

Research paper

Automatic assessment in engineering mathematics: evaluation of the impact

Antti Rasila, Linda Havola, Helle Majander and Jarmo Malinen

Department of Mathematics and Systems Analysis,

Aalto University School of Science and Technology, Finland

antti.rasila@tkk.fi, linda.havola@tkk.fi, helle.majander@tkk.fi, jarmo.malinen@tkk.fi

Abstract

We study the impact of a web-based automatic assessment system STACK in teaching mathematics for engineering students. We describe several uses of automatic assessment which have been tested in Aalto University during the past few years. We measure the impact of e-assessment to learning outcomes in engineering mathematics. This question is motivated by the practical need to show that the system is, in general, worth the effort invested but also our wish to better understand the learning process. The secondary aim is to obtain information about the different factors affecting the learning outcomes that would be useful in further improving mathematics teaching. Our goal is to show that such system can significantly activate students, allow much increased flexibility in practical arrangements of teaching, and facilitate innovative practices in, e.g., diagnostic testing and grading students' work.

Keywords: Automatic assessment, mathematics, progressive assessment

1. Introduction

Computer aided assessment (CAA) system STACK has been used in Aalto University School of Science and Technology since 2006. The system consists of a computer algebra system (CAS) for evaluating symbolic expressions, a web-based user interface, and a database for storing the exercise assignments and the student solutions. STACK is an open source software licensed under the GPL [7]. It was originally developed by C. Sangwin [16, 17] in the University of Birmingham, but the system has been further adapted for the requirements of engineering mathematics courses in Aalto University [8, 14]. For a technical description of the system and basic examples about its applications, we refer to [8] and [13]. Since the initial testing in 2006, the system has been taken into use for almost all engineering mathematics courses at Aalto University. In fact, we believe that we are the largest user of STACK in the world at the time of writing this paper.

We consider three particular applications of STACK. First, we briefly outline results from the diagnostic mathematics starting skills testing by using STACK that has been introduced to all our new students in 2008 and 2009. Second, we study experiences on automatically assessed exercise assignments on the course Mat-1.1210 Basic course in mathematics S1. It is the first of the three compulsory mathematics courses for electrical

and telecommunications engineering students. About 200 first year students enrol to this course each year, and automatic assessment has been used since year 2007. Third, we discuss the motivation and results of the experimental course Mat-1.2991 Discrete mathematics which was taken by 58 students. About half of the students were computer engineering majors. In this course, web-based automatically assessed problem assignments constituted also an essential part of the final grade. The goal of this experiment was to activate students, and to balance their workload more evenly throughout the duration of the course. In the both courses, STACK was used as a component in blended learning [6]; i.e., traditional lectures and exercise sessions were used together with e-assessment. We remark that also pure e-learning approaches have been experimented with [2] but in too small scale to provide sufficient data for statistical analysis.

2. Review of literature

CAA has been relatively popular in teaching computer science, and the impact of e-assessment has been studied in [1, 9] and [19]. To our knowledge, no wide-scale research of the impact of using such system in teaching mathematics, at least not at university level, has been pursued earlier.

E-learning methodologies in teaching university level mathematics have been studied by M. Nieminen [11], although his recent PhD thesis does not involve e-assessment. In the study, course results were compared by covariance analysis: scores of the final tests were scaled to correspond to each other by means of the item response theory. The main conclusion was the following: there is no statistically significant difference between the results of the students who studied on an on-line course compared to those who were attending traditional lecture-based teaching. Some problems with the technology were reported; the training portal proved unsuitable for studying mathematics. These findings underline a need for specialised software (such as STACK) for teaching mathematics.

3. Research problems and methodology

The main objective of this research is to measure the impact of e-assessment to learning outcomes in engineering mathematics. This question is motivated by the practical need to show that the system is, in general, worth the effort invested but also our wish to better understand the learning process. The secondary aim is to obtain information about the different factors affecting the learning outcomes that would be useful in further improving mathematics teaching.

