STACK assignments in university mathematics education

Mäkelä Ari-Mikko
Project Researcher, M. Sc. (tech)
Tampere University of Technology
Tampere, Finland
E-mail: ari-mikko.makela@tut.fi

Ali-Löytty Simo
University Lecturer, D. Sc. (tech)
Tampere University of Technology
Tampere, Finland
E-mail: simo.ali-loytty@tut.fi

Humaloja Jukka-Pekka
Doctoral Student (tech)
Tampere University of Technology
Tampere, Finland
E-mail: jukka-pekka.humaloja@tut.fi

Joutsenlahti Jorma
Senior Lecturer in Mathematical Education, Ph. D.
University of Tampere
Tampere, Finland
E-mail: jorma.joutsenlahti@staff.uta.fi

Kauhanen Janne
Lecturer, Ph. D.
Tampere University of Technology
Tampere, Finland
E-mail: janne.kauhanen@tut.fi

Kaarakka Terhi
Lecturer, Ph. D.
Tampere University of Technology
Tampere, Finland
E-mail: terhi.kaarakka@tut.fi

Conference Key Areas: Mathematics and Engineering Education

Keywords: STACK, web-assisted learning tools

INTRODUCTION

E-learning methods and especially web-based assignments have become more and more common at schools and universities in recent years. [1, 2] For example, at Tampere University of Technology (TUT) information technology and web-based learning tools are continuously being developed in order to improve education in mathematics. The role of e-learning and information technology in education are also increasingly emphasized in the Finnish National Curricula [3].

Students' learning process can be assisted and diversified with the help of e-learning tools such as Moodle. In mathematics education at TUT, Moodle and other such dynamic web-based learning environments have been used for several years for delivering learning material and assignments. However, interactive learning tools that can receive and check students' answers or give instant feedback to students have not been commonly used previously. Some experiments have been made though, and a web-based program for automatic answer checking has been used for more than ten years in the mathematics basic skills test for the first year students of TUT [4].

Now the aim is to utilize such software that delivers assignments, checks students' answers and gives feedback to the students in the engineering mathematics courses. The software that has been used is called STACK (System for Teaching and Assessment using a Computer algebra Kernel) which can be integrated into Moodle [5]. STACK assignments have been created as a part of STEM education material bank project Abacus which is currently carried out in collaboration between technical universities in Finland and Portugal [6]. The project aims at producing high-quality co-developed materials for the partners.

STACK is a system that uses computer algebra system Maxima to deliver assignments and establishes the mathematical properties of expressions entered by the student. In STACK, students have to write their answers as mathematical expression instead of just selecting the correct answer as in the multiple choice question in which the student may guess the correct answer. It was first created by Christopher Sangwin and the first version was published in 2005. In STACK, one can create structured questions including randomized elements. Thus, it is possible to generate individualized exercises for the students, and even if they would work in groups, each student has to solve their own exercises rather than copy the answers from one another. In addition, written feedback which may include mathematical computations based on the student's answer can be generated in STACK assignments as necessary. [2, 5]

Race & al. [8] state that feedback is a vital criterion for a good web exercise. In a computer aided system such as STACK the feedback is also instant. For example, in mathematics exercises students can immediately find out whether their answers are correct or not, and possibly what mistakes they have made. Feedback guides the students to identify their errors and revise them. It can also motivate the students to

try again after giving a wrong answer, and this learning by trial and error can be done in the comfort of privacy. Furthermore, with the help of automatically generated feedback, the students can self-review their learning process. Since the feedback in STACK is automatically generated, students can solve problems and get tips as feedback on their answers whenever and wherever they want.

STACK was adopted as a part of exercises in four basic mathematics courses at TUT in the semester 2015-2016. The courses were Engineering Mathematics 1-2 (EM1-2) and Honours Mathematics 1-2 (HM1-2). There were around 130 students enrolled on HM and 180-210 on EM courses. Each course lasted for one period (7+1 weeks). The contents of the courses covered roughly SEFI 1 level excluding probability calculus [7] and are worth five credits each. Altogether, all the students of TUT are required to complete 15 credits of SEFI 1 level mathematics courses and 12 credits of courses that measure up approximately the level of SEFI 2.

In this paper we introduce a fairly new way to deliver mathematics problems that students can solve independently. The students' activity in STACK during the lecture week and its relation to the exam grades were examined. We found out that there is a correlation between the exam grades and the time that the answers in STACK were submitted. It was also discovered that the HM students made their submissions earlier than the EM students.

Rasila & al. [9] studied some of these questions in their research paper that concerned the use of STACK in a basic mathematics course. They observed that the number of participants in STACK exercises was higher than the one in the traditional exercises throughout the course. It was also discovered that the activity in STACK exercises peaked near the submission deadlines. Rasila & al. [10] found that the correlation between the amount of exercise problems solved and exam scores did not hardly differ when traditional and STACK exercises were compared. They also made the unsurprising observation that more the STACK and traditional problems were solved the better the exam grade in general. Same result was also observed by Paiva & al. [11].

