Increasing Students' Awareness of Their Behavior in Online Learning Environments with Visualizations and Achievement Badges

Tapio Auvinen, Lasse Hakulinen, and Lauri Malmi

Abstract—In online learning environments where automatic assessment is used, students often resort to harmful study practices such as procrastination and trial-and-error. In this paper, we study two teaching interventions that were designed to address these issues in a university-level computer science course. In the first intervention, we used achievement badges, a form of gamification, to reward students for submitting early, solving exercises with few iterations, and completing the exercises with full points. In the second intervention, we used heatmap visualizations that show a prediction of the student's success if he or she continues to behave in the same way, based on data from earlier courses. The results from a controlled experiment show that the heatmap visualizations had an impact on how early the students submitted their exercises and how many points they earned from them. We also observed that *performance approach* goal orientation correlated with an interest towards the achievement badges whereas *performance avoidance* correlated with an interest towards the visualizations. The results indicate that increasing students' awareness of their behavior can have a positive impact on their study practices and results. However, the same type of feedback may not suit all students because of individual differences.

1 Introduction

Learning activities in all levels of education are increasingly based on blended or distance learning settings, instead of working in a classroom or laboratory. Moreover, students do not merely read text or follow video clips from the web, discuss the learning content with their peers in a discussion forum or prepare documents alone or jointly with peers. They are increasingly working interactively with tools that can provide them with some form of feedback on their actions. In a simple form, it could be automatic feedback on quizzes. More elaborate feedback includes, for example, analysis of correctness, style, structure or efficiency of programming submissions [1], [2], working with interactive visualization or simulation tasks, e.g. [3], [4], [5], [6], or getting personalized feedback on one's work, e.g. [7], [8].

A central aspect of this development is that students will increasingly work alone and face-to-face communication time with peers or teachers is reducing. An extreme version of this development takes place in Massive Open Online Courses (MOOCs) where the whole course operates in distance learning mode, and direct interaction with the teacher or tutor is missing totally or is very limited (although many MOOCs use peer communication to mitigate the missing teacher contact).

Such a change in learning arrangements strongly emphasizes students' *self-regulation skills*, i.e., students' motivation to study, and capability to apply their metacognitive skills to plan and monitor their working and progress [9]. The use of self-regulation strategies is known to correlate with academic performance, e.g. [10], [11], [12], [13]. Furthermore, research suggests that the relationship is causal, and interventions to foster self-regulation improve performance [14], [15], [16], [17]. Not all students are naturally good at regulating their learning. They overestimate their skills [18], use few metacognitive strategies unless explicitly scaffolded [13], [19], [20] and may not use metacognitive aids even if such are provided [21]. This suggests that lack of self-regulation is a source of underachievement and it is important to investigate ways to facilitate it in the context of online learning environments where students are required to independently take responsibility of their working.

Independent studying can be supported by automatic feedback, which can partially substitute the lack of human tutoring and feedback. In programming education, automatic assessment tools have already been used for decades [1], [2], [22]. An analysis of how students work with such tools has revealed several problems in the study process. Many students are inclined to procrastination, which has a negative effect on their performance [23], [24], [25], [26]. Moreover, inefficient learning strategies, like trial-and-error in solving tasks, when immediate feedback is available, are common [3], [27], [28].

This paper continues our work, where we focus on designing such forms of automated feedback that would have a positive effect on students' *studying practices* in online learning environments, and thereby improve learning results. In our previous work, we explored the use of gamification in an algorithm simulation environment to draw students' attention to beneficial study practices [29]. Our tool provided students with achievement badges as

• *The authors are with the Department of Computer Science, Aalto University, Espoo, Finland.*
E-mail: {tapio.auvinen, lasse.hakulinen, Lauri.Malmi}@aalto.fi.

Manuscript received 5 Nov. 2014; revised 11 May 2015; accepted 26 May 2015. Date of publication 3 June 2015; date of current version 14 Sept. 2015.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TLT.2015.2441718

1939-1382 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

rewards when they, for example, started working early or ensured the correctness of their solutions before submitting instead of working in a trial-and-error fashion. A controlled experiment revealed that badges had a positive, statistically significant effect on some aspects of students' behavior. Further analysis revealed that badges are most preferred by students with certain *achievement goal orientation* profiles, i.e., motivations towards certain types of goals and rewards [30].

In this paper, we explore if providing students with visualizations of their own behavior can improve their study practices by simply increasing students' awareness of their own behavior and thereby facilitating self-reflection. We provided students with heatmap visualizations that allowed them to compare their own behavior to the behavior of the students of the past courses. We tracked several dimensions of student behavior, such as how close to the deadline they completed the exercises, the number of resubmissions they used to solve exercises and how frequently they submitted revised solutions, and plotted the past students' average final exercise points as a function of their behavior. Students viewing the visualizations saw their current behavior plotted over the heatmap, offering a prediction of their final success if they keep on working in the course in the same way.

We evaluated the effect of the visualizations in a randomized experiment where half of the students saw the heatmaps and the other half acted as a control group. We investigated the change in students' behavior before and after the visualizations were introduced by analyzing the log data from the learning environment.

In addition, we compared students' responses to the visualization experiment and the previous experiment with achievement badges to study if different kinds of students responded to these two forms of feedback. More specifically, our research questions were:

1) Can students' behavior be influenced by visualizing their behavior, even though the visualizations are fully voluntary elements and not tied to grading?
2) What kinds of students, in terms of performance and achievement goal orientation, respond to the visualizations vs. gamified feedback?

The rest of this paper is organized as follows: In Section 2, we introduce related work regarding gamification, visualizations of students' behavior, and goal orientation theory. In Section 3, we describe the experiments and the visualization method. In Section 4, we present the results, and in Section 5 discuss our interpretation of the results. Finally, Section 6 concludes the paper.

2 Related Work

2.1 Gamification

Gamification can be defined as the use of game design elements in non-game contexts [31]. Gamification is often used in an attempt to increase user engagement in a system. The approach has been shown to have positive results, as Hamari et al. [32] report in their review about gamification. In educational contexts, however, different students may respond differently to the same gamification method [33], [34], [35]. Moreover, gamification might reduce internal motivation towards the activity by replacing internal motivation with external [36], especially if the external rewards are seen as controlling, as opposed to informational [37].

Achievement badges are one commonly used form of gamification. Hamari and Eranti define them as an optional challenge provided by a meta-game that is independent of a single game session and yields possible reward(s) [38]. Other definitions exist as well, but typically achievement badges provide additional challenges that are separate from the core functionality of the system. Furthermore, badges are a form of feedback about users' progress and behavior within a system. Moreover, often the badges have a social aspect e.g. in the form of public leaderboards.

There is evidence that achievement badges can have an effect on students' behavior. Denny [39] studied the effect of badges and found out that they had a positive effect on the quantity of student's contributions during the course. Domínguez et al. [33] reported that students using a gamified e-learning environment with achievement badges earned better scores in practical assignments while performing poorly on written assignments. Abramovich et al. [34] found evidence suggesting both positive and negative effects of using badges. They further state that the effects of the badges depended on the learners' abilities. It should be noted that the effect of using achievement badges can strongly depend on the users and the context [32]. Furthermore, even in the same context, subtle changes in the presentation of badges can have an impact on their effect [40].

2.2 Dashboards

A dashboard offers the user a visual overview of the state of a system. For example, the operator of a power plant can use a dashboard to confirm with a glimpse that the plant is operating normally. In education, dashboards can give the teacher an overview of the state of the class, or give the student an overview of his/her progress. In this section, we discuss three kinds of visual dashboards that have been developed for educational purposes: ones that give the instructor an overview of the course, ones that give the student an understanding of his/her knowledge level in different topics, and ones that help the student to reflect on his/her behavior (e.g. time usage).

Instructor's dashboards can, for example, visualize students' average performance in order to reveal topics that cause problems to many students, help to analyze which learning materials students prefer, or visualize the progress of individual students to identify those at risk of falling behind [41], [42], [43], [44], [45]. The tracked attributes typically include e.g. forum activities, how much time students use in different activities, at what time do they work, which learning materials or assignments students have accessed or completed, and their scores.

Students' dashboards usually visualize similar attributes as those aimed for teachers, but their aim is to help students to regulate their own learning. Visualizations can roughly be divided into two categories: those that visualize students' knowledge, and those that visualize behavior. Knowledge visualizations are usually used in adaptive learning environments which maintain a learner model representing the student's mastery of different topics. The model can be built

based on the correctness of quiz answers, access to course materials etc. The learner model can be visualized for the student as an *open learner model* to support reflection and help the student to decide which learning materials or activities should be accessed next [46].