It is a difficult question in itself what we should understand by learning outcomes. In principle, there are three main philosophical world views one should consider here: positivist, constructivist and pragmatist. Positivists hold a deterministic view about the expected causes that determine effects or outcomes of human actions. The positivist viewpoint emphasizes the role of the underlying causes (or laws) to be discovered using experiments and statistical testing of data. Constructivists, on the other hand, hold the assumption that individuals seek to understand the world where they live and work in their own terms, thus making it crucial for the researcher to describe their subjective experiences. Consequently, methodologies related to constructivist studies are usually

qualitative. The third position, arising from the philosophy of Peirce and others, is pragmatism. As a world view, pragmatism refers to actions, situations, and observed consequences rather than inferences from preceding events and circumstances as in positivism. There is a concern of what works, and how. Instead of focusing on methods, pragmatists emphasize the research problems and use all available means to understand them. [5]

In this study, we adopt the pragmatist position for practical and philosophical reasons given below. First, it would be difficult to arrange a large scale experiment in real world conditions that would be controlled enough to provide reliable and systematic data of experiments leading to positive knowledge. Indeed, skills and attitudes of new students change from year to year, and it is problematic to accurately measure if this has relevance to the conclusions. However, the starting skills test (considered later in this paper) is a partial solution, but it has been only available since 2008. We do not have comparable earlier data. Subjective learning experiences involving automatic assessment are certainly interesting, and they remain an object for future studies. On the other hand, the constructivist view is at odds with the practical motivations of our research which have an inherit perspective of an outside observer – we aim to show that automatic assessment is applicable and useful in large scale teaching. During the past few years, we have gathered comprehensive data concerning first and second study year, covering both coursework and success in examinations. The main research methodology in present study is statistical analysis of data observed in the real world conditions, supplemented with interviews.

We also take somewhat controversial view that learning outcomes are accurately measured by the standard tests used for grading students. This view can be defended by the practical motivation of our study: the success of teaching is mainly measured by the same metrics. Obviously, this view has its limitations as essential qualitative changes may remain undetected. For example, some studies [9, 19] indicate that e-assessment may stipulate thinking skills and facilitate deep learning. Because such change is not necessarily revealed in usual university mathematics exams, questions of this type are beyond the scope of the present research. As a secondary topic, we study the students experiences with the system, and how they prefer to use it. Questions about the costs and human resource requirements are discussed, too.

4. The basic skills test

New engineering students have been tested for their basic skills in mathematics in autumns 2008 and 2009. The same test will be used also in autumn 2010. The main advantage of test is that it the same problems are used annually, enabling comparisons to the data from previous years. The test problems were originally created in Tampere University of Technology but the assessment system used there was different because of software licence issues [18]. At Aalto University, STACK was used for the test which consists of 16 randomised problems covering the most important topics in high school mathematics. Because the test was a part of a compulsory course for engineering students nearly all new students were tested. Testing took places in a computer classroom. There was an instructor present supervising the test and answering technical questions. The test results are summarized in Figure 1. The test scores are mainly used as a normalising factor in this study; for a more complete review of results see [12, 18].

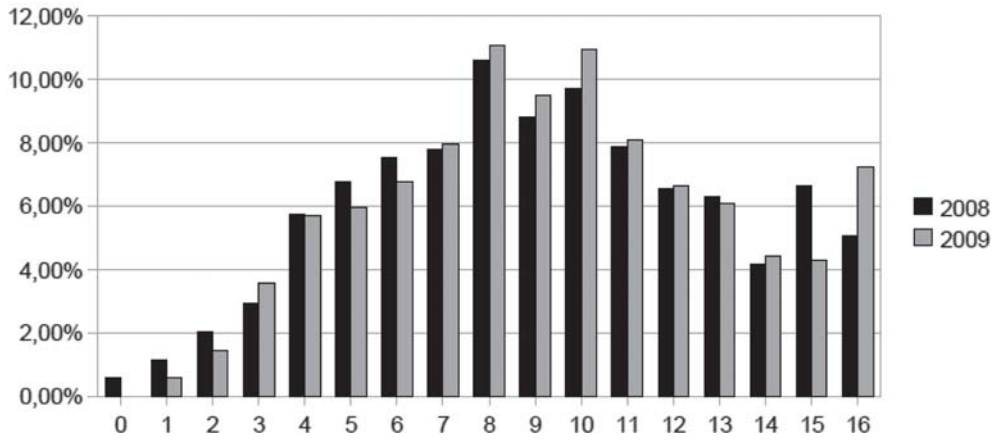


Figure 1. Distribution of the scores in the basic skill test of mathematics: years 2008 (N=889) (black) and 2009 (N=843) (gray). The length of the pillar describes the proportion of the total population with the score (0–16).

5. Experiences from the course S1

Basic course in mathematics S1 is the first of the three compulsory mathematics courses for electrical and telecommunications engineering students. It is intended to provide the basic skills needed in the degree program concerning the subject matter of the course. To contents of the course are complex numbers, matrix algebra, linear systems of equations, eigenvalues, differential and integral calculus for functions of one variable, introductory differential equations and Laplace transforms. Automatic assessment with STACK was first implemented on the course in 2007, and the same problems have been used on the course thereafter. The course also includes lectures and traditional exercise sessions supervised by an instructor. All lectures and exercises on the course are voluntary; students can choose only to participate on exams.

