing the instructor's workload too much? According to our previous study (Suhonen & Tiili 2021), the students spent a considerably long time in interaction with an automatically scored exam if they were allowed to. In that study, the students could retake an automatically scored "basic level exam" as many times as they liked or needed to pass the course. The total time they spent was many times higher than they traditionally do in paper exams. This finding encouraged us to further increase the weight of automated, recurring exams in the course. So, the solution to the question stated was, in our case, to use automatically scored, recurring, week exams that the students were able to take three times and increase their weight in the grading of the course. The exam arrangements are described in more detail in the next chapter.

Automatic assessment provides students with immediate feedback on their performance, allowing them to identify their strengths and weaknesses and adjust their learning strategies accordingly. It also provides consistent grading, reducing the possibility of subjective bias or grading errors that can occur when grading manually. Studies have found that students generally have positive attitudes towards automated assessment. For example, according to a study by Ardid and Vidaurre, student comments were generally positive, especially on ease-of-use and its usefulness during the learning process to diagnose the level achieved. On the other hand, there were also some criticisms, especially in terms of clarity of the questions and the rigidity of the automatic scoring (Ardid and Vidaurre, 2018). Overall, automatic scoring of exam answers can improve efficiency, consistency, and fairness in grading, while providing immediate feedback to students and saving instructor’s time.

2 METHODOLOGY

2.1 Engineering physics course - mechanics

The recurring, automated weekly exams were piloted on a bachelor’s level elementary engineering mechanics online course that was offered nation-wide. The platform was Moodle, and the setup was asynchronous implementation with weekly deadlines for exams and one final deadline for measurement assignments and final exam. In the beginning, the course had 167 active participants of which roughly half were students in Tampere University of Applied sciences, the rest were students in other universities of applied sciences in Finland. The course lasted for 10 weeks, and it had 6 weekly (or topic) exams which formed 60% of the course’s final grade. The rest of the points came from either from final exam or from measurement assignments. Measurement assignments are one-topic, relatively simple tasks, which doesn’t need very complicated equipment. In online courses, the equipment has to be easily available at home. Measurement assignments are described in more detail in our previous study (Suhonen 2021). With the week exams alone, it was possible to pass the course with two lowest grades (1-2). The maximum grade is 5. If the students aimed at better grades than 1-2, they needed also take the final exam or measurement assignments according to their own choice.

2.2 Weekly exams

On this piloted course, students had three attempts for each weekly exam and could use all materials during the online exams. Each attempt lasted a maximum of 2 hours, and students were able to retake the exam immediately, although they were encouraged to study between attempts. Each weekly exam had a deadline after which it was closed, and students were provided with a video to explain the solution to the problems. With these arrangements, we have been able to create an environment in which students view exams as an opportunity to learn and grow, rather than simply as a test of their level of knowledge. Additionally, the independence of time and place, together with multiple attempts, reduces the stress and anxiety that the examination situation could otherwise induce.

The weekly exams mainly consisted of the following types of questions: 1) multiple-choice questions requiring conceptual thinking, such as force diagrams; 2) problems based on diagrams, graphs, and measurement data; and 3) problems requiring mathematical solutions. The randomization of initial values was accomplished by using STACK exercises in Moodle. In all cases, there were multiple versions of the same type of question and/or the initial values for the problem were randomized. In this way, we tried to eliminate the possibility of the trial-and-error method in the exam, as well as reduce the feasibility of copying from peers. It was likely that the students had different versions of the questions each time they attempted.

One example of the questions is shown in the figure 1. It shows the question translated into English (A), the question as the students saw it (B) and part of the 12 different versions of the graph for the same question (C). Even though the question was the same in all attempts in this case, the graph changed.