We begin by introducing the research questions in section 1. The research methods are discussed in the second section in which we also show how does the STACK assignment appears to the student. In section 3 we introduce and analyze the results observed in this study. Finally in section 4 we compare our results to the other studies and draw the conclusions about the results.

1 RESEARCH QUESTIONS

This study concerns the use of STACK in the courses mentioned above. Especially we are interested in

- do the points gathered and the time of the last submission in STACK exercises affect the exam grades?
- when and for how long do the students solve the STACK assignments?
- how does the activity in STACK differ between HM and EM students?

These questions were also studied in a master's thesis that considers students' use of STACK in EM1 and HM1 [12].

2 METHODS

There were six weekly exercises in the courses mentioned above. Each exercise included from eight to ten assignments of which three or four were implemented with STACK. In STACK assignments, students were able to give their answers in Moodle. Students had one week to solve the exercises and return their answers. In HM1 the students had to also submit the full, written solutions for the STACK problems in PDF to Moodle. In general, student's attempts in STACK exercises were not limited, so it was always possible to have another try until the correct answer was found. Other assignments were solved in the weekly problem classes which were arranged in the university campus.

In *Fig. 1* it is shown how does a STACK assignment look like to the student. First, the problem and syntax hints are given. Below those, there are the input fields with the white background. Once the student has written the answers and checked them, the feedback is shown on the yellow background. In this feedback it is told whether the answer is correct together with the marks for the submissions. Also, hints are given if the answer is incorrect.

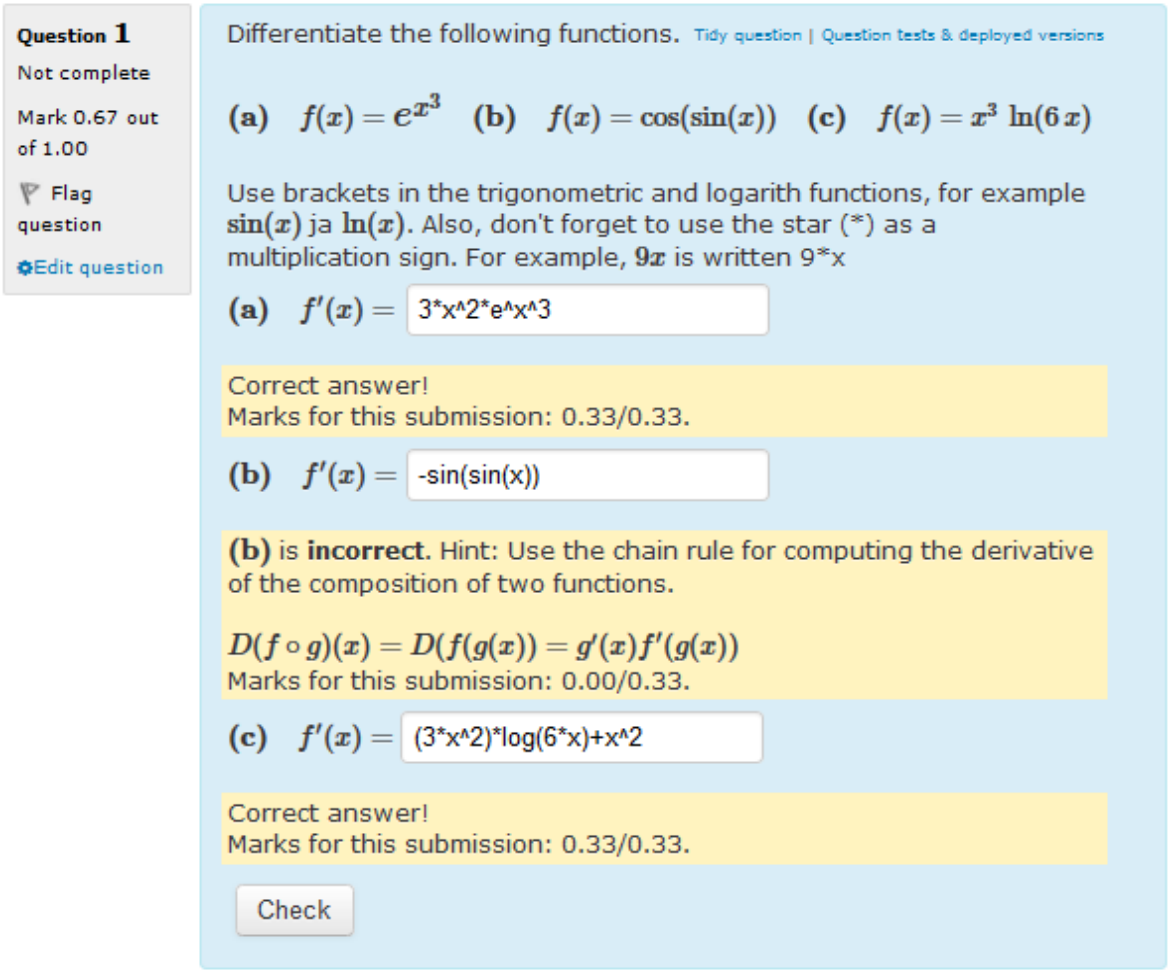

Fig. 1. STACK assignment in use. Once the student has given the answers in the input fields and checked them, STACK gives the feedback on the yellow. background.