Table 1. Spearman’s rank correlation between the basic skills test, the exercise and the exams scores on the course S1 on years 2007–2009. P-values are less or equal to $p=0.0000$, except for basic skills test in 2009 where $p=0.0002$.

Year	Basic skills	Traditional	STACK
2007	n/a	0.49	0.57
2008	0.45	0.67	0.71
2009	0.35	0.69	0.66

Statistical analysis of the results from this course (see Table 1) shows that the amount students training with the system has a significant correlation to their scores from exams. Clearly, the number of problems a student tried to solve explains the success in examinations much better than the starting skills, supporting the popular belief that mathematics is mostly learned by practising with many problems. Web-based problems have a better correlation to success in exams than traditional ones in 2007. The reason for this is probably plagiarism, which is much harder with e-assessment if randomisation is used. Interestingly, the difference vanishes after 2007, pointing to a possible change in the study culture. The student activity as increased significantly, in particular among the best students (Table 2). It is more difficult to assess actual effects of STACK to the learning outcomes. We have observed certain improvement in students skills in examinations. The level of improvement seems to be most significant among the best students, and in routine test problems that can be solved algorithmically. However it is difficult to quantify the effects of this. Independent studies [15] have shown a significant increase in the proportion of new students in telecommunications engineering who pass a basic course in mathematics in their first study year since e-assessment was introduced. The student activity hours (for all mathematics courses using STACK, years 2009–2010) are illustrated in Figure 2. As it was found in [13] it seems that many students prefer to work outside the office hours, possibly because of schedule conflicts. Flexibility of schedules is a key advantage of e-learning over traditional classroom teaching.

Table 2. The percentage of automatically assessed (above) and traditional (below) exercise assignments solved by students. Numbers are sorted presented by the grade given (0–5), where 0 means failing the course. The general level of activity among the failing students is very low.

	0	1	2	3	4	5
2007	11.60% 3.78%	17.97% 7.77%	33.02% 20.19%	31.19% 9.40%	64.04% 26.84%	79.68% 61.61%
2008	13.20% 4.79%	23.62% 13.56%	36.55% 16.15%	49.56% 28.85 %	65.60% 56.81%	74.89% 58.44%
2009	14.62% 3.77%	23.28% 10.00%	38.78% 29.20%	49.53% 50.48%	51.16% 68.22%	78.32% 92.48%

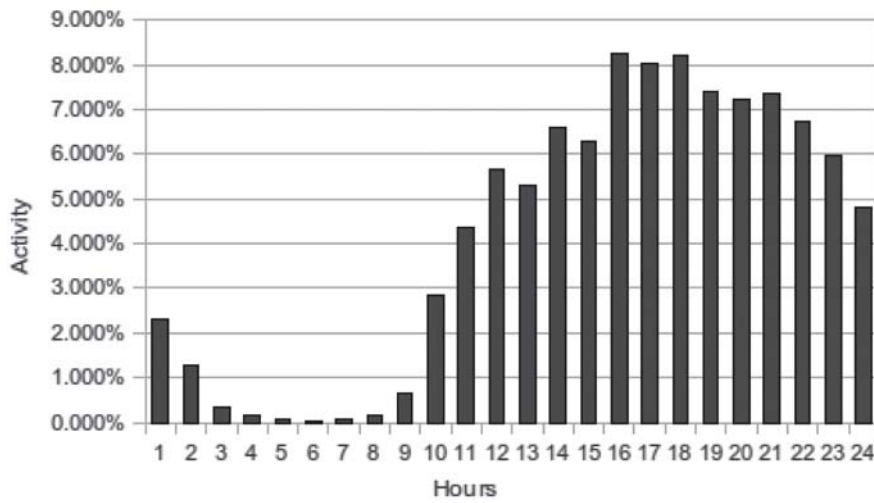


Figure 2. Student activity hours e-assessment system in Aalto University for nine mathematics courses using STACK. The relative frequency of submitted student solutions by hour. Total 93339 students submissions have been registered in 2009–2010.