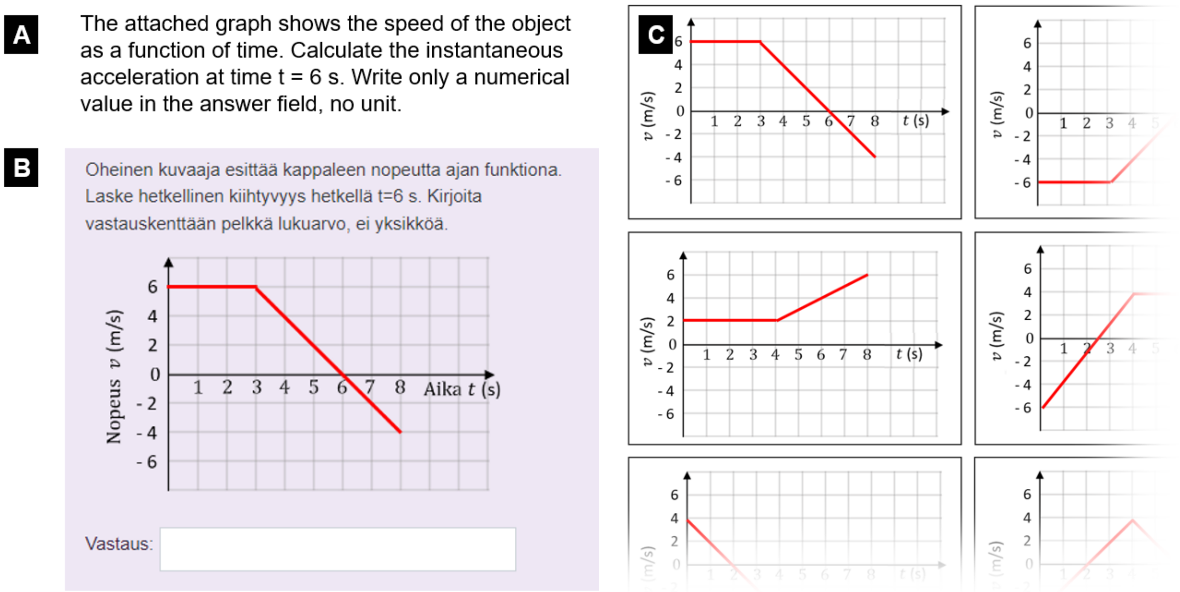

Fig. 1. Example of a week exam problem. A) The question translated into English. B) The question as the students see it. C) Part of the 12 different versions of the graph for the same question.

Another example of a week exam problem is shown in Fig. 2. This question had five similar, but slightly different set-ups (shown in B and C) of the problem. Here also the initial values are randomized (shown with red boxes in B).