All the student activity related to the STACK assignments was saved in the Moodle logs. Data was analysed with Matlab by means of educational data mining [13]. We used regression analysis to explore how the activity in STACK predicts the grades in the exams. Moreover, statistical classification was used to determine, how did the students who succeeded well in the exam act in Moodle while exercising with STACK compared to those who got lower grades. Students' experiences about STACK were also surveyed in the master's thesis with a poll that was arranged in the EM1 and HM1 courses. We will refer to these polls in this paper.

3 FINDINGS

We observed that the students earned points from STACK exercises quite well. That is, the students kept solving the problems until they got the correct answers. When both two courses from EM and HM are considered, the average percentage of full points gained in all the exercises is 93 % in EM and 96 % in HM. It seems that the students were able to solve even the more challenging assignments quite well with the help of hints in STACK. Of course, the students could have assisted each other as well when solving the problems. The comparison between the points gathered in STACK and the exam grades is shown in *Fig. 2* in which the weekly means for each grade group are evaluated.

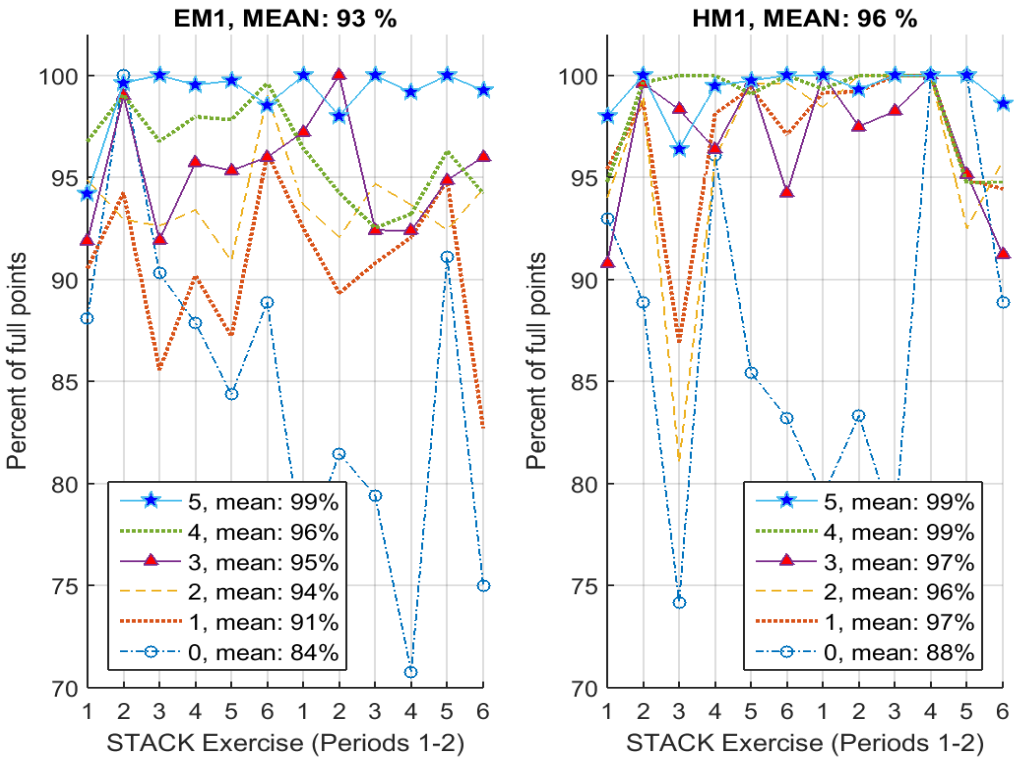

Fig. 2. The percentage of full points gathered in STACK in each exercise week for each grade group. Students who did not pass the exam also gathered less points in STACK exercises. Conversely, those who succeed in the exam also gathered the highest points in STACK. The difference is clearer in EM than in HM.

Certainly those who did not pass the exam (grade 0) also gathered fewer points in STACK than the others. Roughly, the better the grade the more points gathered in the weekly STACK exercises. The difference is more explicit in EM, but a same kind of observation can be made in HM as well. The correlation coefficients between grades and mean percentage of points gained in STACK were 0.33 in EM1 (p-value < 0.001) and 0.35 (p < 0.001) in EM2. In HM1-2 the correlations were 0.35 (p < 0.001) and 0.24 (p = 0.03), respectively. The correlations are Spearman rho's and p-values were computed with the alternative hypothesis 'correlation is not zero'. The Pearson linear correlation coefficients and their p-values were nearly the same as Spearman rho's. Only the p-value for HM2 were somewhat greater, 0.12.

Study shows that the HM students submitted their final answers in the STACK exercises generally earlier than the EM students. This can be clearly seen from *Fig. 3* in which it is shown how the times of the last submissions are distributed during the lecture week. We notice that the final submissions were mostly made within the 24 hours before the deadline. However, while in EM more than 50 % of the students made the last submissions during the last day almost every lecture week, in HM the students made the submission earlier. In some weeks, most of the last submissions in HM were made in the first day the exercises were published.

