

The effect of providing learning analytics on student behaviour and performance in programming: a randomised controlled experiment

Author(s)

Hellings, Jan; Haelermans, Carla

DOI

[10.1007/s10734-020-00560-z](https://doi.org/10.1007/s10734-020-00560-z)

Publication date

2022

Document Version

Final published version

Published in

Higher Education

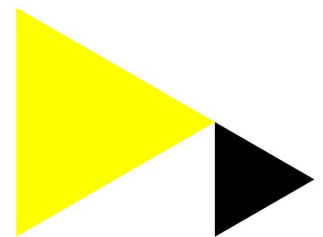
License

CC BY

[Link to publication](#)

Citation for published version (APA):

Hellings, J., & Haelermans, C. (2022). The effect of providing learning analytics on student behaviour and performance in programming: a randomised controlled experiment. *Higher Education*, 83(1), 1-18. <https://doi.org/10.1007/s10734-020-00560-z>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: <https://www.amsterdamuas.com/library/contact>, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.



The effect of providing learning analytics on student behaviour and performance in programming: a randomised controlled experiment

Jan Hellings¹ · Carla Haelermans²

Published online: 10 June 2020

© The Author(s) 2020

Abstract

We use a randomised experiment to study the effect of offering half of 556 freshman students a learning analytics dashboard and a weekly email with a link to their dashboard, on student behaviour in the online environment and final exam performance. The dashboard shows their online progress in the learning management systems, their predicted chance of passing, their predicted grade and their online intermediate performance compared with the total cohort. The email with dashboard access, as well as dashboard use, has positive effects on student behaviour in the online environment, but no effects are found on student performance in the final exam of the programming course. However, we do find differential effects by specialisation and student characteristics.

Keywords Randomised controlled experiment · Learning analytics dashboard · Student behaviour · Student performance · Higher education

Introduction

Many policymakers and higher education institutes (HEIs) are concerned with the low completion rates in higher education, which are often ascribed to the high number of student dropouts in the first years of the bachelor education. In the USA, only a little over half of the

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10734-020-00560-z>) contains supplementary material, which is available to authorized users.

✉ Carla Haelermans
Carla.haelermans@maastrichtuniversity.nl

Jan Hellings
j.f.hellings@hva.nl

¹ Department of Computer Science, Amsterdam University of Applied Sciences, Wibautstraat 2-4, 1091 GM Amsterdam, The Netherlands

² Research Centre for Education and the Labour Market (ROA), School of Business and Economics, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

students graduate from a bachelor's degree within 6 years (Symonds et al. 2011). In the Netherlands, only 40 % of the students graduate within 5 years (from a 4-year program) and in the first year, 36% of the students drop out (Ministry of Education Culture and Science 2016). Recent research has shown that nowadays students seem less prepared and often have a different study attitude than before, leading to low performance (Beattie et al. 2018). It is also often said that a lack of motivation and self-regulatory skills from the students might be the reason for the high dropout numbers (Kitsantas et al. 2008), as well as that they have difficulties predicting or knowing during the course, whether they are doing well or not. These issues are prevalent in higher education in general, but specifically in engineering education. Therefore, the management and teacher bodies at the HEIs are looking for ways to better inform students, and increase involvement and self-regulation by students, to make students more responsible, have them study more and to decrease dropout rates. Given the increasing use of online environments, in combination with the desire to increase self-regulation and involvement of students, one option is to use visualisation technique such as dashboards based on learning analytics, in combination with early warning systems, to signal to students how they are doing on their activities in online environments and to remind them to study regularly (Bos 2016) (Lauría et al. 2013), and by sending regular emails with access links to the dashboard and the learning analytics, which might also serve as a reminder for their coursework (similar to what has been done previously in the literature, see, e.g. Beattie et al. 2018; Oreopoulos and Petronijevic 2018).