6. Continuous evaluation with automatic assessment

Encouraged by our good experiences about e-assessment, an experimental course Discrete mathematics was set up at the spring semester 2010 (see also [3, 10]). The main idea was that the exercise assignment would form a significant portion of the final grade – a student could even pass the course without going to an exam. This approach follows the progressive, or continuous, assessment model [4, p. 192–193]. The model is certainly not new but it is difficult (or at least resource intensive) to implement effectively on a large course because of often resulting plagiarism. Again, the blended learning model was used: classroom lectures and face-to-face exercise sessions were held alongside the e-assessment, although use of STACK was extensive compared to the earlier experiments. The grading used on the course is illustrated in Figure 3.

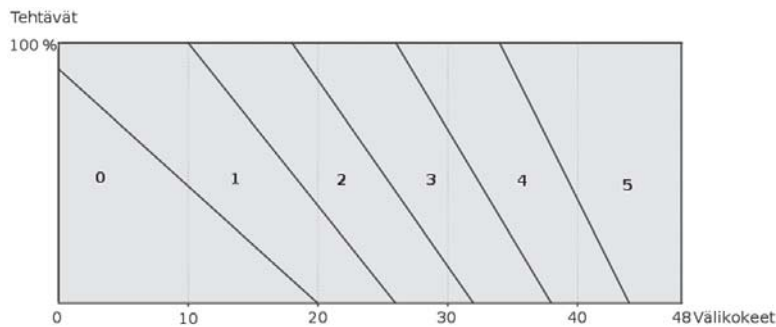


Figure 3. The grading system on the course Discrete mathematics: proportion of exercises solved is on the y-axis and exam score (0 – 48 points) is on the x-axis. The grades are 0 (fail) and 1–5, where 1 is the least passing grade and 5 is the best.

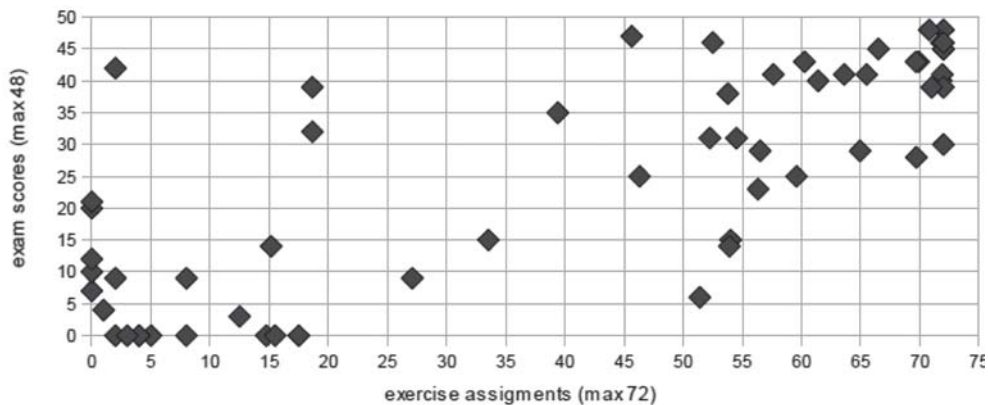


Figure 4. Student scores from exams and exercises by the time of the mid-term examination. About 29% of students have solved more than 90% of exercises.

Scores from exams and exercise assignments are illustrated in Figure 4. It is clear that the grading system for the course is highly motivating for students. Correlations of exercises and exam scores are given in Table 3. There are some examples of students who could solve a problem assignment when working with the e-assessment system, but could not solve a very similar one in the exam. This is particularly surprising because solutions to the problems cannot be easily copied when using e-assessment, and thus it is likely that the students solved their problem assignments by themselves. A likely explanation for such failure is stress in the examination situation, but this question requires further investigation.

Table 3. Spearman’s rank correlations between exercise activity and scores from the exam scores. The results are similar to those of the course S1.

Correlations	Traditional	STACK
Exam score	0.69	0.73

After the course, feedback was collected from students. Questions were asked by using a five point Likert scale, but there was also an option for free form feedback. Overall, the feedback from the course was overwhelmingly positive both regarding the course arrangements and the technology. For example, only one student agreed, and nobody strongly agreed with the statement “STACK system was difficult to use”. Based on the feedback, most of the students saw STACK as very useful for learning basic mathematical concepts and techniques, although many wished for even more comprehensive feedback concerning submitted solutions. On the other hand, students generally believed that learning advanced theoretical concepts and applications still requires face-to-face interaction with teacher. This is a key argument for using the blended learning model as in the pilot course. A more comprehensive analysis of the data is given in [10]. The grading system will be further piloted on other courses in the near future.