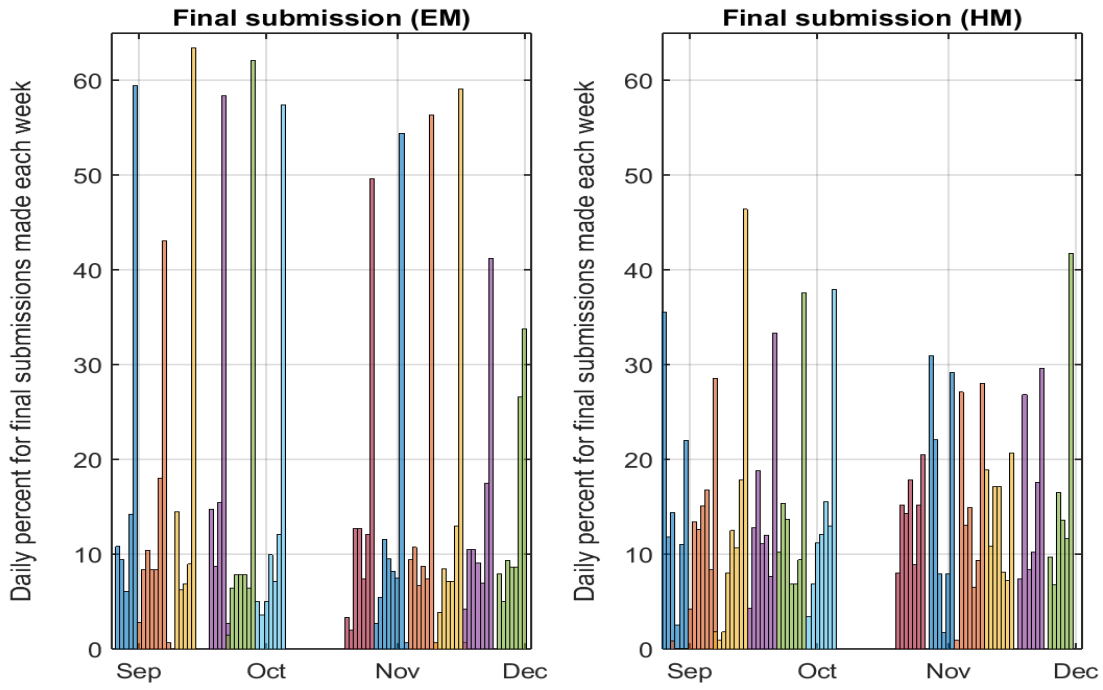

Fig. 3. The final submissions made in STACK exercises focus on the last 24 hours before the deadline. Different colours represent the separate lecture weeks. In general, the HM students made the last submissions earlier than the EM students. The empty gaps in the figure indicate the exam weeks when lectures or exercise sessions were not held.

It was observed that there is a relation between the exam grades and the time in which the last submissions in STACK are made. The students who made the submissions earlier performed better in the exam. This can be clearly seen from *Fig. 4* which includes median times of the last submissions in each exercise with respect to the exam grades given in the first period. The correlation coefficients between grades and median times of submission were 0.29 in EM1 (p-value < 0.001) and 0.30 (p < 0.001) in EM2. In HM1-2 the correlations were 0.43 (p < 0.001) and 0.26 (p = 0.018), respectively. Again, the correlations are Spearman rho's and p-values were computed with the alternative hypothesis 'correlation is not zero'. The Pearson linear correlation coefficients and their p-values hardly differed from Spearman rho's.