The aim of the study at hand is to analyse the effectiveness of a learning analytics dashboard (LAD), with weekly email with a personalised link to the dashboard, aimed to signal students about their study progress as well as about their chances of success, by showing visual performance indicators of the student and the average of the cohort. We are interested in the effects on both online behaviour and performance of students as well as their final performance in the course. Our study is carried out among 556 freshman computer science students of a Dutch University of Applied Sciences who attend a Java programming course to acquire their basic Java programming skills in 2015. About half of these students ($n = 276$) were conditionally random assigned to the treatment group of having access to the LAD during the 8-week course, and the other half ($n = 280$) to the control group, without dashboard. The students in the treatment group received a weekly email with a personalised link to the LAD. All students are supposed to weekly perform exercises in the online environment of the course. We find that the weekly email and corresponding dashboard access (intent-to-treat, ITT), as well as dashboard use (two-stage least squares, 2SLS), have positive effects on student performance in the online environment during the course (standardised effect sizes of 0.2 and 0.25), but not on student behaviour in the online environment, or on student performance in the final exam or their chances of passing the programming course. However, heterogenous analysis shows that final exam grades are significantly higher for treatment students in the specialisation software engineering and are significantly lower for treatment students in game development (the two largest specialisations).

So far, there is not much evidence that using learning analytics dashboards for the student will influence student behaviour and ultimately exam performance. There are only few causal studies on the effectiveness of learning analytics dashboards (LAD) (e.g. Dodge et al. 2015; Kim et al. 2015; Park and Jo 2015), as also emphasised by Bodily and Verbert (2017). These few studies show mixed results, and each has their flaws. The study of Park and Jo (2015) is one of the few examples that analysed the effects of an LAD in higher education. In this study, students from two classes were randomly assigned to the treatment group and received a LAD

(73 observations in total, randomisation at the class level). No significant effects were found on student learning outcomes, but it is very well likely that this is due to the low number of observations, and therefore low power, of this study. Dodge et al. (2015) also perform a randomised experiment among 883 students using learning analytics in two courses with historically low pass rates. They find no effect on grades of the treatment in general, but they do see an effect on low-SES students, for one of the two courses. In another study, Kim et al. (2015) investigated the effect of a LAD in a management Statistics e-learning course at a private university located in Korea. The treatment group receiving the LAD scored 12% higher than the control group. However, there seems to be no random assignment, so it is unclear whether the found effect can be attributed to the LAD or whether there may have been selection into the treatment group that explains this effect.

Apart from these causal studies, there is larger body of correlational studies looking into learning analytics dashboards, sometimes combined with an early warning system (e.g. Hu et al. 2014; Verbert et al. 2013), finding mixed results of the relationship between the learning analytics dashboard and student performance and motivation. The recent overview study of Teasley (2017) sums this up nicely as well. From these studies, it seems that a certain type of intervention might be beneficial for student performance, if certain requirements of the dashboard, such as integrating data from student information systems, are fulfilled (e.g. Teasley 2017, and references therein).

So, although quite some studies on learning analytics dashboards exist, only few provide causal evidence on the effects of LADs and none of these causal studies includes prediction models for the content of the LAD. Furthermore, the results of the few causal and mostly correlational studies are mixed and none of these experimental studies has been performed in a Western context, making these results not automatically applicable to the situation in most European countries. Although quite some correlational studies exist from these countries pointing in the direction of a positive effect, we cannot draw causal conclusions from these studies. Lastly, to our best knowledge, there is no other study combining the learning analytics dashboard with an early warning system, using prediction models with data from the LAD and from student characteristics to inform students about their chance of success. The only two studies (Hu et al. 2014 and Verbert et al. 2013) that include prediction models with a LAD are correlational and not causal. Therefore, this study aims to add to this body of knowledge with a randomised experiment testing the effectiveness of sending a weekly email with a link to a learning analytics dashboard (with prediction models therein) in a freshmen programming course in higher education in the Netherlands. Additional analyses have shown that success in this programming course relates in a disproportionately high way to success in the first year of the bachelor's degree (significant odds ratio of 12), making it an extremely relevant course for student success. Therefore, the main contribution of our study to the literature and to policy is the unique combination of causal analysis of LADs combined with prediction models in a Western setting in higher education on the effect of providing students with weekly emails with a link to their learning analytics dashboard in their programming course. Another contribution is that the combined treatment of the email and the dashboard is a low-cost innovation that is easy to implement. Furthermore, other innovative features of the dashboard (besides the prediction models) are that it has a signalling function (with an early warning system using prediction models) and includes elements of competition.