Knowledge visualizations can simply be progress bars or bar charts illustrating the student's progress in different topics [47], [48], [49], [50], or more complex representations of e.g. Bayesian student models [51], trees that represent hierarchical knowledge [52], or concept maps that show how topics are connected to each other [53], [54], [55]. Students can also be allowed to view their own progress relative to peers to warn if they are falling behind (social comparison) or to help them discover what learning activities others are performing (social navigation) [56], [57], [58].

Student modeling makes it possible to automate the monitoring of students' progress. TADV [59] monitors student models to detect students who are falling behind or skipping topics, and automatically suggests feedback messages to be sent by the teacher. A controlled experiment revealed a slight learning gain in the treatment group. A system by Chen et al. [60] also monitors student models and sends SMS messages suggesting what to study next, remind about deadlines and suggest suitable peers as mentors. In a controlled experiment, the treatment group received higher scores in weekly tests and had a higher number of completed exercises. The learning gain was highest for the low-performing students.

Dashboards that aim to improve students' understanding of their behavior usually visualize the time spent in different activities, the number of accessed learning resources, forum activities, etc. Many visualize both the student's and the peers' activity (either average or each peer separately) to allow students to notice if their behavior differs from the norm.

COMTELLA [61] is a peer-to-peer system where students can share web resources such as files, articles or links. Various visualizations have been tried in the system to trigger social comparison and to encourage users to participate and contribute. In one experiment, users were visualized as spheres whose size represents the activity of each user. A significant increase in user contributions was observed after the visualization was introduced. In other experiments, users have been visualized e.g. as stars with varying size and brightness or avatars with different facial expressions [62].

Moodog [63] visualizes students' total material views, spent time and forum activities. It allows students to compare their activity to peers and notice if he/she is falling behind. Emails are automatically sent suggesting learning materials that student has not yet accessed. student activity meter (SAM) [64] visualizes students' time usage, accessed resources and forum activities. It allows students to compare where, when and how much they have spent time compared to an average student, and recommends materials that peers have used. CAMera [65] tracks and visualizes user's activities and time spent in application programs (e.g. word-processor or email) and websites. It also visualizes student's social network based on sent and received messages. It is aimed as a tool for reflection, allowing students to see on which activities or resources they have spent time and when.

StepUp! [66] tracks time usage in the Eclipse integrated development environment and websites. It visualizes reached goals, distribution of activity per day, time spent in different tools or activities, and allows students to compare their own activity to peers. A survey revealed that students considered the dashboard useful and thought that it helped them to reach goals. However, a few were disappointed that others could see their activity and suggested that tracking can cause stress and consequently decrease productivity. Most students perceived the feeling of being observed as a negative aspect. Furthermore, the authors noted that the users did not find the dashboard beneficial enough to use it regularly.

Course Signals [67] predicts students' success based on their performance in the course so far, amount of interaction with the learning environment, demographic characteristics and past academic history (GPA, test scores). Students receive periodical emails that visualize their risk level as traffic lights and suggest various help resources on campus. The authors compared the results from courses before and after introducing CourseSignals and observed an improvement in the grade distributions. Furthermore, students who took CourseSignals enabled courses had a lower drop out rate than others. Students reported that their motivation was positively affected by the system and very few considered the messages demotivating. However, the system was not evaluated in a controlled experiment, and thus there is a possibility of confounding variables that differentiate the courses using the system from the others.

Wäschle et al. [68] evaluated the effect of visualizing students' procrastination levels. In the first experiment, university students completed a weekly self-report scale on procrastination in an online learning environment, and a randomized treatment group was shown their weekly procrastination levels as line charts. Students with visualizations procrastinated significantly less. In the next experiment, the students were divided into three groups: no visualizations, visualizations and fake visualizations with random data. Again, the students with the actual visualizations procrastinated significantly less than the others. No effect was observed on performance in either study.

The systems presented in this section can be categorized into four categories:

1) Systems that visualize student's behavior and progress for the instructor who can then monitor and support students. This approach does not scale up to large courses.
2) Open learner models that visualize students' knowledge levels.
3) Systems that monitor students automatically and send notifications to at-risk students.
4) Dashboards that visualize students' behavior in order to support self-reflection. Verbert et al. [69] conclude in their survey of such dashboards that few of them have been evaluated for their impact on students' actual behavior or learning. Especially controlled experiments with randomized groups are rare.

The purpose of our study is to evaluate with a controlled experiment if the fourth approach, supporting self-regulated learning by visualizing students' own behavior, actually has an effect on their behavior.

TABLE 1
Example Items from The Goal Orientation Survey

Orientation	Example item
Mastery-intrinsic	To learn as much as possible is an important goal for me at school.
Mastery-extrinsic	It is important for me to get good grades.
Performance-approach	It is important to me that others see me as capable and skillful.
Performance-avoidance	I try to avoid situations in which I may appear incapable or stupid.
Avoidance	I aim to do only the mandatory schoolwork and no more.

The actual survey was in finnish.

2.3 Goal Orientation

Goal orientations define "why and how people are trying to achieve various objectives and refer to overarching purposes of achievement behavior" [70]. Rather than focusing on *what* people try to achieve, goal orientation characterizes individual's *motivation* to try to achieve various objectives. In this paper, we discuss goal-orientation in the context of education although it is applicable in other domains as well. Nicholls [71] and Dweck [72] originally identified *mastery* and *performance* goals. *Mastery* goals refer to students' strive to increase competence while *performance* goals refer to the strive to demonstrate competence relative to others. Nicholls [73] later added the *avoidance* goals to take into account that not all students are motivated but prefer to avoid challenging tasks and minimize effort. Performance orientations can be further divided into *approach* and *avoidance* components which refer to students' strive to demonstrate competence versus trying to avoid demonstrating incompetence relative to others [74], [75]. Niemivirta separates mastery orientation into a *mastery-intrinsic* orientation where a student studies for the sake of learning and a *mastery-extrinsic* orientation where a student evaluates mastery in terms of external criteria such as grades or feedback [76]. Goal orientations are not mutually exclusive but each individual exhibits a mixture of different goals with varying intensities [77].

Goal orientation has been shown to predict students' motivation and performance in educational settings [78]. It also plays a role in how students respond to achievement badges [30], [34]. In our earlier study [30], we studied how students with different goal orientations responded to badges in a blended university-level computer science (CS) course. In the beginning of the course, students responded to a goal orientation survey instrument by Niemivirta [76] which distinguishes five goal orientations. Some example statements from the survey are given in Table 1. Badges were introduced in the learning environment in the middle of the course and students gave feedback in the end. The results showed that those who found the badges most motivating had different goal orientation profiles than the other students. They, for example, had higher *performance approach* and *mastery extrinsic* orientations. We also observed that the students who were the most motivated by badges were high-performing even before the badges were introduced. In the present visualization study, we used the same goal orientation instrument which makes it possible to compare if different students, in terms of goal orientations, respond to badges or visualizations.

3 Methods

In this paper, we analyze data from the online learning environment of the Data Structures and Algorithms course at Aalto University in Spring 2013 and Spring 2014. In 2013, the learning environment had achievement badges for all students, whereas in 2014, we conducted a controlled experiment where half of the students saw visualizations of their own behavior. Both courses had a goal orientation survey in the beginning and a feedback survey in the end. Research question 1 about the effects of visualizations on student behavior is answered by analyzing data from the 2014 heatmap study. Research question 2, what kinds of students respond to visualizations versus badges, is answered by comparing students' responses in the two studies.

Section 3.1 gives an overview of the course and the online learning environment. Section 3.2 describes details of the badge study and Section 3.3 details of the heatmap study. Section 3.4 describes the analysis methods for comparing students' responses to the two interventions.

3.1 The Course and the Learning Environment

Data Structures and Algorithms is a bachelor's level computer science course that is mandatory for CS major students, as well as for minor students from some other study programs. The course has lectures and closed lab sessions on-campus, a team project work, and homework exercises completed in an online learning environment. Only students' behavior in the online learning environment is studied in this paper. The homework exercises consist of automatically assessed program and algorithm simulation exercises. The online learning environment is described in detail in [79] and the algorithm simulation exercises in [5].

In an algorithm simulation exercise, a student is shown the pseudo code of an algorithm, and the task is to manipulate the visualization of a data structure with the mouse in order to demonstrate what steps the algorithm would take. The student immediately receives feedback on which steps were correct. Exercises can be submitted after the deadlines as well, but the points of late submissions are reduced by 50 percent. The exercises are initialized with random input data each time, which makes it possible to allow students to attempt the exercises as many times as they want.