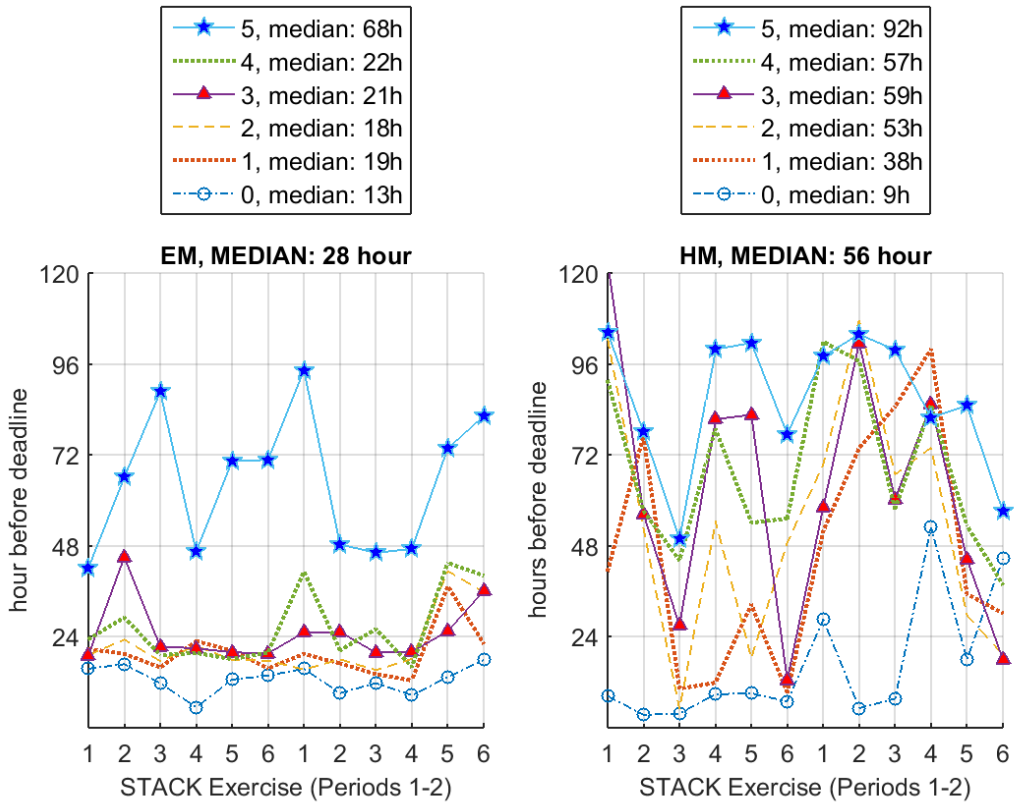

Fig. 4. Times of the final submission (median) in each exercise compared to the exam grades. Note that the times of submissions through the periods are compared to *the first period* exam grades. The figure shows that HM students made the submissions earlier than EM students. Also, those who made the submissions earlier generally were more successful in the exam.

The difference is more explicit in Engineer Mathematics. Those who got the best grades in the exam submitted their final answers earlier than the others. Also, in both EM and HM those who got the lowest grade (0), generally made the latest submissions. There is also a difference between EM and HM. In *Fig. 3* we already observed that the students in HM made their submissions earlier compared to those in EM. The median time of the final submissions in EM was 28 hours and in HM 56

hours before the deadline. Roughly, the students in HM generally made the submission one day earlier than the EM students.

The means and standard deviations for the times of the final submissions for each grade are shown in Table 1. Again, we analyse the times of the submission with respect to the *first period exam grades*, so we can explore the development for the same population in all the grade groups. That is, the times of the final submissions in EM1-2 and HM1-2 are sorted with respect to the grades given in EM1 and HM1, respectively. In *Table 1*, the results for each of the grades are compared with the other grades in order to determine whether the differences in the times are statistically significant in the significance level of 1 %. The differences between the two courses for each of the grades are also compared in both EM and HM.

Table 1. The means and standard deviations of the times of the final submissions for each grade. The figures in parentheses imply statistically significant differences in the significance level of 1 % between the times of the final submissions for separate grades. For example, in EM1, times of the final submissions for the zero graders differ significantly with those who got the grades 3, 4 and 5. The star implies the significant difference for the same populations in the different courses.

Grade/ Course	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
EM1	26 ± 13 (3,4,5)	26 ± 8 (2,3,4,5)	39 ± 7 (1,5)	41 ± 8 (0,1,5, *)	44 ± 7 (0,1,5)	66 ± 10 (0,1,2,3,4)	42 ± 40
EM2	29 ± 15 (3,4,5)	27 ± 7 (2,3,4,5)	46 ± 10 (1,5)	52 ± 7 (0,1,5, *)	52 ± 10 (0,1,5)	68 ± 9 (0,1,2,3,4)	50 ± 47
HM1	22 ± 14 (1,2,3,4,5)	45 ± 14 (0,3,4,5,*)	52 ± 21 (0,5)	61 ± 16 (0,1,5)	65 ± 9 (0,1)	81 ± 14 (0,1,2,3)	58 ± 49
HM2	36 ± 18 (1,2,3,4,5)	60 ± 15 (0,5, *)	62 ± 17 (0,5)	62 ± 15 (0,5)	70 ± 19 (0)	81 ± 14 (0,1,2,3)	64 ± 47

From the table above, we see that those who were given grade 5 in EM1, also submitted their final answers in STACK somewhat earlier than the others and the difference to all the other grades is statistically significant. The same goes for HM1 except that the difference between grades 4 and 5 is not significant in the 1 % significance level. However, the results for those who did not pass the exam (grade 0) differ significantly with all the others in HM. Altogether, it is appropriate to say that the earlier the final submissions, the better the grade. The results do not differ significantly between the two courses, except for grade 3 in EM and grade 1 in HM.

The correlation between the amount of the solved STACK assignments and the exam grades were studied in Mäkelä's thesis [12]. This correlation was compared to

the correlation between the amount of solved *traditional* exercises (those solved in the weekly problem classes at the university) and the exam grades. We noticed that these correlations were almost the same. From this we may assume that solving the STACK exercises was as important as solving the traditional exercises in order to succeed in the exam.

The time that the students spent between the first click in STACK and the last submission was also explored. Since the amount of assignments in the exercises differed, the duration in each exercise was divided by the amount of the assignments. The results for each exercise are shown in *Fig. 5*.