In the remainder of this paper, we first present the context of the study, describing both the learning analytics dashboard treatment and the Dutch higher education system in which this experiment took place. The “Data and methodology” section presents the data and

methodology, describing also the identification strategy, compliance with the assignment to treatment group and the dashboard use. In the “Results” section, we describe the results. We also perform some additional analyses on potential heterogenous effects. The “Conclusion and discussion” section consists of the conclusion and a discussion.

Context of the study

The treatment

The treatment consists of a learning analytics dashboard (LAD) for freshmen students in a Java Programming Course at a University of Applied Sciences in the Netherlands. The Java Programming course is a blended learning course (Bos 2016) with an online practice environment, which consists of both a Moodle (<https://moodle.org/>) course with quizzes and practical assignments and a Myprogramming lab (MPL) e-text environment. The learning analytics dashboard uses the behaviour of the student in the Moodle and Myprogramming lab environment to make predictions about the student’s chance of passing the course, their predicted grade, and to show students their progress in the online environment. In the learning analytics dashboard, the students get information on their expected grade (and thereby chances of passing the course) and the expected grade of their fellow students (based on the predictions of the whole cohort) and about their progress in the Moodle and Myprogramming Lab, again compared with their fellow students. Treatment students receive a weekly email with a direct link to their personal dashboard, to remind them of the existence of the dashboard and send them information about their expected result.

In Fig. 1, an example is shown of the dashboard that students had access to if they were in the treatment group. The learning analytics dashboard shows at the top (1) the expected grade ranging from 1 to 10 and (2) the average grade of the total cohort. Below this, it presents the expectation with respect to whether the student will pass the course, and the certainty with which this prediction is made. In the middle, it shows a progress bar (3) with the total online progress in Moodle and MPL. At the bottom, there are two pie charts showing the progress in (4) Moodle and (5) Myprogramming Lab with the averages of the total cohort. As Fig. 1 shows, the dashboard contained many elements that have previously also been used in studies on learning analytics dashboards and/or early warning systems, as is discussed in Jivet et al. (2017) and shown in their Table 2.

Tempelaar et al. (2015) and Hu et al. (2014) found in their research a strong relation between the online behaviour and the course results and it is therefore expected that providing students with information on their online behaviour and performance so far (resulting in the prediction of the grade for the course) will increase the online behaviour and via that way the results in the programming course. As Teasley (2017) discusses, the visual representation is a meta-cognitive tool (Durall and Gros 2014) that relates to increased awareness, self-reflection and sense-making (Bodily and Verbert 2017; Verbert et al. 2014). Furthermore, previous research shows that a visual representation of study progress might be an effective way of informing students (e.g. Tufte 1991). Informing students about their study progress is said to be a good stimulator for students to increase their work effort and time spent studying (e.g. Bandiera et al. 2015; Chevalier et al. 2014). Furthermore, providing a comparison with performance of fellow students is shown to relate to increased competition which is also said to have positive effects on the subsequent (practice) behaviour of students (e.g. Lam et al.