On the first visit to the learning environment, students were asked to answer a goal orientation survey by Niemivirta [76]. In the survey, each goal orientation scale consists of three items that students rated using a seven point Likert scale. A composite score for each orientation was calculated as the average of the three items. The survey also contained filler items that are not used in this study. Students were

TABLE 2
Completion Criteria of the Badges in the Three Categories and Difficulty Levels: Bronze (B), Silver (S), and Gold (G)

Category	Criteria	Value of X B	S	G	Example icon
Time management	50 % of points of the round earned at least X day(s) before the deadline	1	3	7	
Learning	Round completed with X % score	50	75	100	
Carefulness	X assignments with perfect score using only one attempt	5	10	15	

allowed to submit an empty form in order to avoid bogus answers. The goal orientations and examples of the items are shown in Table 1. The Cronbach alpha reliabilities were 0.89 for mastery-intrinsic orientation; 0.90 for mastery-extrinsic orientation; 0.74 for performance-approach orientation; 0.85 for performance-avoidance orientation; and 0.79 for avoidance orientation.

After the course, feedback was collected about the interventions in the course feedback questionnaire. The questionnaire contained numerical questions with a five-point Likert scale (0 = completely disagree, 4 = completely agree) as well as an opportunity to give open ended feedback.

3.2 Badge Study

In 2013, the course had 15 program simulation and 39 algorithm simulation exercises that were divided into eight rounds with deadlines roughly one week apart. The exercises were graded so that 50 percent of points were required for a passing grade 1, and 90 percent were required for the best grade 5. The course had slightly different arrangements for CS major and minor students but the simulation exercises were the same. The simulation exercises formed 20 percent of the final course grade for major students and 30 percent for minor students. The rest was determined by the exam (40 percent), a team project for minors (30 percent) and closed labs for majors (40 percent weight), which are not analyzed in this paper.

Badges were introduced in the beginning of round five. They were divided into three categories: *learning*, *time management*, and *carefulness*, and each badge type had three difficulty levels: *bronze*, *silver*, and *gold*. The completion criteria for the badges are described in Table 2. *Learning* and *time management* badges were given separately for each round, whereas *carefulness* badges included all the rounds from 5 to 8. Recently earned badges were visible in the sidebar of the exercise page, and all the available and earned badges were shown on a separate *badge summary page*. Visits to the badge summary page were logged. The technical implementation of the badge system is described by Haaranen et al. [80] in more detail.

3.3 Heatmap Study

In 2014, the course had 45 algorithm simulation and nine program simulation exercises divided into eight rounds with deadlines roughly one week apart. The algorithm simulation exercises made up 18 percent of the final grade. The rest was determined by the program simulation exercises (2 percent), closed labs (32 percent) and the exam (50 percent), which are not analyzed in this paper. The heatmap visualizations were only available for the algorithm simulation exercises because sufficient data of students' behavior was not available from previous years for program simulation exercises. Therefore, students' behavior in the program simulation exercises is excluded from the heatmap study. The grading scales were the same for major's and minor's students, and the grading scale for the simulation exercises was the same as in 2013 (50 percent required to pass).

When registering to the online learning environment, each student was randomly assigned to a treatment or a control group with a 50/50 probability. The heatmaps appeared on round two for students in the treatment group, while the first round acted as a baseline. Students were able to switch between eight different visualizations and the selections were logged.

To evaluate the effect of the visualizations on student behavior, we compared the change in their behavior between the baseline and the intervention rounds, between treatment and control. A control group is necessary because differences in the difficulty of the weekly rounds cause changes in students' behavior irrespective of the intervention.

Because previous work suggests that mostly high-performing students are the most interested in badges, we also compare high- and low-performing students separately. The final points from the baseline rounds cannot be used for separating high- and low-performing students because of a ceiling effect, i.e., many students receiving full points. Instead, students are divided into high- and low-performing ones based on their average points of the first attempt to each exercise. Bonferroni correction is used when comparing high and low-performing students because it is a post-hoc test.

3.3.1 Heatmap Visualizations

A heatmap is a visualization where the color of a point represents the output value as a function of the coordinates that represent inputs. We used heatmaps to visualize past students' *final exercise points* based on their behavior, for example, how early they started to work and how many attempts they made per exercise, on average. The heatmaps were generated using data from the same course over the previous four years (2010-2013), during which the exercises and their grading policy remained mostly the same. When a student views the visualization, his/her own behavior is plotted over the heatmap with a "You are here" label. Thus, the visualization provides a prediction of the student's final points given the behavior so far, alerting the student if he/she is studying in a way that has led to poor outcomes in the past. For example, a student submitting at the last minute and resorting to a trial-and-error mode with multiple repeated attempts would see the "You are here label move to a red area on the heatmap, whereas a student employing good study practices would stay on a green area. The position of the label is updated as the student submits more exercises.

Students' behavior is represented by five variables, described in Table 3. The variables were selected so that they address the same aspects as the achievement badges: avoiding procrastination and carefully checking the solutions before submitting. A two-dimensional heatmap can be drawn using two of the variables at a time. Students are

TABLE 3
Variables Used in the Visualizations

variable	description
Attempts	Average number of attempts per exercise
First submission	Average points of the first attempts to each exercise (relative to maximum points)
Interval	Average time interval between consecutive attempts to the same exercise (in minutes, 2 h cutoff)
Improvement	Average improvement between consecutive attempts to the same exercise (relative to maximum points)
Earliness	Average distance of the submissions to the deadline (in days)

able to select from six pre-determined pairs of variables. In addition there are two one-dimensional line charts visualizing only one variable at a time. Example visualizations are shown in Fig. 1.

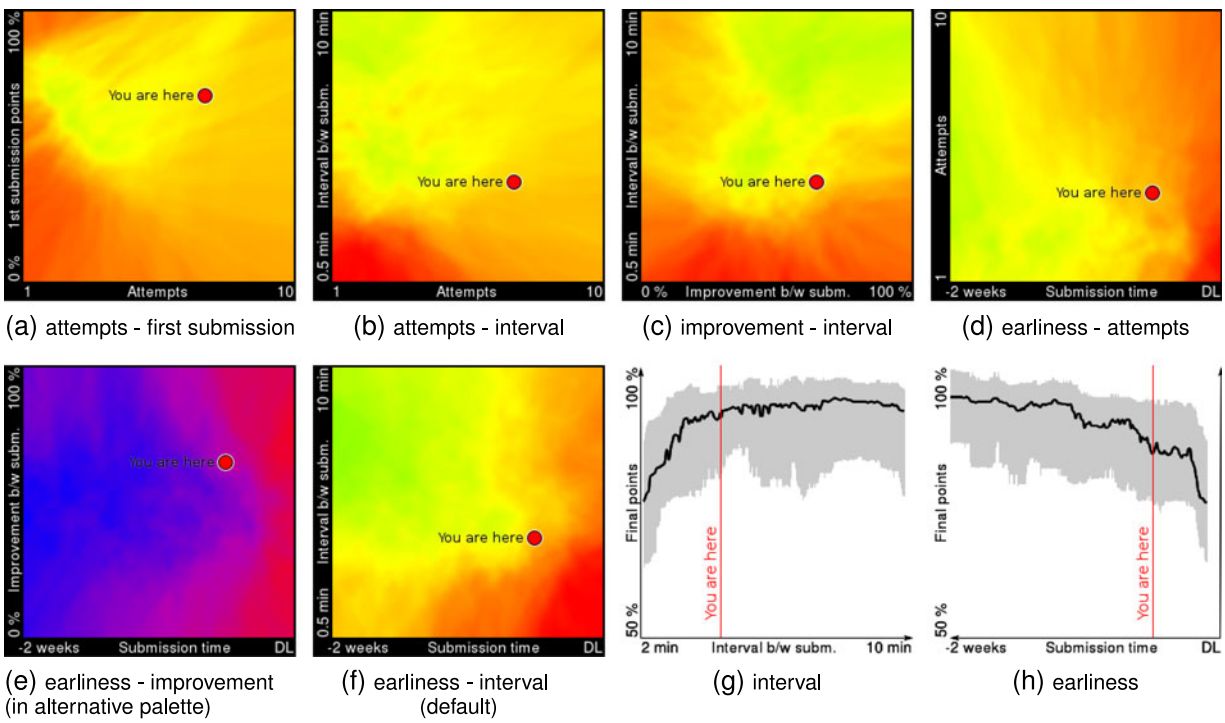
The student's main view of the learning environment is shown in Fig. 2. The visualization is placed on the right of the list of exercises. Below the visualization there are two dropdown menus for selecting from alternative visualizations and palettes. Below the dropdown menus are explanations of the axes and the palette, and then an explanation of the meaning of the heatmaps.

