ADAPTIVE MULTIPATH EXERCISE IN THERMODYNAMICS

M. Harjula[a,1], V. Havu[b], J. Malinen[c], P. Salo[d]

[a] Department of Mathematics and Systems Analysis, Aalto University, Finland, 0000-0002-0622-5177
[b] Department of Applied Physics, Aalto University, Finland
[c] Department of Mathematics and Systems Analysis, Aalto University, Finland, 0000-0001-9117-8450
[d] Department of Applied Physics, Aalto University, Finland, 0000-0002-7749-9116

***Conference Key Areas**: Teaching mathematics and physics in engineering education, and Digital tools and AI in engineering education*
***Keywords**: Adaptive exercises, STACK, Stateful*

ABSTRACT

A teaching experiment is carried out in a university-level thermodynamics course using adaptive and interactive e-learning material, created in the new Moodle question type *Stateful* that extends the traditional STACK question type with inner-loop adaptivity. The system collects data that is used to algorithmically classify the students to different groups according to their behaviour in solving problems. It is observed that the classification of this data is consistent with a parametric model prediction based on the interaction data from *Stateful*. Moreover, for two principal student groups out of three, the parametric model explains the final course grade quite well. The proposed methods and the results obtained can be used in learning analytics and for allocating specific pedagogical actions to students in different groups.

[1] *Corresponding Author*
M. Harjula
matti.harjula@aalto.fi

DOI: 10.5281/zenodo.17631772

1 INTRODUCTION

In this article, a new e-learning concept for university-level elementary physics is discussed in terms of a pedagogical experiment that is carried out in an authentic setting. Earlier e-learning platforms have many benefits that include better reaching the students and giving them support without the need of teachers for scheduled learning; see, e.g., (VanLehn 2006). Automatic assessment systems such as STACK (System for Teaching and Assessment using a Computer algebra Kernel) have been used for a long time in conjunction with traditional teaching methods in a way that improves *active* and *cooperative learning* as well as *formative* and *elaborative feedback* (Hattie & Timperley, 2007; Shute, 2008; Attali, 2015). STACK is suitable for teaching general mathematical content with individually randomised exercises. It provides feedback for students and automatic assessment of students' answers for teachers; see, e.g., (Sangwin, 2013; Rasila, Malinen, & Tiitu, 2015; Sangwin & Harjula, 2017; Paiva, Ferreira, & Frade, 2017; Kulmala, Kokkonen, & Kontro, 2021; Erskine & Mestel, 2018; Nakamura, Fukazawa, & Takahara, 2013; Nakamura, Taniguchi, & Takahara, 2014). All students, however, see essentially the same problem and are expected to solve it following the same steps in STACK.

The adaptivity of electronic study material to different students' needs is a particularly important issue as well. The novel Moodle question type *Stateful* was originally drafted as a modification of STACK in (Harjula, Malinen, & Rasila, 2016), and it makes it possible to interactively link traditional STACK exercises as separate *scenes* into a network; see also (Altieri, Horst, Kallweit, Landenfeld, & Persike, 2020). The transition to the next scene is determined by rules, according to which students' answers in earlier scenes have been classified. At any moment, the student's past action in *Stateful* has affected the internal state of the exercise, thus leading to future adaptations. Hence, the *Stateful* question type makes it feasible to create dynamically adaptive e-learning materials that have an underlying story or a plot, emphasising strategic skills in problem solving while adapting to varying levels of students' skills. Thus, *Stateful* makes it possible to give students automated, personalised, and versatile feedback beyond the capabilities of the traditional STACK – resembling a form of an intelligent dialogue at its best. Exercises based on the *Stateful* question type are able to introduce serious gamelike features in a non-game setting (Deterding, Dixon, Khaled, & Nacke, 2011; Devlin, 2011). From the computer science point of view, *Stateful* exercises are state machines that accept external control in the form of input from the student.

We use a single, adaptive, multipath exercise on thermodynamics that has been realised using *Stateful*; i.e., the *Stateful* Diesel Exercise (henceforth, SDE) that describes thermodynamics of a four-stroke Diesel engine. Such an engine is particularly suitable for realisation in the *Stateful* question type using multiple scenes. Not only does the action of the engine consist of four strokes but also its thermodynamical state diagram comprises four stages where understanding of different thermodynamical processes is required. In essence, the Diesel cycle provides an underlying plot for SDE, and both formative and elaborative feedback is provided within the interactive narrative of SDE. It is worth observing that SDE implements the pedagogy of *mastery learning* in the sense of (Bloom, 1984).