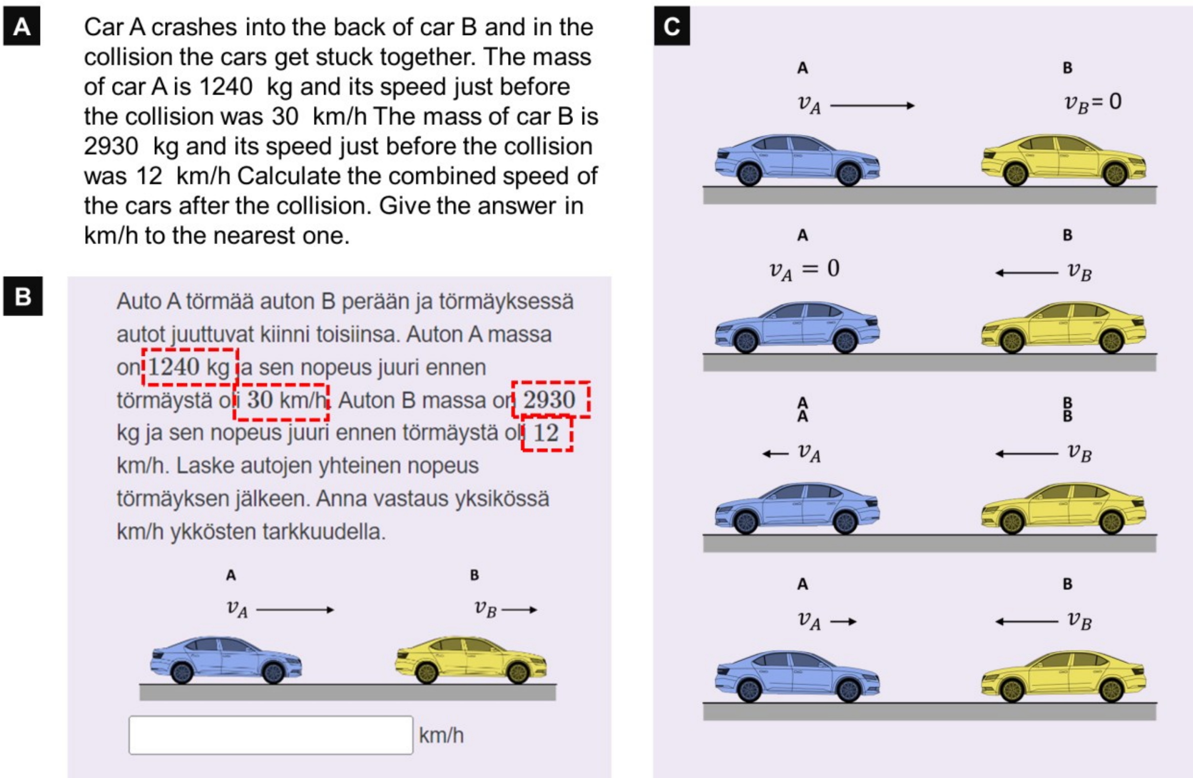

Fig. 2. Another example of a week exam problem. A) The question translated into English. B) The question as the students see it (excluding the red boxes). The randomized values are here indicated with the red boxes. C) Four other versions of the problem set-up.

The maximum number of points for each week exam was 6 and it was earned by solving 3-6 problems (depending on the exam). The scoring was automatic, which is easily done in multiple choice questions but a bit more laborious in other types. In mathematical solutions, the typical erroneous answers were fed to the answer tree in STACK-exercises in Moodle to yield partial points for partially correct solutions. This of course requires that the instructor knows the typical errors in advance. It is also possible to adjust the automatic scoring system after having a look at the student answers. To prevent losing all points for some typing or other small error, the student were asked to check their own answers after publishing the solution video for each weekly exam. If the students felt that they had lost points unfairly, they could ask the instructor for manual scoring.

3 RESULTS

3.1 Week exam results

There were 6 weekly (or topic) exams during the course. The last weekly exam was still open during writing this paper, and it is thus omitted in the analysis. Regarding the first five weekly exams, the number of students, average times and average points for each attempt are show in the table 1 and in the figure 3. The exams yielded similar data and the trends are roughly similar in each exam. The data shows that the majority of the students retook the exam at least once and more than one third on average used all attempts. In all exams, there are slight increases in the average points according to the number of attempts. The average time spend in interaction with one exam goes down from roughly 40 mins of the first attempt to a bit less than 30 mins of the third attempt. There is naturally variation in the times, since the topics and questions are different for each week. Altogether this means that on average, a student has spent 81 mins in one week exam, taken 2.1 attempts and spent 7.6 hours doing all the week exams. This is a remarkable time spent on high-intensity working on an exam situation.

Table 1. Attempts, durations and points for first five weekly exams. The maximum number of points for each week exam was 6 and the maximum time for each attempt 120 min.

Exam attempt	Number of students	Average Time (min)	Average points
1.1	167	53	2.64
1.2	119	38	2.76
1.3	81	34	3.27
2.1	166	41	2.82
2.2	120	32	2.87
2.3	86	27	3.13
3.1	159	41	2.84
3.2	94	29	3.06
3.3	55	26	3.42
4.1	155	40	2.62
4.2	117	24	2.56
4.3	79	18	3.02
5.1	149	33	2.19
5.2	117	28	2.16
5.3	81	27	2.87