Heatmaps are generated with a nearest neighbor smoother [81] using each student of a past course as a data point. Students are mapped to a 2D plane based on the two selected variables that represent their behavior. Meaningful ranges must be predefined for the scales based on the course characteristics (e.g. few students need more than 10 attempts or submit more than two weeks before the deadline). Students are inserted into a spatial data structure that allows to efficiently find *m nearest neighbors*, which in our implementation is a k-d tree. The value of each pixel determined by the average of m nearest data points (i.e., most similar past students). In our implementation with 1,197 data points (i.e., students), $m = 100$ nearest neighbors are used for smoothing. A higher value of m would result in a smoother heatmap with less detail, whereas a smaller number would produce a more detailed map more susceptible to outliers. The output (final points) is mapped to a color ramp so that minimum points for a passing grade (50 percent points) are mapped to red, 75 percent points to yellow and 100 percent points to green. An alternative red/blue palette is available for the color blind. The one-dimensional linecharts are generated in an analogous way, so that the output is shown as the y-value instead of color. In addition, the lower and upper quartiles are shown in gray.

For example, the top-right corner of the earliness-attempts heatmap represents students who on average submitted right before the deadline using 10 attempts. The color of the corner is determined by the average final points of 100 past students who were closest to such behavior. Because their final points were poor, the area is colored red.

3.4 Comparing Students' Responses to Badges and Visualizations

Research question 2 asks if different students respond more strongly to badges or visualizations. We characterize



(a) attempts - first submission (b) attempts - interval (c) improvement - interval (d) earliness - attempts

(e) earliness - improvement (in alternative palette) (f) earliness - interval (default) (g) interval (h) earliness

Fig. 1. Heatmap and linechart visualizations available to students. In subfigures a-f, the color represents past students' final points as a function of the two variables representing behavior. In subfigures g-h, the thick line represents final points and the gray area lower and upper quartiles. The variables are explained in Table 3.

Fig. 2. Student's view of the learning environment with a heatmap visualization on the right. The red dot indicates the student's own behavior.

how much interest a student exhibits towards the interventions by measuring how many times they visited the badge summary page (badge views) or switched the heatmap visualizations (heatmap clicks). Another alternative would have been to use self-reported values from the feedback survey, but that would have reduced the sample size considerably. The badge views and heatmap clicks are normalized by dividing them with the number of attempted exercises to take into account that students with less overall activity in the system have less opportunities to interact with the interventions.

To study if high- or low-performing students responded more strongly to the interventions, we calculate the correlation between the students' baseline performance and badge views/heatmap clicks. Baseline performance is characterized by two alternative metrics: final points from the baseline rounds and points from the first attempts to each exercise on the baseline rounds. The first attempt points are included because the final points suffer from a ceiling effect. To study the role of goal orientations, we calculate the correlations between each goal orientation variable and badge views/heatmap clicks.

4 Results

In both data sets, students who have no submissions to the baseline or intervention rounds, or did not answer the goal orientation survey are excluded. Students who submitted the first rounds late, after the intervention was activated, are excluded as well because no baseline can be measured for them. Recap submissions (submissions after the deadline to exercises that have already been completed with higher points) are also ignored in the analysis.

The 2013 course had 339 students, out of which 55 were rejected because of missing goal orientation answers and 29 because of missing exercise activity. Totally, N = 254 students are included in the badge study.

The 2014 course had 263 students, out of which 12 were rejected because of missing goal orientation answers and 36 because of missing exercise activity. Totally, N = 215 students are included in the heatmap study, with 109 of them in the treatment group.

4.1 The Effect of Visualizations

Students were divided into high- and low-performing ones based on their average points of the first attempt to each exercise, using the median as the threshold (53 percent of maximum points).

Table 4 describes the change in students' behavior after the heatmaps were introduced. A statistically significant difference is observed in the change of *earliness* of the high-performing students.

The *earliness* values (average distance of submissions to deadline) are visualized in Fig. 3. The figure shows the earliness of all, low- and high-performing students round-by-round. In high-performing students, the treatment group starts at the same value on the baseline round, and gradually starts to submit earlier compared to control. In low-performing students, there seems to be little difference between treatment and control.

TABLE 4
Change in Students' Behavior from Baseline to Intervention (Student's T-Test, Two-Tailed), and Difference in Final Points (Wilcoxon Rank-Sum Test, Two Tailed, Because of Non-Normal Distribution)

variable	perf	median CONTROL	median TREATMENT	mean CONTROL	mean TREATMENT	SD CONTROL	SD TREATMENT	df	t / W	p
earliness-diff	all	−0.3474	−0.2568	−0.2040	0.5333	5.4040	7.3681	192.4	−0.8348	0.4049
earliness-diff	low	0.3481	−0.8853	0.1976	−0.9209	5.7356	8.9907	79.2	0.7496	0.9114
earliness-diff	high	−0.6891	0.6009	−0.6443	1.7835	5.0340	5.3913	106.9	−2.4310	**0.0334** *
attempts-diff	all	0.2500	0.3974	0.1197	0.2830	1.0686	1.4250	194.6	−0.9487	0.3440
attempts-diff	low	−0.0238	0.1042	−0.1609	−0.1620	1.2798	1.8387	83.8	0.0034	1.0000
attempts-diff	high	0.3494	0.5000	0.4273	0.6655	0.6597	0.7663	106.7	−1.7435	0.1682
relative-points-diff	all	−0.0562	−0.0089	−0.1086	−0.0703	0.1676	0.1566	212.7	−1.7318	0.0848
relative-points-diff	low	−0.0745	−0.0354	−0.1320	−0.0956	0.1727	0.1884	98.4	−1.0312	0.6100
relative-points-diff	high	−0.0351	0.0000	−0.0829	−0.0485	0.1595	0.1206	94.6	−1.2602	0.4214
final-points	all	1821.0000	1860.5000	1665.0560	1761.8080	350.7423	297.0740		W = 4778.5	**0.0272** *
final-points	low	1713.0000	1794.0000	1615.8448	1695.3792	372.6924	336.7861		W = 1226.0	0.5611
final-points	high	1863.7778	1938.0000	1718.9990	1818.9135	319.8789	247.1286		W = 1181.0	0.1230

Bonferroni Correction is used in the Post-hoc Tests (Low- and High-Performing Students). Relative Points *Refers to Earned Points Divided by Available Points. SD = standard deviation, df = degrees of freedom. N = 215.*

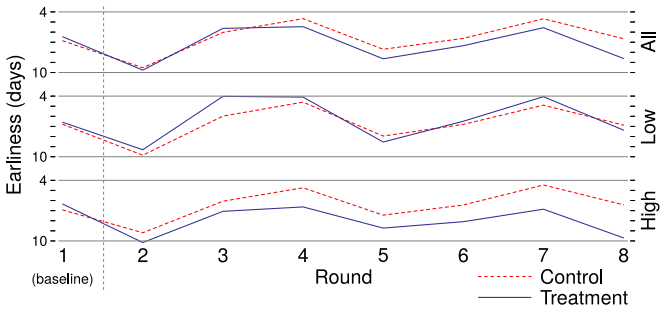


Fig. 3. Earliness of all, low- and high-performing students round-by-round in the heatmap study.

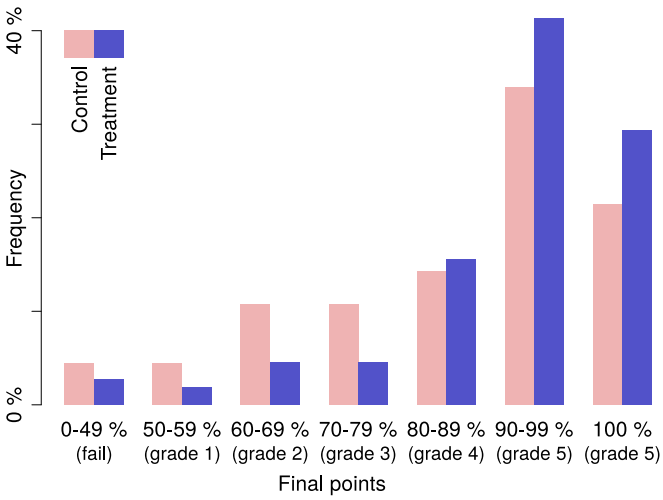


Fig. 4. Histogram of final points in the heatmap study.