There is a slight difference between the EM and the HM students. Median duration in EM was 2.4 hours per assignment while in HM it was 4.0 hours. However, in the second period the median durations in HM were lower than in EM. It seems that while the median durations in EM were pretty similar throughout both periods, the HM students spent less time in STACK in the second period for a single assignment.

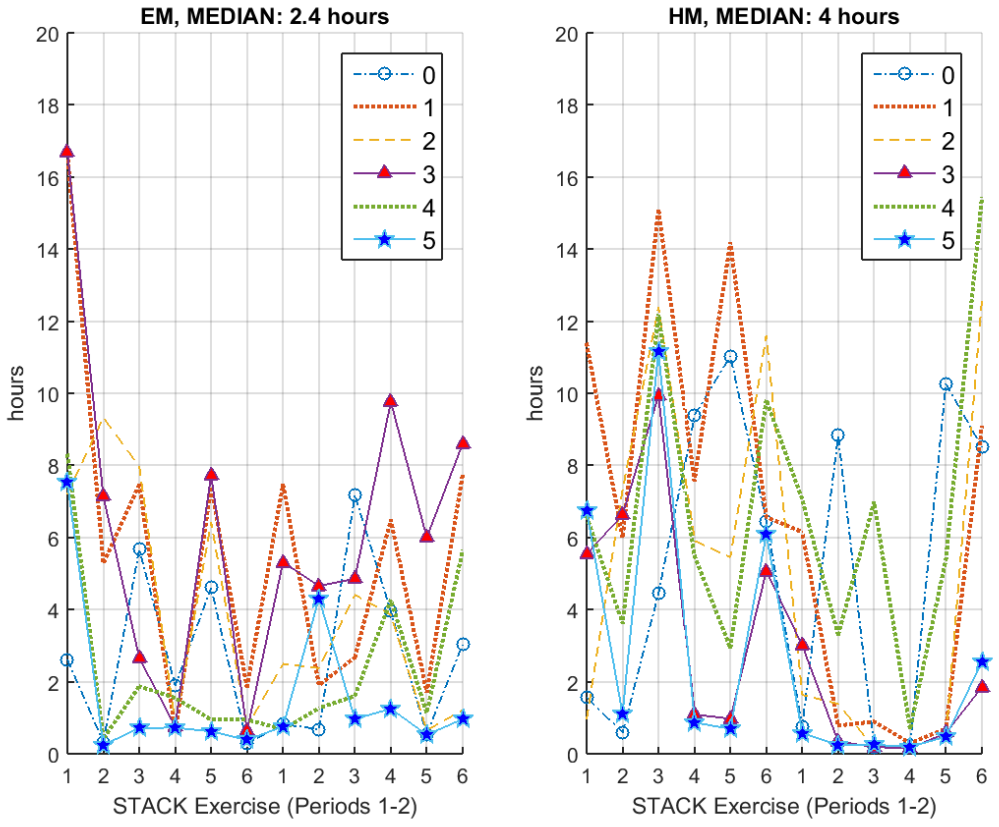

Fig. 5. The time spent (median) in STACK exercises from the first click to the last submission. Since there were different amount of assignments in the exercises, the time has been divided by the amount of the assignments. HM students seem to spend more time per assignment. It was also discovered that the durations among the HM students decreased significantly in the second period.

In Fig. 6 it can be seen how many students returned their answers within a given time interval from the first click. The diagram indicates the means for all exercises for three grade groups. From the figure we see that relatively small percent of students return their final answers within half an hour, except HM2 in which the percent is bigger in all grade groups. The most of the students that returned answers in STACK made their final submissions within 24 hours from the first click.

Most of the students answered at least something in STACK apart from the zero graders in EM. Under 60 % of them left their answers in STACK while over 90 % of the other students made the submissions. Slightly greater amount of zero graders returned their answers in HM. Students who got grades 1-3 were nearly as active in solving STACK problems as those who obtained grade 4 or 5. However, the latter group made their submissions earlier on average. The difference is clearer in EM. Over half of the 4-5 grade population spent under four hours in STACK, which can be considered the time required for a single exercise session, while the portion is clearly under 40 % among the other grade groups.