It is characteristic for *Stateful* that very large amounts of student interaction data is produced which opens up new possibilities for *learning analytics* using statistical methods, parametric modelling, and perhaps even machine learning in future. Perhaps, not all students benefit much from adaptive, interactive materials such as SDE, and some of the students may not even be willing to interact with a computerised system. In order to understand students' reactions, we first classify the students according to the different ways how they respond to SDE. The classification is then related to students' learning outcomes by statistical tools and parametric modelling. More precisely, classification of students' behaviour is possible based on the paths they went through SDE, or whether they reached an end scene at all. Additional data for learning analytics can be obtained from the actual scene transitions, the changes in the state within scenes, and the steps on the path where the student gained points. We use *self-organising maps* (Kohonen, 2013) (SOM) and a novel parametric model called *fitted metric* (as explained in Section 3.2) as classifiers and a statistical predictor for various measures of learning outcomes.

2 COURSE ARRANGEMENT AND TEACHING EXPERIMENT

2.1 Student population, course and its grading

The e-learning experiment was arranged in Aalto University, Finland, in conjunction with an undergraduate thermodynamics course for first-year students of mechanical engineering. In the final grade of the course, the exam accounted for 30%, STACK 27%, *Stateful* 3%, the laboratory sessions for 20%, and the pre- and post-lecture work accounted for 10% each. In autumn semester 2020, 182 students enrolled to the course, 163 participated in the final exam, and 151 passed. Due to COVID-19 pandemic, the exam was held online in "open-book" format where the students had access to all the material. The results of the exam were quite similar with the results from previous years when traditional exams were arranged.

All course participants were used as test subjects, and they were informed about the study. Of the 182 participants, 140 (77%) were male and 42 (23%) female. The median of student age was 20.7 years. The student demographics were obtained from the university records, and no further demographic information was collected. Since there is a deviation from the principle of informed consent, an ethical review was carried out before the research (Aalto University Research Ethics Committee, D/652/03.04/2020).

For this work, the most relevant learning objective of the course are cyclic thermodynamic processes. For the analysis carried out, only the exercises and the exam are used as a measure of learning outcomes since they produce numerical data on student behaviour that is directly suitable for quantitative methods.

2.2 Diesel cycle as a *Stateful* exercise

The e-learning experiment involved a single *Stateful* exercise (SDE) about the Diesel cycle which serves as an exemplar of a non-trivial thermodynamic cycle for the purposes of the course. This cycle is used to model a four-stroke Diesel engine; for further details, see e.g. (Giancoli, 2008). The Diesel cycle was translated into the *Stateful* format consisting of several scenes as shown in Fig. 1. Most of the scenes are part of five *review loops* that consist of useful elaborative feedback (see (Attali & van der Kleij, 2017)) for those who are deemed to need it. The feedback is given as hints that help the student forward.