TABLE 5
The Mean Number of Times each Chart was Displayed to a Student

chart	chart displays
earliness - interval (default)	39.52
earliness - interval (selected)	22.32
earliness	22.21
attempts - first submission	12.62
earliness - attempts	6.99
attempts - interval	5.26
earliness - improvement	4.94
interval	4.73
improvement - interval	3.50

4.1.1 Effect on Performance

No statistically significant change is observed in the earned exercise points (*relative-points-diff* in Table 4). However, there was a strong ceiling effect with 65 percent of students reaching full points on the baseline round. In fact, the high-performing students who earned maximum points from the first round can only have a zero or negative change on the later rounds. Therefore, we compared the absolute final points between treatment and control as well (*final-points* in Table 4). The treatment group earned 6 percent more points during the course ($p = 0.0272$). Fig. 4 shows the point distributions of treatment and control. In treatment, a bigger proportion of students earned grades 4 and 5 or reached maximum points.

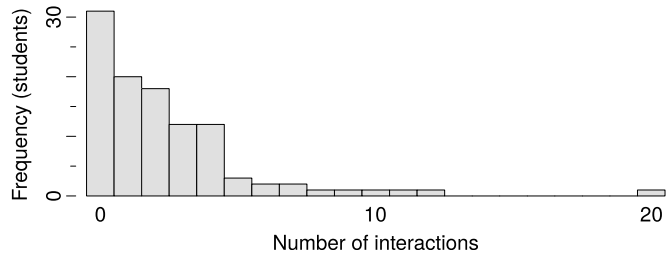


Fig. 5. Histogram of the number of separate times students interacted with the visualizations.

TABLE 6
Spearman's Correlations between Students' Performance and Normalized Badge Views / Heatmap Clicks

intervention	N	performance	r^2	p
badges	254	total points	0.046	0.0035 *
		first attempts	0.071	0.0003 *
heatmaps	109	total points	0.001	0.7823
		first attempts	0.060	0.0345 *

Total Points = *Earned total exercise points before the intervention.* First Attempts = *average points from the first attempts to each exercise before the intervention.*

4.1.2 Interaction with the Visualizations

Table 5 shows how many times each chart was shown to a student, on average. This includes times when a student explicitly selected the chart and when the previously selected or the default chart was shown. This way, we take into account the case that a student selects the most interesting chart and leaves it on for the remainder of the course. The table shows separately the displays of the default chart in case a student has never explicitly selected a chart, and the cases where a student has explicitly selected that particular chart. The most popular charts were the *earliness-interval* chart which was the default, and the one-dimensional *earliness* chart, which together account for 84 percent of the chart displays.

Fig. 5 shows a histogram of how many separate times students interacted with the visualizations. In this histogram, consecutive clicks during the same page load are counted as one interaction. Thirty students (28.4 percent of treatment) did not interact with the charts at all, and the median number of interactions was 2. The mean number of different charts viewed per interaction was 5.7.

4.2 What Kinds of Students Respond to Visualizations or Badges?

To study if high- or low-performing students respond to the interventions more strongly, we calculate Spearman's correlations between their baseline performance and badge views/heatmap clicks (Table 6). Both interventions have a weak but statistically significant correlation when using the first attempt points as the performance metric. Students' baseline performance and interactions with the interventions are visualized in Fig. 6. Students with the most interactions with the interventions were mostly high-performing already before the interventions.

In order to study if goal orientations explain students' responses to the interventions, we calculated the correlations between the goal orientation variables and badge views/heatmap clicks. The correlations are reported in

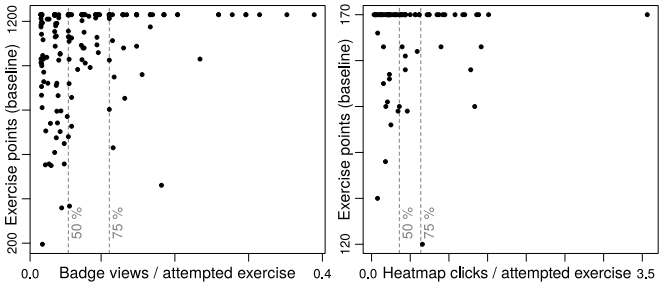


Fig. 6. Students' interest towards the interventions versus their exercise points before the interventions. The medians and the top quartiles are shown with dashed lines.

Table 7. With badges, there are statistically significant correlations to *mastery extrinsic* and *performance approach*. With heatmaps, on the other hand, there is statistically significant correlation to *performance avoidance*. With badges, *performance approach* explains about 7 percent and *mastery extrinsic* about 3 percent of the variance. With heatmaps, *performance avoidance* explains about 6 percent of the variance.

4.3 Feedback

Feedback about the interventions were collected in the end of the courses in the course feedback surveys. In the badge study, 138 out of the 254 students gave feedback (54.3 percent). In the heatmap study, 52 out of the 109 students in the treatment group gave feedback (47.7 percent).

Numerical feedback answers are reported in Tables 8 and 9. The majority of students report that neither badges nor heatmaps affected their behavior. 26.1 percent agree that badges affected their behavior, and 9.6 percent agree that heatmaps did. Few students report that badges or heatmaps disturbed them.

Students were also asked how often they studied the visualizations during the course. The results are shown in Table 10. Only 11.5 percent of treatment group reported studying the visualizations at least during every round, which is in line with the recorded activity (Fig. 5). The majority of the students reported studying the visualizations only a few times during the course.

TABLE 7
Spearman's Rank Correlations between Goal Orientations and Normalized Badge Views/Heatmap Clicks

orientation	study	r^2	p
mastery int.	badge	0.002	0.5265
	heatmap	0.000	0.8865
mastery ext.	badge	**0.028**	**0.0229** *
	heatmap	0.036	0.1016
perf. appr.	badge	**0.068**	< **0.001** *
	heatmap	0.022	0.2066
perf. avd.	badge	0.003	0.4340
	heatmap	**0.062**	**0.0308** *
avoidance	badge	0.006	0.3051
	heatmap	0.010	0.3885

N = 254 (badges), N = 109 (heatmaps).

5 Discussion

Students' goal orientations were correlated to how much they interacted with the interventions, which we consider an indicator of how strongly they respond to this kind of feedback. It is not surprising that students' interest towards badges was correlated with *mastery extrinsic* orientation because badges are extrinsic indicators of performance. However, it is interesting that badges resonated with the *performance approach* orientation while heatmaps resonated with *performance avoidance*. We interpret that this is in line with goal orientation theory. Achievement badges reward for completing additional challenges and provide a way to demonstrate high performance. They do not, however, punish for failure because it makes no difference if the student barely misses the badge criterion or fails completely. Heatmaps, on the other hand, do not reward for high perfor-

TABLE 8
Numerical Feedback to the Badge Study

feedback item	0	1	2	3	4
Trying to earn badges had an effect on my behavior	**41.3** %	21.7 %	10.9 %	15.2 %	10.9 %
I found the badges motivating	19.6 %	**25.4** %	17.4 %	**25.4** %	12.3 %
Badges disturbed my work	**60.9** %	14.5 %	15.2 %	6.5 %	2.9 %
I was satisfied with the criteria for awarding badges	3.6 %	8.7 %	**42.8** %	28.3 %	16.7 %
Visual look of the badges was good	2.9 %	6.5 %	21.7 %	**42.0** %	26.8 %
I think that badges should be used in the next year's course as well	8.0 %	12.3 %	29.0 %	21.0 %	**29.7** %

(0 = completely disagree, 1 = somewhat disagree, 2 = cannot say, 3 = somewhat agree, 4 = completely agree). N = 138.

TABLE 9
Numerical Feedback to the Heatmap Study

feedback item.	0	1	2	3	4
The charts affected how I did exercises and submitted them	**53.8** %	21.2 %	15.4 %	7.7 %	1.9 %
The charts gave me useful information of my own activities	26.9 %	17.3 %	**28.8** %	23.1 %	3.8 %
The charts disturbed my work	**59.6** %	13.5 %	21.2 %	3.8 %	0.0 %
The charts were easy for me to interpret	13.5 %	21.2 %	9.6 %	**32.7** %	23.1 %
In think that the charts should be used in the next year's course as well	7.7 %	5.8 %	**40.4** %	30.8 %	15.4 %

(0 = completely disagree, 1 = somewhat disagree, 2 = cannot say, 3 = somewhat agree, 4 = completely agree). N = 52.

TABLE 10
Feedback about Students' Activity with the Heatmaps

How often did you study the charts?	students
At almost every exercise	3.8 %
At almost every round	7.7 %
A few times during the whole course	40.4 %
One or two times during the whole course	38.5 %
Never	7.7 %

$N = 52$.

mance but rather warn students for behavior that typically predicts poor outcomes. Therefore, it is natural that students who worry about being inferior to others express more interest towards the heatmaps.