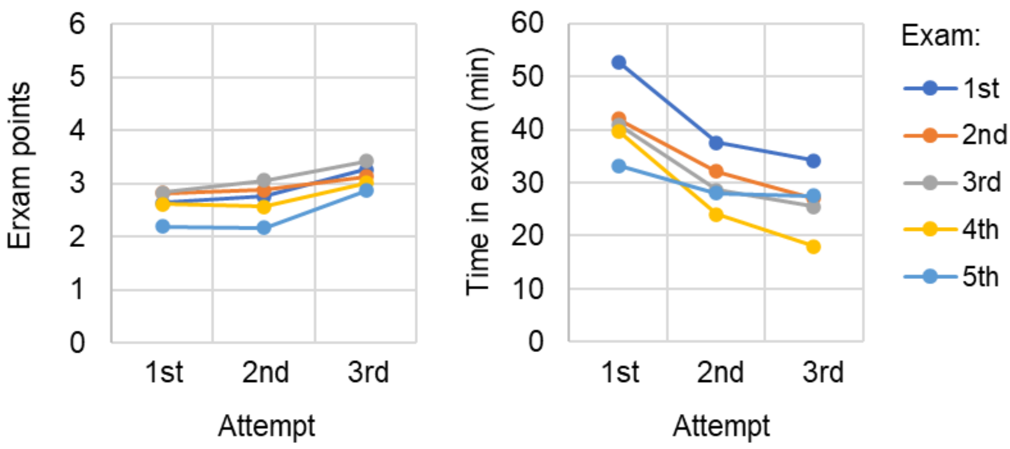

Fig. 3. Exam points (left) and spent time (right) as a function of attempt for the first five weekly exams.

The first exam is investigated a bit more deeply here, and the results are presented in the figures 4 and 5. The point distributions of the first week exam are shown in the Figure 4. It shows the ovarell effect of retakes on the points.The final point distribution is clearly higher than that of the first attempt.

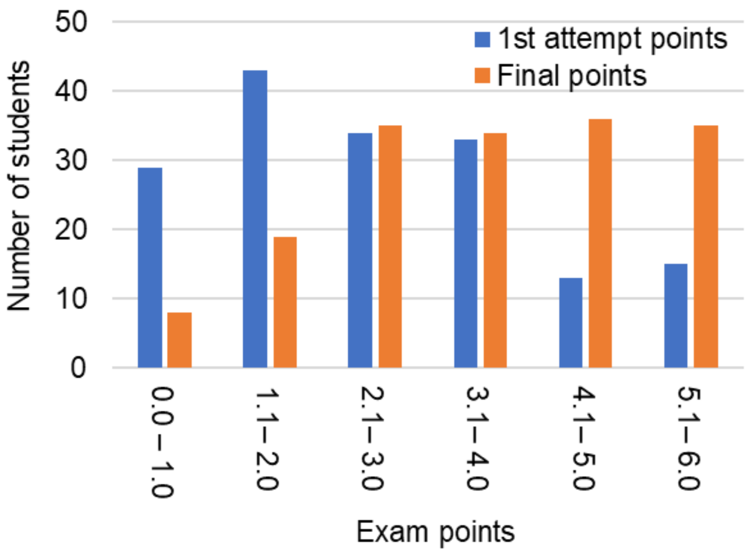

Fig. 4. The points distribution after 1st attempt, and the final points distribution in the first weekly exam.

Fig.5. shows student points and exam time categorized according to the success at the first attempt. The graph data contains only those 81 students who took all three attempts in first weekly exam. We can see that those students who got the lowest points (0-2) at first attempt had a highest increase in their points with successive attempts. Those who scored averagely (2-4) or highly (4-6) at their first attempt actually did worse at the second attempt. This suggests that they didn’t fully realize their need to study in between the attempts and maybe they just had an other try. Anyhow, students used their last attempt more wisely (on average) and got higher points at the last try than with the second try. Those who scored lowest in the first

attempt, used more time in the second and third attempt than those who scored averagely or highly.