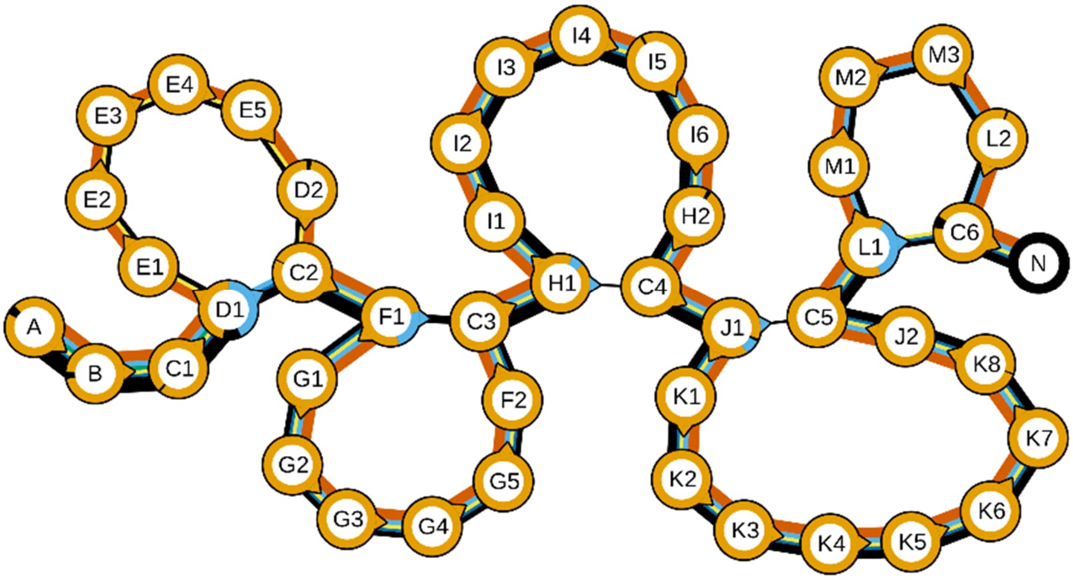

Fig. 1. Student paths through SDE consisting of the main path and five review loops, also indicating the general design (i.e., the topology) of the exercise.

In SDE, the entry scene A introduces the Diesel cycle but has no questions; see Fig. 1. Then all students are taken to a warm-up scene B where several Multiple Choice Questions (MCQ) about the relevant thermodynamical processes are posed. Next, the first main scene C1 is entered where the problem is set and numerical values are given. The first scene where students can end up in different paths is D1. This is a STACK-like question where the student is asked to solve for the maximum pressure of the Diesel cycle. If the right answer is given, the student is taken back to the modified main scene C2 from where scene F1 is reached. In case of a wrong answer in scene D1, the student is taken to a review loop E1–E5 consisting of several MCQs on how to calculate the maximum pressure. In Fig. 2 (left panel), the student view of the first scene E1 of the review loop is shown. After completing the review loop, the student is asked to solve again for the maximum pressure in a modified scene D2 with a STACK-like question. The difference between scenes D1 and D2 is that the student has all the results of the review loop E1–E5 at their disposal in D2; see Fig. 2 (right panel).

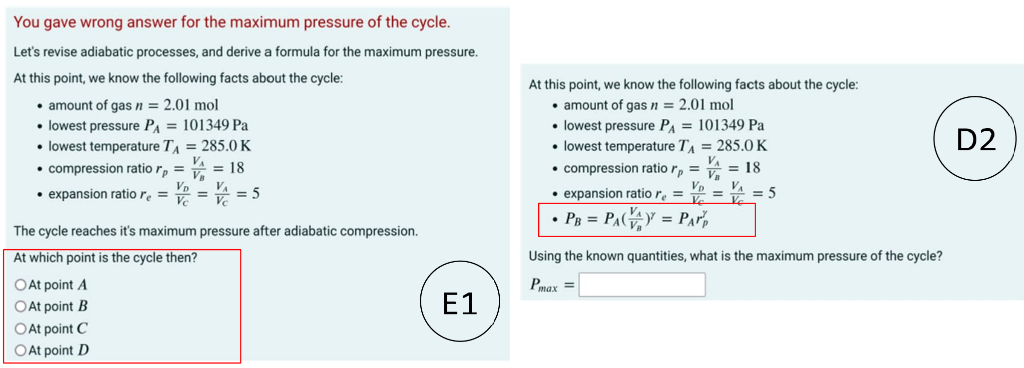

Fig. 2: Two student views from SDE where the nodes E1 and D2 appear in Fig. 1.

A similar process is repeated for the remaining four quantities of interest in the Diesel cycle, i.e., the maximum temperature of the cycle (F1/F2), heat input into the cycle (H1/H2), heat expelled from the cycle (J1/J2) and the thermodynamical efficiency of the overall cycle (L1/L2). All of these have dedicated review loops that trigger in case of a wrong answer. Finally, the student exits via the final scene (N).