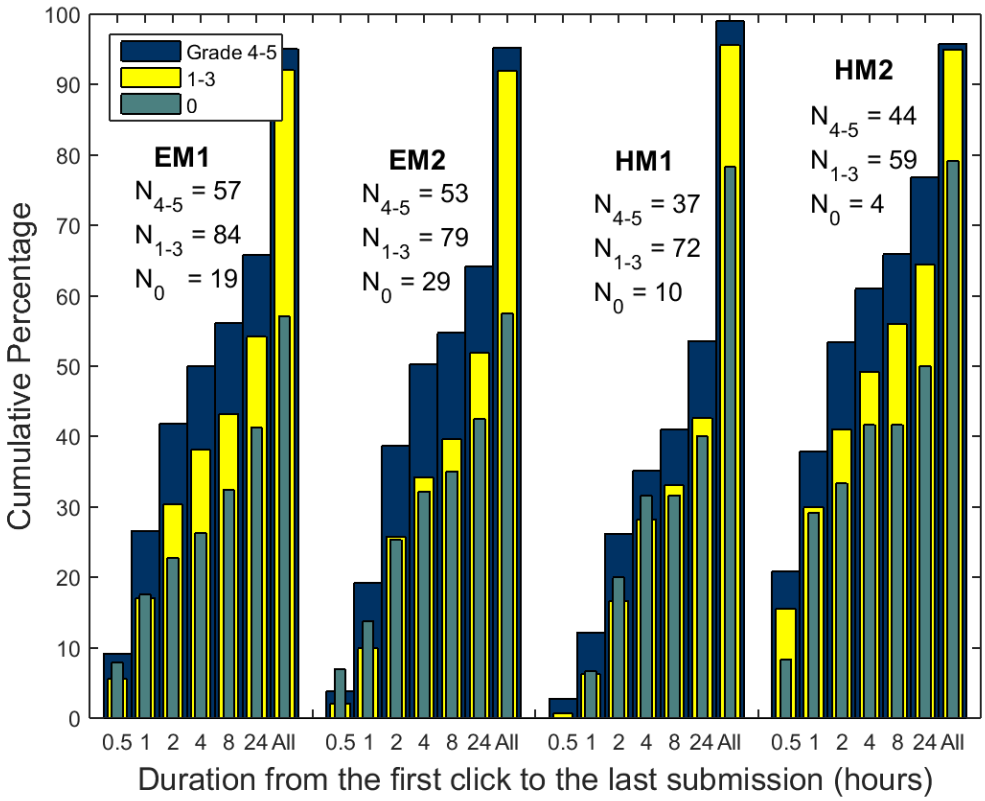

Fig. 6. Cumulative percentages of the durations in STACK from the first click to the final submission. Quite small amount of students returned their final answers within half an hour. Note the difference between HM1 and HM2: clearly, the students spent less time in STACK in HM2, while in EM there is virtually no difference between the courses.

As can be seen from Fig. 6, the durations in STACK exercises were quite similar in EM1 and EM2. However, there is a notable difference between HM1 and HM2. Clearly a greater amount of students made their submissions under two, four, eight or 24 hours in HM2 than in HM1. This explains the difference that was observed in the median durations in Fig. 5 between these courses. The difference is highly significant: according to a Student's t-test, the p-value is less than 0.001 when the durations for all the students in HM1 and HM2 were compared.

The difference between HM1 and HM2 may presumably be explained by the PDF submissions that took part in the HM1 course. Although solving the problems and giving the answer into STACK should not differ whether they were required to make the PDF submission or not, the students may have been considering assignments more thoroughly and writing the solutions more carefully while solving the STACK exercises. This kind of behavior might explain the differences in the durations in HM between the two periods.

4 CONCLUSIONS

In this paper we focused on STACK, a web-based system for delivering engineering mathematics problems that automatically checks the answer and gives feedback to the student. We contemplated how the students act while solving problems in STACK and is there a relation between exam marks and the time when the submissions in STACK are made or the points gathered in STACK. The differences between Honours and Engineering Mathematics students were also studied.

The first research question considered whether the points gathered and the time of the last submissions in STACK exercises affect the exam grades. It was discovered that, indeed, there is a correlation between the points gathered in STACK exercises and the exam marks. Students who earned points well were more likely to achieve the highest grades as well. On the other hand, zero graders distinguished from others by collecting the least points in the STACK exercises. This result agrees with the observations made by Rasila & al. [10] as well as Paiva & al. [11].

Along with the points gathered, there was also a correlation between the time the last submissions were made and the exam grades. The students who gave their final answers in STACK earlier than others also performed generally better in the exam, while those who did not pass the exam made the latest submissions. We may not draw a causal connection between these two, yet the result implies that students who are more interested in mathematics and solving problems, and who do not procrastinate with solving the problems and giving their answers, also get the better grades.

The second objective was to determine when and how long do the students solve the STACK assignments. We observed that the activity in STACK is greatest near the deadline, that is, the most of the students made the last submission less than 24 hours before the deadline. Rasila & al. made the same unsurprising discovery in their research [9]. It was also found out that HM students made the submissions roughly

one day earlier than EM students. The times of the last submissions in HM distributed way more uniformly during the lecture week.

Those who got the highest grades in the exam also spent less time in STACK while solving the problems than those who were not that successful in the exam. However, the differences were not as definite as in the times of the last submissions. It was observed that there were no statistically significant differences in the time the students spent in STACK between EM and HM students. However, we found out that the difference in the durations was highly significant between HM1 and HM2 courses. The students spent clearly less time in HM2. Unlike in HM2, in HM1 the students had to also submit written solutions of the problems in PDF to Moodle in addition to submitting their answers into STACK, which probably explains why the students spent more time in STACK in HM1.

Most likely, the PDF submissions decreased the use of WolframAlpha and such calculators. Therefore, if we want students to solve the problems without any calculators or software, requiring full written solutions, not just final answers, is a fine way to prevent the use of such tools. The question is, whether we want this or not. On the other hand, it is reasonable to rehearse the use of tools like Matlab and Maxima. Also, the aim is to create problems that cannot be solved just by putting the problem in WolframAlpha. Sangwin and Köcher have discussed more about these kinds of issues in their recent paper [14].