The explicit use of competition as a motivational tool is sometimes seen as a characteristic feature of gamification [82]. However, our achievement badge implementation did not have leaderboards or any other means of comparing one's achievements with others. Badges were merely shown to the students privately as a form of feedback and an incentive to employ good study practices. Nevertheless, badges resonated with performance approach that is associated with the desire to demonstrate competence compared to others. We hypothesize that this might be because the badge system is implicitly seen as competitive even without explicit competitive elements. More precisely, getting a badge might make the student feel that he or she performed better than others who did not get a badge. Furthermore, students may have compared their badges face-to-face.

The students who were the most interested in badges or heatmaps were mostly high-performing already during the baseline rounds (Fig. 6). This can be expected with badges because the additional challenge provided by them is only meaningful for high-performing students. Low-performing students, in contrast, may be struggling to complete the exercises in the first place and do not seek for additional challenges.

The effect of the heatmap visualizations was strongest on the already high-performing students (Table 4). It is known that self-regulating activities are correlated with performance, i.e., the best students are typically self-regulating and vice versa [83]. We believe that because the low-performing students were less interested in monitoring and planning their learning, they paid less attention to the visualizations, and thus the visualizations failed to affect their behavior. We hypothesize that the heatmap visualizations can support students who are already self-regulating but they do not create a desire to self-regulate if it does not already exist. Chen et al. [60] noticed that sending email reminders improved the results of low-performing students. It could be that external regulation works better for the less self-regulating students. Arnold and Pistilli [67] also report that CourseSignals, which implements external regulation in the form of email remainders, improved students' grades.

The treatment group earned more exercise points during the heatmap intervention. The heatmaps, which are calculated from the previous courses' data, show that students who submit earlier reach higher points. The intervention caused some students to start submitting earlier than control, and this change may in fact have improved their results. On the other hand, it could be that the visualizations simply made the students pay more attention to their activity, including the number of earned points, which made them try harder. Similarly, seeing their own position plotted on the chart may have made the students feel that their activity is monitored which made them try harder.

The majority of the students interacted with the heatmap visualizations only a few times during the course (Fig. 5 and Table 10). Furthermore, the majority of the displayed visualizations were the one or two dimensional *earliness* charts. Moreover, about a third of students considered the charts difficult to interpret. Thus, a fewer number of simpler visualizations might be more efficient than the current heatmaps.

In both studies, the majority of students disagreed to the intervention having an effect on their behavior. This suggests that the effect seen in the data comes from a small group of students. However, this is based on a self-reported value and both badges and heatmaps may potentially affect students' behavior unconsciously. Further studies are needed to determine the effects on individual students rather than the average difference between treatment and control.

It seems that heatmap visualizations can be used to steer students' behavior for the better, and we have found similar effects with badges in a previous study [29]. However, with badges, the effect may come from the desire to achieve rewards, while with the heatmaps, it may come from the fear of failure. Thus, when designing automated feedback for online learning environments, we must take into account what types of people respond to it and why. Even though we can affect students' behavior with the visualizations, we do not know how it affects their spirit and well-being. The *performance avoidance* orientation has been linked to negative attributes such as exhaustion at school and sense of inadequacy as a student [84], and avoidance goals in general correlate with lower self-esteem and lower subjective well-being [85]. Thus, while some type of feedback may improve results it may have harmful side-effects that are not observable in the log data.

5.1 Limitations

Eleven percent of students did not answer the goal orientation survey which may cause students with certain goal orientation profiles to be under- or overrepresented in the data. For example, *avoidance* oriented students might be more likely to skip the survey. Similarly, certain types of students may be more likely to give feedback which could cause a bias in the results.

Both the badge and heatmap interventions tried to affect multiple aspects of students' behavior (earliness, number of attempts, number of completed exercises). However, some of the goals may be conflicting and different students may be optimizing different aspects of their behavior at the expense of others. This may reduce the strength of the observed effects. Furthermore, the control group may have been aware of the intervention which may have affected their behavior as well and reduced the difference between treatment and control.

A direct comparison between the two studies is challenging because there are slight differences in the course arrangements. However, we do not expect slight differences in the exercises to significantly affect students' responses to badges or heatmaps because they are not tied to the content of the exercises. These differences might have a significant influence on students' behavior, but in this paper, we do not compare their behavior between the two studies.

Finally, the distribution of exercise points is very skewed with more than half of the students reaching the highest grade. This makes it more difficult to evaluate the effect of the heatmaps on students' performance because many of them had little room to improve. We also had to divide students to high- and low-performing ones by their first attempt points. Therefore, by *high performing students* we refer to those who earn high points with few attempts, whereas *low-performing students* may include those who earn high points but require many attempts.

6 Conclusions

In this paper, we analyzed the effect of showing students visualizations of their own behavior in an online learning environment in order to help them pay attention to beneficial study practices such as starting early or avoiding trial-and-error behavior. We also compared what kinds of students, in terms of achievement goal orientation and performance, responded to the visualizations versus an earlier experiment with achievement badges that were used for the same goal.

The results show that students' behavior was affected simply by visualizing their behavior, even without giving tangible rewards such as exercise points. The heatmap visualizations had an impact on how early the high-performing students submitted their assignments. Furthermore, students who saw the visualizations earned higher total points from the exercises.

There were differences in the goal orientations of students who exhibited interest towards badges or heatmaps. *Performance approach* oriented students were more interested in badges whereas *performance avoidance* oriented students were more interested in the heatmaps. Therefore, we believe that the two forms of feedback affect different kinds of students. In both studies, the students who exhibited the most interest towards the feedback were already high-performing before the interventions. However, not all high-performing students were interested in the badges or heatmaps. More research is needed to find ways to support low-performing students as well. Furthermore, a within-subject experiment with equally difficult baseline and treatment rounds would allow us to study the effect of the interventions on individual students.

References

[1] K. M. Ala-Mutka, "A survey of automated assessment approaches for programming assignments," *Comput. Sci. Educ.*, vol. 15, no. 2, pp. 83–102, 2005.

[2] P. Ihantola, T. Ahoniemi, V. Karavirta, and O. Seppälä, "Review of recent systems for automatic assessment of programming assignments," in *Proc. 10th Koli Calling Int. Conf. Comput. Educ. Res.*, New York, NY, USA, 2010, pp. 86–93.

[3] J. Helminen, P. Ihantola, V. Karavirta, and L. Malmi, "How do students solve parsons programming problems?: An analysis of interaction traces," in *Proc. 9th Annu. Int. Conf. Int. Comput. Educ. Res.*, 2012, pp. 119–126.

[4] T. Kangasniemi, S. Olkinuora, P. Joensuu, O. Natri, P. Qvist, M. Ketola, J. Brusin, H. Virtanen, M. Närhi, R. Jokela, E. Palomäki, H. Tiitu, and K. Nordström, "Designing virtual laboratories: Decarboxylation reactions, vacuum distillation and virus identification by pcr in the lablife3d second life laboratory," in *Proc. 5th Int. Conf. Comput. Supported Educ.*, 2013, pp. 661–672.

[5] L. Malmi, V. Karavirta, A. Korhonen, J. Nikander, O. Seppälä, and P. Silvasti, "Visual algorithm simulation exercise system with automatic assessment: TRAKLA2," *Informatics Educ.*, vol. 3, no. 2, pp. 267–288, 2004.

[6] J. Sorva, J. Lönnberg, and L. Malmi, "Students' ways of experiencing visual program simulation," *Comput. Sci. Educ.*, vol. 23, no. 3, pp. 207–238, 2013.

[7] P. Brusilovsky and S. Sosnovsky, "Engaging students to work with self-assessment questions: A study of two approaches," *ACM SIGCSE Bull.*, vol. 37, no. 3, pp. 251–255, 2005.

[8] P. Brusilovsky, S. Sosnovsky, M. V. Yudelson, D. H. Lee, V. Zadorozhny, and X. Zhou, "Learning SQL programming with interactive tools: From integration to personalization," *ACM Trans. Comput. Edu.*, vol. 9, no. 4, p. 19, 2010.

[9] R. Azevedo and J. G. Cromley, "Does training on self-regulated learning facilitate students' learning with hypermedia?" *J. Edu. Psychol.*, vol. 96, no. 3, p. 523, 2004.

[10] B. J. Zimmerman and M. M. Pons. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *Am. Edu. Res. J.* [Online]. *23 (4)*, pp. 614–628. Available: http://aer.sagepub.com/content/23/4/614.abstract

[11] S. Bergin, R. Reilly, and D. Traynor, "Examining the role of self-regulated learning on introductory programming performance," in *Proc. 1st Int. Workshop Comput. Edu. Res.*, New York, NY, USA, 2005, pp. 81–86.