2.3 Assessment of student activity in STACK and *Stateful*

Classical STACK questions can be assessed in terms of “raw points” (as defined by the question logic) and the number of attempts the student makes. In addition to such data, there is *path information* available in *Stateful*, describing the order the student has visited or skipped scenes of SDE shown in Fig. 1. The path information does not affect the assessment of the course, and it is not visible to the student. The coloured lines between the scenes in Fig. 1 reflect the path data, since they indicate the fractions of students continuing on specific branches. The path information be easily numerically quantified and combined with other types of performance data by statistical means and optimisation as was done in this work. The actual grading logic of SDE for course assessment was simplistic, and points were given from successfully completing specific key steps where all of points could be collected no matter how many attempts it took. The student is shown the cumulative score in real time.

The parametric model given in (Harjula et al., 2024) can be used to grade SDE either by fixing the model parameters manually or by estimating the model parameters with the objective to follow a chosen target variable in course assessment. In the latter case, the model is called *fitted metric*, and it can be interpreted as a statistical predictor of students’ learning outcomes. The fitted metric contains persistent information about students’ qualities that is often called a *student model,* e.g., in (VanLehn, 2006).

2.4 Software

The statistical computing was done using the R-system v.4.1 which was extended by kohonen-package v.3.0 for including the SOM algorithms that were used with default settings and a rectangular adjacency grid. *Stateful* v.1.0.2 and STACK v.4.3.5 were used on a Moodle environment based on v.3.8.2+. The optimal parameter estimation for fitted metric modelling was carried out using Octave v.6.3.0. The e-learning material, i.e., SDE, was created in the experimental development environment Eleaga Editor v.r1906.

3 RESULTS

3.1 Clustering of path data from *Stateful*

Apart from a minor complication due the review loop structures in Fig. 1, the general design of SDE is linear. This makes the vectoring of the student interaction data for statistical processing quite straightforward. Self-organising maps (SOM) (Kohonen, 2013) were chosen as the clustering method because it provides a geometric map of the data where similar data vectors tend to be adjacent, helping the interpretation of observations. The size of the adjacency grid (namely 3x2) was selected based on an iterated robustness experiment on randomly selected partial data. Similar key clusters were consistently detected, from which we identified three principal groups $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ of students that were consistently present in all clustering experiments, leading to following observations:

Group $\mathcal{A}$: Students who skipped the first and/or the second review loops, concerning maximum pressure and temperature in Diesel cycle, respectively. 48 students.

Group $\mathcal{B}$: Students who made large numbers of attempts, especially in the warmup MCQ part, and continued through all review loops. 39 students.

Group $\mathcal{C}$: Students who gave up without reaching the end scene. 23 students.

The student groups $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ amount to 76% of the test subject population, 110 students in total. The three remaining clusters in 3x2 adjacency grid configuration are of smaller size, consisting of 35 students that were not included into $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ by SOM. For example, there were 16 students who went through all the review loops in Fig. 1 except the final loop dealing with thermodynamical efficiency. These "ungrouped" students are too few so as to make statistical inferences in this experiment. We note that the general idea of clustering (or profiling) students is by no means a new approach: see (Kwarikunda et al., 2022) and the references therein.

3.2 Analysis of student data

The results of our analysis are shown in Fig. 3 where three pairs of statistical variables are plotted against each other. The membership in groups $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ is indicated by colours, and the centroids of the groups are shown by filled symbols.

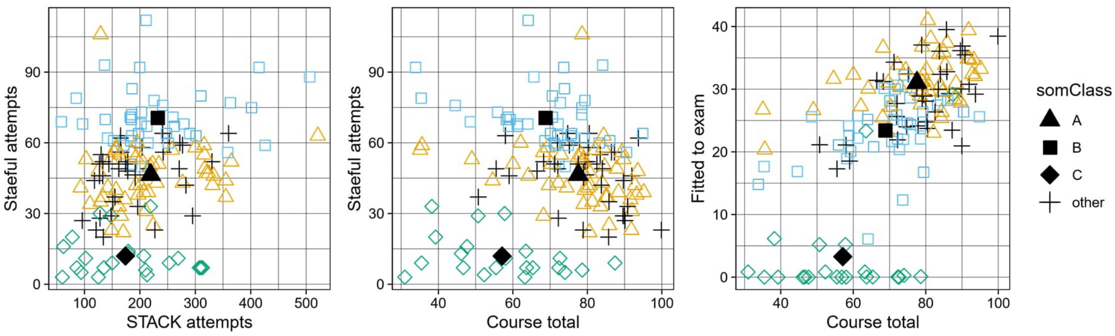

Fig. 3: Scattergrams of variable pairs where students' behaviour in Stateful has been described by groups $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. The group centroids have been drawn with filled symbols, and they would correspond to grades 4 for $\mathcal{A}$, 3 for $\mathcal{B}$, and 2 for $\mathcal{C}$ if the usual grading range 0…5 of the course were used.