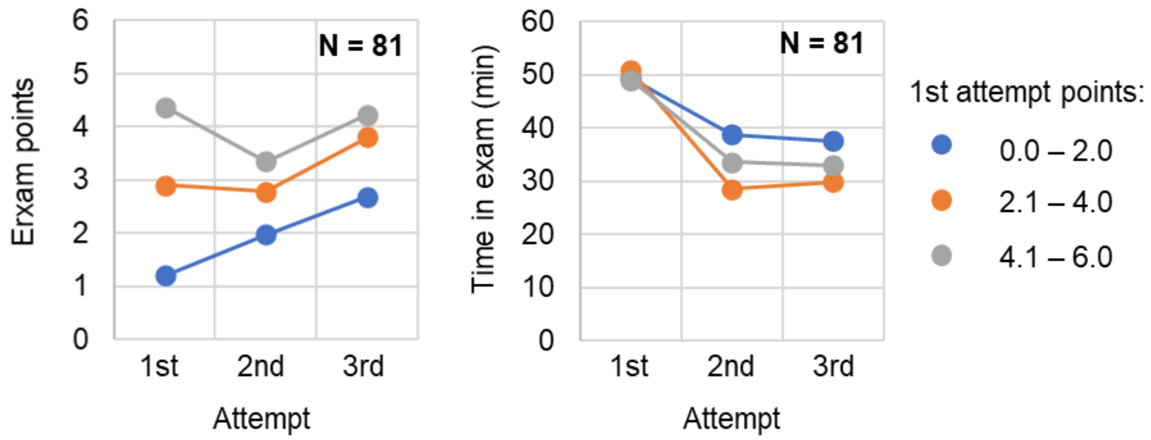

Fig. 5. Exam points (left) and spent time (right) as a function of attempt for the first week exam. The data is categorized according to the success at first attempt.

3.2 Student feedback

The student experiences of this recurring, automatically scored, weekly exams were surveyed using a short online questionnaire. There were 71 answers (43 %) and of them 70 reported that they had used more than one attempt in the weekly exams. This means that more than half of those students who have used many attempts in the exams have answered (see Table 1). Even though the answer percentage was not very high, the responses given truly represent student experiences with the many-attempt weekly exam. The survey contained multiple choice statements on 5-point Likert scale and open-ended questions. The summary of student answers to different statements are presented in the figure 6.