[12] L. Barnard-Brak, V. O. Paton, and W. Y. Lan, "Profiles in self-regulated learning in the online learning environment," *The Int. Rev. Res. Open Distrib. Learn.*, vol. 11, no. 1, pp. 61–80, 2010.

[13] H. K. Ning and K. Downing, "A latent profile analysis of university students' self-regulated learning strategies," *Studies Higher Edu.*, pp. 1–19, 2014, to be published.

[14] J. Hattie, J. Biggs, and N. Purdie. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Rev. Edu. Res.* [Online]. *66(2)*, pp. 99–136. Available: http://rer.sagepub.com/content/66/2/99.abstract

[15] B. W. Tuckman, "The effect of learning and motivation strategies training on college studentsi achievement," *J. College Student Develop.*, vol. 44, no. 3, pp. 430–437, 2003.

[16] R. Azevedo and J. G. Cromley, "Does training on self-regulated learning facilitate students' learning with hypermedia?" *J. Edu. Psychol.*, vol. 96, no. 3, pp. 523–535, 2004.

[17] M. Bannert, M. Hildebrand, and C. Mengelkamp, "Effects of a metacognitive support device in learning environments," *Comput. Human Behavior*, vol. 25, no. 4, pp. 829–835, 2009.

[18] E. R. Steinberg, "Cognition and learner control: A literature review, 1977-1988," *J. Comput. Based Instruct.*, vol. 16, no. 4, pp. 117–121, Oct. 1989.

[19] R. Azevedo, "Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning," *Edu. Psychol.*, vol. 40, no. 4, pp. 199–209, 2005.

[20] K. Falkner, R. Vivian, and N. J. Falkner, "Identifying computer science self-regulated learning strategies," in *Proc. Conf. Innovation Technol. Comput. Sci. Educ.*, New York, NY, USA, 2014, pp. 291–296.

[21] S. Narciss, A. Proske, and H. Koerndle, "Promoting self-regulated learning in web-based learning environments," *Comput. Human Behavior*, vol. 23, no. 3, pp. 1126–1144, 2007.

[22] C. Douce, D. Livingstone, and J. Orwell, "Automatic test-based assessment of programming: A review," *J. Edu. Resources Comput.*, vol. 5, no. 3, p. 4, 2005.

[23] S. H. Edwards, J. Snyder, M. A. Pérez-Quiñones, A. Allevato, D. Kim, and B. Tretola, "Comparing effective and ineffective behaviors of student programmers," in *Proc. 5th Int. Workshop Comput. Edu. Res. Workshop*, New York, NY, USA, 2009, pp. 3–14.

[24] N. J. Falkner and K. E. Falkner, "A fast measure for identifying at-risk students in computer science," in *Proc. 9th Annu. Int. Conf. Int. Comput. Educ. Res.*, 2012, pp. 55–62.

[25] J. B. Fenwick Jr, C. Norris, F. E. Barry, J. Rountree, C. J. Spicer, and S. D. Cheek, "Another look at the behaviors of novice programmers," *ACM SIGCSE Bull.*, vol. 41, no. 1, pp. 296–300, 2009.

[26] N. Michinov, S. Brunot, O. L. Bohec, J. Juhel, and M. Delaval, "Procrastination, participation, and performance in online learning environments," *Comput. Edu.*, vol. 56, no. 1, pp. 243–252, 2011.

[27] S. H. Edwards, "Using software testing to move students from trial-and-error to reflection-in-action," *ACM SIGCSE Bull.*, vol. 36, no. 1, pp. 26–30, 2004.

[28] V. Karavirta, A. Korhonen, and L. Malmi, "On the use of resubmissions in automatic assessment systems," *Comput. Sci. Educ.*, vol. 16, no. 3, pp. 229–240, Sep. 2006.

[29] L. Hakulinen, T. Auvinen, and A. Korhonen, "The effect of achievement badges on students' behavior: An empirical study in a university-level computer science course," *Int. J. Emerging Technol. Learn.*, vol. 10, no. 1, pp. 18–29, 2015.

[30] L. Hakulinen and T. Auvinen, "The effect of gamification on students with different achievement goal orientations," in *Proc. Int. Conf. Learn. Teaching Comput. Eng.*, Apr. 2014, pp. 9–16.

[31] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "gamification"," in *Proc. 15th Int. Acad. MindTrek Conf.: Envisioning Future Media Environ.*, New York, NY, USA, 2011, pp. 9–15.

[32] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? A literature review of empirical studies on gamification," in *Proc. 47th Hawaii Int. Conf. Syst. Sci.*, Hawaii, USA, 2014, pp. 3025–3034.

[33] A. Domínguez, J. S. de Navarrete, L. de Marcos, L. Fernández-Sanz, C. Pagés, and J.-J. Martínez-Herráiz, "Gamifying learning experiences: Practical implications and outcomes," *Comput. Edu.*, vol. 63, pp. 380–392, 2013.

[34] S. Abramovich, C. Schunn, and R. M. Higashi, "Are badges useful in education?: It depends upon the type of badge and expertise of learner," *Edu. Technol. Res. Develop.*, vol. 61, no. 2, pp. 217–232, Mar. 2013.

[35] L. Haaranen, P. Ihantola, L. Hakulinen, and A. Korhonen, "How (not) to introduce badges to online exercises," in *Proc. 45th ACM Tech. Symp. Comput. Sci. Educ.*, New York, NY, USA, 2014, pp. 33–38.

[36] S. Nicholson, "A user-centered theoretical framework for meaningful gamification," in *Games+Learning+Society 8.0*, Madison, WI, USA, 2012.

[37] E. L. Deci, R. Koestner, and R. M. Ryan, "Extrinsic rewards and intrinsic motivation in education: Reconsidered once again," *Rev. Edu. Res.*, vol. 71, no. 1, pp. 1–27, 2001.

[38] J. Hamari and V. Eranti, "Framework for designing and evaluating game achievements," in *Proc. 5th Int. Conf. Digital Res. Assoc. Conf.: Think Des. Play*, Hilversum, Netherlands, 2011, pp. 122–134.

[39] P. Denny, "The effect of virtual achievements on student engagement," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, 2013, pp. 763–772.

[40] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 687–698.

[41] V. Podgorelec and S. Kuhar, "Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments," *Elektronika ir Elektrotechnika*, vol. 114, no. 8, pp. 111–116, 2011.

[42] O. Scheuer and C. Zinn, "How did the e-learning session go? The student inspector," *Frontiers Artif. Intell. Appl.*, vol. 158, p. 487, 2007.

[43] L. Ali, M. Hatala, D. Gašević, and J. Jovanović. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Comput. Edu.* [Online]. *58(1)*, pp. 470–489. Available: sciencedirect.com/science/article/pii/S0360131511002089

[44] R. Mazza and V. Dimitrova. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *Int. J. Human-Comput. Studies* [Online]. *65(2)*, pp. 125–139. Available: sciencedirect.com/science/article/pii/S1071581906001303

[45] R. Mazza and L. Botturi. (2007, Apr.). Monitoring an online course with the GISMO tool: A case study. *J. Interactive Learn. Res.* [Online]. *18(2)*, pp. 251–265. Available: http://www.editlib.org/p/22972

[46] S. Bull and J. Kay, "Student models that invite the learner in: The SMILI:() open learner modelling framework," *Int. J. Artif. Intell. Edu.*, vol. 17, no. 2, pp. 89–120, 2007.

[47] A. T. Corbett and J. R. Anderson, " Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling User-Adapted Interaction*, vol. 4, no. 4, pp. 253–278, 1994.

[48] F. Linton and H.-P. Schaefer, " Recommender systems for learning: Building user and expert models through long-term observation of application use," *User Modeling User-Adapted Interaction*, vol. 10, no. 2/3, pp. 181–208, 2000.

[49] S. Bull and A. Mabbott, " 20000 inspections of a domain-independent open learner model with individual and comparison views," in *Proc. 8th Int. Conf. Intell. Tutoring Syst.*, 2006, vol. 4053, pp. 422–432.

[50] A. Mitrovic and B. Martin, "Evaluating the effect of open student models on self-assessment," *Int. J. Artif. Intell. Edu.*, vol. 17, no. 2, pp. 121–144, 2007.

[51] J.-D. Zapata-Rivera and J. E. Greer, "Interacting with inspectable Bayesian student models," *Int. J. Artif. Intell. Edu.*, vol. 14, no. 2, pp. 127–163, 2004.