We see from Fig. 3 (left panel), comparing the student activity between STACK and *Stateful*, that students in group $\mathcal{C}$ show much less interest than groups $\mathcal{A}$ and $\mathcal{B}$ in working with STACK or *Stateful*. Otherwise the activities correlate positively. The students in group $\mathcal{B}$ took full advantage of the feedback by going through all of the review loops in SDE, and their activity is thus highest.

We compare the students' total course points with the student activity in *Stateful* in Fig. 3 (middle panel). This indicates that group $\mathcal{A}$ has highest total points even though group $\mathcal{B}$ is exposed most to the formative feedback in the review loops of SDE. It is worth observing that the correlation between the total points and activity in SDE is negative for groups $\mathcal{A}$ and $\mathcal{B}$ (Pearson $r = -0.29$ with $p = 0.05$ for $\mathcal{A}$, and $r =$

−0.31 with p = 0.06 for $\mathcal{B}$). We hypothesise that the additional feedback provided by SDE in review loops is to a greater extent received by students who find the course challenging.

Compared to the clustering described in Section 3.1, an essentially different approach can be taken to process the path and score data from *Stateful*. The parametric model given in (Harjula et al., 2024) can be used to grade SDE with the objective to follow a chosen target variable in course assessment. Using nonlinear optimisation in Octave, the 72 model parameters in the fitted metric for SDE were optimised so as to best predict the somewhat modified final exam grading, see (Harjula et al., 2024). The result of this work is given in Fig. 3 (right panel) where the outcome of the fitted metric model has been plotted against the total course points. We observe first that these statistical variables have a strong correlation (Pearson r = 0.61 with p = 0.0001) over the whole student population. The correlations are significant for both the groups $\mathcal{A}$ and $\mathcal{B}$ as well (Pearson r = 0.43 with p = 0.002 for $\mathcal{A}$, and r = 0.43 with p = 0.006 for $\mathcal{B}$) but not for group $\mathcal{C}$. Secondly, it can be clearly seen that most of the students in group $\mathcal{C}$ did not finish SDE: hence, the fitted metric fails to predict the total grade for these students because the information in SDE interaction data is too sparse. Thirdly, SDE interaction data contains a fair amount of useful information about students in groups $\mathcal{A}$ and $\mathcal{B}$ in contrast to group $\mathcal{C}$. It can be observed that the fitted metric successfully separates groups $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in Fig. 3 (right panel). Moreover, the students not classified to any of the groups $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ appear uniformly among $\mathcal{A}$ and $\mathcal{B}$ but not at all among $\mathcal{C}$ in Fig. 3 (right panel) whereas the same students appear among $\mathcal{A}$ but not among $\mathcal{B}$ and $\mathcal{C}$ in Fig. 3 (left and middle panels).

4 CONCLUSIONS

An e-learning material for thermodynamics was introduced for first-year university students in mechanical engineering. The material describing the Diesel cycle (referred to as SDE) was developed using the novel *Stateful* question type. A teaching experiment was carried out using 182 of the students as test subjects. A large amount of interaction data was collected from the students in SDE and from their performance in the other parts of the course.

The collected e-learning data is rich enough for robustly classifying 76% of the students into three principal groups by using Self-Organising Maps (SOM) as a clustering algorithm as opposed to, e.g., (Kwarikunda et al., 2022) where a complicated statistical procedure was used. These groups have statistically significant differences both in other course metrics and also in observed behaviour in problem solving challenges provided by SDE as reported in (Harjula et al., 2024) in greater extent. A parametric model, i.e., the fitted metric, was applied the same learning data from SDE, and it was observed to give a consistent, refined picture with the clustering experiment by SOM.