7. How much does it cost?

A question of practical importance is: how much does it cost, and is it worth the investment? According to our experience, creating a set of randomised, pedagogically meaningful problems for a full-semester 10 ECTS credit course required about three months of programming work. It should be noted that few people have both technical skill and teaching experience required for creating meaningful problem assignments. We have found that a system where the responsible teacher (lecturer) of the course works together with a programmer leads to a result which is good from both the pedagogical and technical point of view. STACK itself is free open source software, but running it requires a computer server. On the other hand, using STACK saves work after it has been properly set up, and thus fewer teaching assistants are required. By using this baseline analysis, we have found that the cost of creating a STACK exercises and introducing the system to a new course is paid back in four to five years.

8. Conclusions

E-assessment is a highly useful tool that can lead to increased flexibility in teaching. It also provides opportunities for improved feedback for students, diagnostic testing, data gathering and novel practices in practical arrangements of courses. Our experiences have shown that e-assessment is suitable for large scale teaching of engineering mathematics, it does not lead to overwhelming technical problems, and it can be highly motivating for the students. Besides these benefits, the system may lead to cost savings, at least in the long run.

References

- [1] K. M. Ala-Mutka: A Survey of Automated Assessment Approaches for Programming Assignments, *Computer Science Education*, Volume 15, Issue 2 (2005), 83–102.
- [2] L. Blåfield: *Matematiikan verkko-opetus osana perusopetuksen kehittämistä Teknillisessä korkeakoulussa*. Master's Thesis. University of Helsinki, 2009. (Finnish)
- [3] L. Blåfield, H. Majander, A. Rasila, P. Alestalo: Verkkotehtäviin pohjautuva arviointi matematiikan opetuksessa. *Tuovi 8 – Hypermedia Laboratory Net Series*. Tampere University, 2010. (Finnish)
- [4] J. Biggs: *Teaching for Quality Learning at University*. 2nd ed. The Society for Research into Higher Education & Open University Press, 2003.
- [5] J. W. Creswell: *Research design: qualitative, quantitative, and mixed methods approaches*. 3rd ed. Sage Publications, 2008.
- [6] R. Garrison and H. Kanuka: Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2) (2004): 95–105.
- [7] *GNU general public license*. Free Software Foundation, 2004.
- [8] M. Harjula: *Mathematics exercise system with automatic assessment*. Master's Thesis. Helsinki University of Technology, 2008.
- [9] M. Joy, N. Griffiths and R. Boyatt: The boss online submission and assessment system. *Journal on Educational Resources in Computing (JERIC)*, 5, Issue 3 (2005).
- [10] H. Majander: *Tietokoneavusteinen arviointi kurssilla Diskreetin matematiikan perusteet*. Master's Thesis. University of Helsinki, 2010. (Finnish)
- [11] M. Nieminen: *Finnish Air Force Cadets in network: experience in use of online learning environment in basic studies of Mathematics*. PhD Thesis. Faculty of Mathematics and Science, University of Jyväskylä, 2008. (Finnish, English summary)
- [12] S. Pohjolainen, H. Raassina, K. Silius, M. Huikkola, E. Turunen: *TTY:n insinöörimatematiikan opiskelijoiden asenteet, taidot ja opetuksen kehittäminen*. Tampere University of Technology. Department of Mathematics, Research Report 84, 2006. (Finnish)
- [13] A. Rasila, M. Harjula, K. Zenger: Automatic assessment of mathematics exercises: Experiences and future prospects. *ReflekTori 2007*, 70–80.
- [14] J. Ruokokoski: *Automatic Assessment in University-level Mathematics*. Master's Thesis. Helsinki University of Technology, 2009.
- [15] J. Ruutu: *Progressing and Promoting Freshman Studies in Communications Engineering – Integrating Students to The Scientific Community*. Master's Thesis. Aalto University, 2010. (Finnish, English summary)
- [16] C. Sangwin: Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and its Applications*, Vol. 23 No 1 (2003), 1–14.
- [17] C. Sangwin: STACK: Making many fine judgements rapidly. *CAME*, 2007.
- [18] K. Silius, T. Miilumäki, S. Pohjolainen, A. Rasila, P. Alestalo et al: Perusteet kuntoon – apuneuvoja matematiikan opiskelun aloittamiseen. *Tuovi 7 – Hypermedia Laboratory Net Series*. Tampere University, 2010. (Finnish)
- [19] J. Sitthiworachart, M. Joy, and E. Sutinen: Success Factors for e-Assessment in Computer Science Education. In C. Bonk et al. (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (2008), 2287–2293.