The third question considered was about differences between HM and EM students. The most substantial difference that was detected was about the time the students give their final answers in STACK. Clearly, HM students made their last submissions earlier than the EM students. The difference of the median durations was 28 hours. HM students also gained a little more points in STACK than EM students, and in average spent less time while solving a single assignment. The points gained or the time spent in STACK did not differ significantly when compared to EM.

The contents of HM courses are more profound and that is why there are probably more students in HM who are more enthusiastic about mathematics. This may explain the results discussed above. According to the survey poll arranged in EM1 and HM1, many HM students considered STACK exercises as introducing problems for the upcoming lectures [12]. On the other hand, some of the HM students felt them a fine way to revise the subjects considered in the lectures. There were not these kinds of mentions in the EM students' answers. This may also explain for its part why the HM students generally made the submissions earlier than the EM students.

According to the survey polls arranged in EM1 and HM1, the students feel fairly positive about solving problems with STACK. This is supported by the observation that the students solved and returned the answers to the STACK problems quite actively. Although many students mentioned the syntax being occasionally somewhat challenging, it was not considered too difficult. 58,7 % of the students (N = 249) disagreed or strongly disagreed and only 4,0 % strongly agreed with the statement "Writing the answer in STACK was too hard".

The written feedback the students gave in the open questions was quite promising. Earlier studies describe the same kind of experiences [9, 10]. The students noted that one can solve problems and learn mathematics with STACK as well as by

solving traditional exercises. Especially the instant feedback and the hints to the wrong answers were acknowledged which verifies Race's & al. [8] observations about the importance of feedback. The students also showed gratitude to the freedom to solve problems in STACK whenever and wherever they wanted. In Mäkelä's thesis [12], it was observed that the points gathered in STACK correlate with the exam grades as well as the points gained in the traditional face-to-face instructed exercises. The students' opinions discussed above confirm this result that was also discovered by Rasila & al. [10].

In addition with earlier studies, this research implies that students have accepted STACK as part of exercises. Regardless, in [10] as well as in [12] the students have noted that the traditional exercises must not be forgotten. Students would rather have a mix of traditional and STACK exercises. As STACK assignments are further developed, students will be able to solve problems without instructor whenever and wherever they want. This kind of independent learning with STACK sets a ground for better online courses. With STACK, it is also possible to execute electrical exams in mathematics courses.

REFERENCES

[1] Rosenberg, M. J., E-learning – Strategies for Delivering Knowledge in the Digital Age, (2001), McGraw-Hill, New York, 343 pages.

[2] Sangwin, C. Computer Aided Assessment of Mathematics, (2013), Oxford University Press, Oxford, 172 pages.

[3] Opetushallitus. Lukion opetussuunnitelman perusteet 2015, 279 pages. http://www.oph.fi/lukion_opetussuunnitelman_perusteet_2015

[4] Pohjolainen, S., Raassina, H., Silius, K., Huikkola, M., Turunen, E., (2006), TTY:n insinöörimatematiikan opiskelijoiden asenteet, taidot ja opetuksen kehittäminen, Research Report 84, 135 pages.

[5] Sangwin, C. Who uses STACK? A report on the use of the STACK CAA system, (2010), http://web.mat.bham.ac.uk/C.J.Sangwin/Publications/2010-3-1-STACK.pdf

[6] STEM Education Material Bank Abacus. https://abacus.aalto.fi/

[7] A Framework for Mathematics Curricula in Engineering Education, (2013), The European Society for Engineering Education, A Report of the Mathematics Working Group, 88 pages. http://www.sefi.be/wp-

content/uploads/Competency_based_curriculum_incl_ads.pdf

[8] Race, P., Brown, S., Smith, B., (2005), 500 Tips on Assessment, RoutledgeFalmer, New York, 184 pages.

[9] Rasila, A., Harjula, M., Zenger, K., (2007), Automatic assessment of mathematics exercises: Experiences and future prospects. Symposium of Engineering Education, Helsinki University of Technology, Helsinki, pp. 70-80.

[10] Rasila, A., Havola, L., Majander, H., Malinen, J. (2010), Automatic assessment in engineering mathematics: evaluation of the impact. Symposium of Engineering Education, Aalto University School of Science and Technology, Espoo, pp. 37-45.

[11] Paiva, R., Ferreira, M., Mendes, A. (2015), Interactive and Multimedia Contents Associated with a System for Computer-Aided Assessment. Journal of Educational Computer Research, pp. 224-256.

[12] Mäkelä, A-M., (2016), E-learning methods in the elementary courses of the university mathematics, Master's thesis, Tampere University Of Technology, 92 pages.

[13] Peña-Ayala, A., (2014), Educational Data Mining, Applications and Trends, Springer International Publishing, New York, 468 pages.

[14] Sangwin, C., Köcher, N., Automation of mathematics examinations, (2016), Computers & Education 94, pp. 215-227.