Similar fitted metric analysis has been carried out by using other measures of learning instead of the (modified) grade of the final exam as the parameter optimisation target. Many reasonable targets seem to lead to essentially similar observations, supporting the hypothesis that the principal groups — produced by the entirely independent method SOM — reflect genuine and identifiable learning related differences in the student population. We conclude that the interaction data from even a relatively simple *Stateful* exercise material (such as SDE) may contain a plenty of useful information about the learning process on a physics course, providing a fruitful starting point for further learning analytics, production of improved e-learning materials, and targeting specific pedagogical measures to students with various learning abilities.

The greater power of expression of *Stateful* comes at a cost, and designing a fair grading logic for a *Stateful* exercise can be a challenge. It is possible that a student does not visit some scenes simply because the student is able to correctly solve the exercise in an elegant way without getting points from the unvisited scenes. Another student may visit some scenes many times, but it is perhaps not desirable to give full points for each repetition. *Stateful* provides sufficient tools for dealing with these situations in a fair manner.

5 ACKNOWLEDGEMENTS

We wish to thank Dr. P. Alestalo, Dr. I. Kontro, and Mr. K. Kulmala for valuable cooperation. This work was partially funded by the Ministry of Education and Culture (Finland) project “ÄlyOppi” (grant number OKM/238/523/2017) in 2018–2021.

REFERENCES

Altieri, M., Horst, J., Kallweit, M., Landenfeld, K., & Persike, M. (2020). Multi-step procedures in STACK tasks with adaptive flow control. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.3944786

Attali, Y. (2015). Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems. *Computers & Education*, *86*, 260–267.

Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education*, *110*, 154–169.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4–16.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining ”gamification”. In *Proceedings of the 15th international academic mindtrek conference: Envisioning future media environments* (pp. 9–15).

Devlin, K. (2011). *Mathematics education for a new era: Video games as a medium for learning*. Taylor & Francis.

Erskine, G., & Mestel, B. (2018). Developing STACK practice questions for the Mathematics Masters Programme at the OU. *MSOR Connections*, *17*(1).

Giancoli, D. C. (2008). *Physics for scientists and engineers with modern physics / Douglas C. Giancoli.* (4th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Harjula, M., Havu, V., Kontro, I., Kulmala, K., Malinen, J., & Salo, P. (2024). *Student behaviour and engagement with adaptive exercises on a thermodynamics course.* Retrieved from https://arxiv.org/abs/2403.04424

Harjula, M., Malinen, J., & Rasila, A. (2016). STACK with state. *MSOR Connections*, *15*(2), 60–69.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.
Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52-65. (Twenty-fifth Anniversay Commemorative Issue)
Kulmala, K., Kokkonen, T., & Kontro, I. (2021). 'Forced to think'—an analysis of how students benefit from online warm-up exercises with feedback. *Physics Education*, *56*(4), 045019.
Kwarikunda, D., Schiefele, U., Muwonge, C. M., & Ssenyonga, J. (2022). Profiles of learners based on their cognitive and metacognitive learning strategy use: occurrence and relations with gender, intrinsic motivation, and perceived autonomy support. *Humanities and Social Sciences Communications*, 9(1), 337–349.
Maxima. (2019). *Maxima, a computer algebra system. version 5.42.2.* https://maxima.sourceforge.io/.
Nakamura, Y., Fukazawa, K., & Takahara, T. (2013). Development of portable STACK: Packaged math e-learning system. In *Proceedings of world conference on e-learning.*
Nakamura, Y., Taniguchi, T., & Takahara, T. (2014, December). Item bank system for the mathematics e-learning system STACK. *Research Journal of Mathematics and Technology*, *3*(2), 77–85.
Paiva, R. C., Ferreira, M., & Frade, M. (2017). Intelligent tutorial system based on personalized system of instruction to teach or remind mathematical concepts. *Journal of Computer Assisted Learning*, *33*(4), 370–381.
Rasila, A., Malinen, J., & Tiitu, H. (2015, 08). On automatic assessment and conceptual understanding. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *34*(3), 149-159.
Sangwin, C. J. (2013). *Computer aided assessment of mathematics*. Oxford, UK: Oxford University Press.
Sangwin, C. J., & Harjula, M. (2017, Feb). Online assessment of dimensional numerical answers using STACK in science. *European Journal of Physics*, *38*(3), 035701.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.
VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.