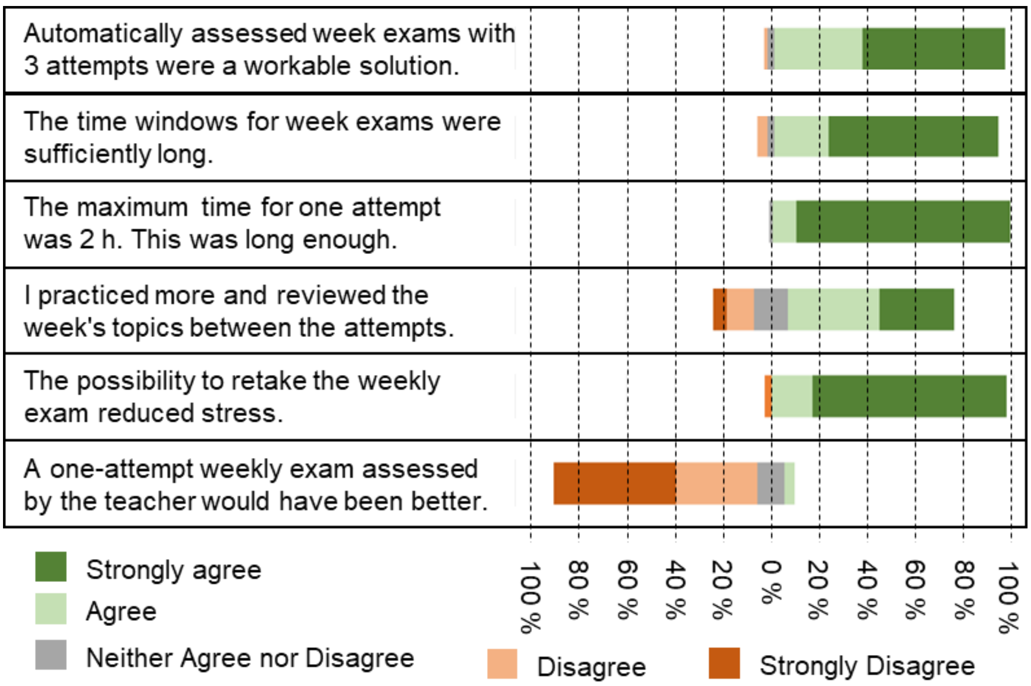

Fig. 6. The summary of student answers to different statements.

Fig. 6 shows the answer percentages to each statement from “strongly disagree” to “strongly agree”. The students found this kind of exam arrangement to work very well (96% agreed) and the time windows and the exam time was considered to be sufficiently long (93% and 99% agreed). The students were encouraged to study between the attempts, but there were no set technical limitations to prevent them from retaking anytime. According to their answers, 69 % (49 respondents) claimed that they had studied between the attempts. Since there were 90-120 student how retook the exam, this means that less than half of them actually did study more between the attempts. The possibility to retake the week exams reduced stress (97% of the respondents). Only a few students (4%) would have preferred to take the traditional exam scored by the teacher, whereas vast majority (85%) disagreed with that idea.

In the open-ended questions, the students were asked “*What was good about these weekly exam arrangements?”* and “*What could be improved in these weekly exam arrangements?*” The answers were analysed and categorized. It was counted, which issues the students brought up and how many times. Only the top four are presented here for both questions. The results indicate that in this exam implementation students valued: improved learning (21 respondents), flexibility (20), reduced stress (15), and immediate feedback (13). According to their answers to question “*What would you improve?*” they would like to improve: Nothing (36 respondents), point loss due to typing/rounding errors (12), more immediate right solutions to problems (8) and tips between the attempts (3). The rounding errors were taken into account in advance by giving exact instructions how to fill in the answers to the questions. This doesn’t unfortunately help to typing errors which easily led to total point loss to that question. What comes to right answers and solution methods, here we needed to balance between learning and assessing. It was chosen that the solution videos were available only after the exam was closed, not immediately after a student had used his/her last attempt. This was a compromise to reduce possibility to use peer’s Moodle account to watch the solution video before own attempt. For learning, some tips or even immediate release of the solution videos would be beneficial, of course.

4 SUMMARY

Automatically scored, recurring weekly exams were piloted in an online engineering mechanics course offered nationally through the Moodle platform. Most students took more than one attempt in the exams, and the majority of students who initially scored low points showed improvement in subsequent attempts. A survey of student feedback found that vast majority of students agreed that this exam arrangement worked well and that retakes reduced stress, was flexible and improved their learning experience and learning outcomes. Almost no one would like to change back to one-attempt exam checked by the instructor.

REFERENCES

Ardid, M., and Vidaurre, A. 2018. “Students’ perception of auto-scored online exams in blended assessment.” *Revista de la Facultad de Educación* 21.2 2018, 21(2), 79-103.

Pascoe, M. C., Hetrick, S. E., and Parker, A. G. 2020. The impact of stress on students in secondary school and higher education, International Journal of Adolescence and Youth, 25(1), pp. 104-112.

Richardson, M., Abraham, C., and Bond, R. 2012. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. Psychological bulletin, 138(2), 353.

Suhonen, S., and Tiili J. 2021. Continuous, versatile and flexible assessment method on engineering physics course. In Proceedings of the SEFI 2021 49th Annual Conference, 13.-16.9. 2021, Berlin, Germany. Technical University of Berlin.

Zunhammer, M., Eberle, H., Eichhammer, P., and Busch, V. 2013. Somatic symptoms evoked by exam stress in university students: the role of alexithymia, neuroticism, anxiety and depression. PloS one, 8(12), e84911.