[52] J. Kay, Z. Halin, T. Ottomann, and Z. Razak, "Learner know thyself: Student models to give learner control and responsibility," in *Proc. Int. Conf. Comput. Edu.*, 1997, pp. 17–24.

[53] V. Dimitrova, "STyLE-OLM: Interactive open learner modelling," *Int. J. Artif. Intell. Edu.*, vol. 13, no. 1, pp. 35–78, 2003.

[54] A. Mabbott and S. Bull, " Student preferences for editing, persuading, and negotiating the open learner model," in *Proc. 8th Int. conf. Intell. Tutoring Syst.*, 2006, vol. 4053, pp. 481–490.

[55] M. Wang, J. Peng, B. Cheng, H. Zhou, and J. Liu, "Knowledge visualization for self-regulated learning," *Edu. Technol. Soc.*, vol. 14, no. 3, pp. 28–42, 2011.

[56] S. Bull and T. Nghiem, "Helping learners to understand themselves with a learner model open to students, peers and instructors," in *Proc. Workshop Individual Group Modelling Methods that Help Learners Understand Themselves, Int. Conf. Intell. Tutoring Syst.*, 2002, pp. 5–13.

[57] I.-H. Hsiao, F. Bakalov, P. Brusilovsky, and B. Knig-Ries, "Progressor: Social navigation support through open social student modeling," *New Rev. Hypermedia Multimedia*, vol. 19, no. 2, pp. 112–131, 2013.

[58] T. D. Loboda, J. Guerra, R. Hosseini, and P. Brusilovsky, " Mastery Grids: An open source social educational progress visualization," in *Proc. 9th Int. Conf. Open Learn. Teaching Edu. Communities*, 2014, vol. 8719, pp. 235–248.

[59] E. Kosba, V. Dimitrova, and R. Boyle, " Using student and group models to support teachers in web-based distance education," in *Proc. 10th Int. Conf. User Modeling*, 2005, vol. 3538, pp. 124–133.

[60] G. Chen, C. Chang, and C. Wang. (2008). Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques. *Comput. Edu.* [Online] *50(1)*, pp. 77–90. Available: sciencedirect.com/science/article/pii/S0360131506000650

[61] J. Vassileva and L. Sun, "Using community visualization to stimulate participation in online communities," *E-Serv. J.*, vol. 6, no. 1, pp. 3–39, 2007.

[62] J. Vassileva, " Motivating participation in peer to peer communities," in *Proc. 3rd Int. Conf. Eng. Soc. Agents World III*, 2003, vol. 2577, pp. 141–155.

[63] H. Zhang, K. Almeroth, A. Knight, M. Bulger, and R. Mayer, "Moodog: Tracking students' online learning activities," in *Proc. World Conf. Edu. Multimedia, Hypermedia Telecommun.*, 2007, vol. 2007, no. 1, pp. 4415–4422.

[64] S. Govaerts, K. Verbert, E. Duval, and A. Pardo, "The student activity meter for awareness and self-reflection," in *Proc. Extended Abstracts Human Factors Comput. Syst.*, New York, NY, USA, 2012, pp. 869–884.

[65] H.-C. Schmitz, M. Scheffel, M. Friedrich, M. Jahn, K. Niemann, and M. Wolpers, " CAMera for PLE," in *Proc. 4th Eur. Conf. Technol. Enhanced Learn.: Learn. Synergy Multiple Disciplines*, 2009, vol. 5794, pp. 507–520.

[66] J. L. Santos, S. Govaerts, K. Verbert, and E. Duval, "Goal-oriented visualizations of activity tracking: A case study with engineering students," in *Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, New York, NY, USA, 2012, pp. 143–152.

[67] K. E. Arnold and M. D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success," in *Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, New York, NY, USA, 2012, pp. 267–270.

[68] K. Wäschle, A. Lachner, B. Stucke, S. Rey, C. Frömmel, and M. Nückles. (2014). Effects of visual feedback on medical students' procrastination within web-based planning and reflection protocols. *Comput. Human Behavior* [Online]. *41*, pp. 120–136. Available: sciencedirect.com/science/article/pii/S0747563214004749

[69] K. Verbert, S. Govaerts, E. Duval, J. Santos, F. Van Assche, G. Parra, and J. Klerkx, " Learning dashboards: An overview and future research opportunities," *Personal Ubiquitous Comput.*, vol. 18, no. 6, pp. 1499–1514, 2014.
[70] A. Kaplan and M. Maehr, " The contributions and prospects of goal orientation theory," *Edu. Psychol. Rev.*, vol. 19, no. 2, pp. 141–184, 2007.
[71] J. G. Nicholls, "Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance," *Psychol. Rev.*, vol. 91, no. 3, p. 328, 1984.
[72] C. S. Dweck, "Motivational processes affecting learning," *Am. Psychologist*, vol. 41, no. 10, p. 1040, 1986.
[73] J. G. Nicholls, M. Patashnick, and S. B. Nolen, "Adolescents' theories of education," *J. Edu. Psychol.*, vol. 77, no. 6, pp. 683–692, 1985.
[74] A. J. Elliot and J. M. Harackiewicz, "Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis," *J. Personality Social Psychol.*, vol. 70, no. 3, pp. 461–475, 1996.
[75] C. Midgley, A. Kaplan, M. Middleton, M. L. Maehr, T. Urdan, L. H. Anderman, E. Anderman, and R. Roeser, "The development and validation of scales assessing students' achievement goal orientations," *Contemporary Edu. Psychol.*, vol. 23, no. 2, pp. 113–131, 1998.
[76] M. Niemivirta, "Motivation and performance in context: The influence of goal orientations and instructional setting on situational appraisals and task performance," *Psychologia: An Int. J. Psychol. Orient*, vol. 45, no. 4, pp. 250–270, 2002.
[77] P. R. Pintrich, "Multiple goals, multiple pathways: The role of goal orientation in learning and achievement," *J. Edu. Psychol.*, vol. 92, no. 3, pp. 544–555, 2000.
[78] A. J. Elliot, H. A. McGregor, and S. Gable, "Achievement goals, study strategies, and exam performance: A mediational analysis," *J. Edu. Psychol.*, vol. 91, no. 3, p. 549, 1999.
[79] V. Karavirta, P. Ihantola, and T. Koskinen, "Service-oriented approach to improve interoperability of e-learning systems," in *Proc. 13th IEEE Int. Conf. Adv. Learn. Technol.*, Jul. 2013, pp. 341–345.
[80] L. Haaranen, L. Hakulinen, P. Ihantola, and A. Korhonen, "Software architectures for implementing achievement badges-practical experiences," in *Proc. Int. Conf. Learn. Teaching Comput. Eng.*, Apr. 2014, pp. 41–46.
[81] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The Am. Statist.*, vol. 46, no. 3, pp. 175–185, 1992.
[82] P. Buckley and E. Doyle, "Gamification and student motivation," *Interactive Learn. Environ.*, to be published.
[83] B. J. Zimmerman. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *Am. Edu. Res. J.* [Online]. *45(1)*, pp. 166–183. Available: http://aer.sagepub.com/content/45/1/166.abstract
[84] H. Tuominen-Soini, K. Salmela-Aro, and M. Niemivirta. (2012). Achievement goal orientations and academic well-being across the transition to upper secondary education. *Learn. Individual Differences* [Online]. *22(3)*, pp. 290–305. Available: sciencedirect.com/science/article/pii/S1041608012000118
[85] A. J. Elliot and K. M. Sheldon, "Avoidance achievement motivation: A personal goals analysis," *J. Personality Social Psychol.*, vol. 73, no. 1, p. 171, 1997.

Tapio Auvinen received the master of science (Tech) degree in computer science from Aalto University, Finland, in 2010. He is a doctoral candidate in the Learning + Technology research group at Aalto University. His research focuses on supporting self-regulated learning in online learning environments.

Lasse Hakulinen received the master of science (Tech) degree in computer science and engineering from Aalto University, Finland, in 2010. He is a doctoral candidate in the Learning + Technology research group at Aalto University. His research focuses on gameful approaches to enrich computer science education (e.g. gamification and alternate reality games).

Lauri Malmi is a professor of computer science at Aalto University. His main research area is computing education research, and he is leading the Learning + Technology research group (LeTech), which focuses on development and evaluation of advanced learning environments and learning tools for programming education. His additional research interests include qualitative investigations of how students understand programming concepts and processes, and surveys in computing and engineering education research literature. He is currently a member of the editorial boards of the *ACM Transactions of Computing Education*, *ACM Inroads*, *Informatics in Education*, and the *IEEE Transactions of Learning Technologies*.