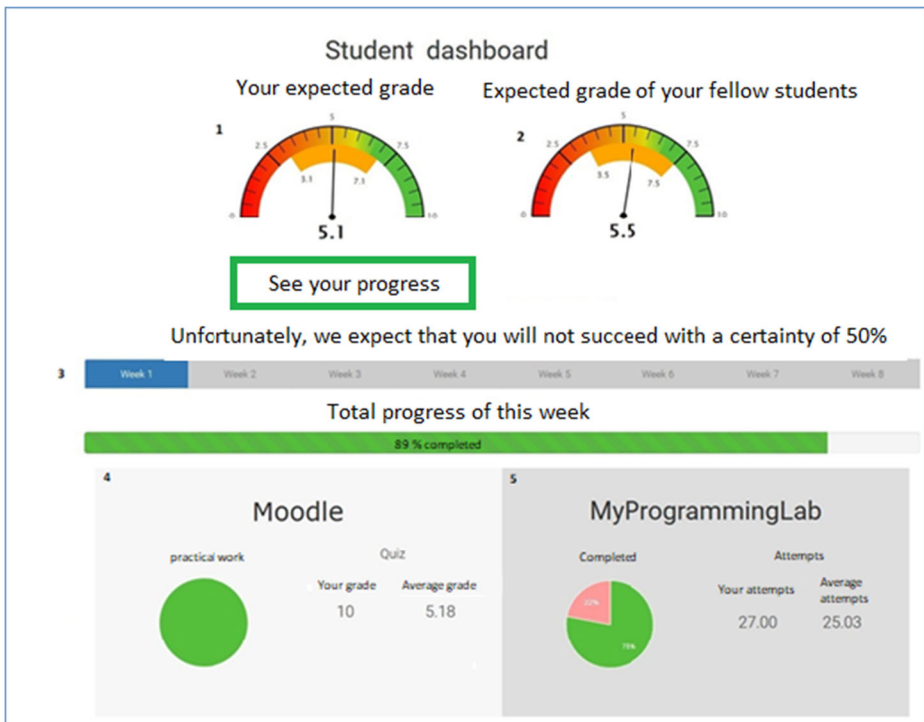


Fig. 1 Dashboard of a student in week 1

2004). At the same time, Teasley (2017) and references therein warns for this competitive element, as it might have unanticipated consequences on motivation and through that on performance (Christy and Fox 2014; Hanus and Fox 2015).

To build the learning analytics dashboard, the online behaviour as well as student background characteristics (which matter for course performance, as we know from the analyses of data from previous cohorts) had to be collected and analysed. The analyses of the online behaviour in the population of the 684 students in the year before the experiment (2014) furthermore showed that a correlation exists between the online behaviour and the result of programming. Students who did all the online exercises scored an average of 6.8 (SD = 2.7) on a scale of 1 to 10 and the ones who did not write all the online exercises scored an average of 4.8 (SD = 3.5) ($t(682) = -7.84, p < 0.00$).

In Verbert et al. (2012), six objectives of learning analytics are distinguished: predicting learner performance and modeling learners, suggesting relevant learning resources, increasing reflection and awareness, enhancing social learning environments, detecting undesirable learner behaviours and detecting affective states, like boredom confusion, frustration and Eureka of learners. In this study, the focus will be on two of them: (1) to predict the learner performance and (2) to increase reflection and awareness of the learner.

The learning performance is predicted using data mining techniques. Dynamic data from the use of the online environment of the Moodle and the Myprogramming Lab in addition to the demographic and historical data can be used for reliable forecasting of student performance (Hu et al. 2014). The prediction models are based on data from the previous cohorts (2014) but are also checked with the 2013 cohort and with the real outcomes of the 2015 cohort, leading

to similar results (the chi-square test showed that the passing rate predicted by the prediction model and the actual passing were highly associated as well as the type I and II errors). We built a linear prediction model to predict the grade, and we used Decision Stump combined with AdaBoost to predict the chance of success in the course (a machine learning model combined with an adaptive boosting algorithm that uses iterations of formative learning outcomes to make the best forecasts, see Wayne and Langley (1992), Rojas (2009) and Hu et al. (2014) for more information on this technique). The prediction models are based on student background variables (prior education level and specialisation) and the variables of online behaviour (the number of practical assignments, the average grade of those assignment, the average grade of the online quizzes and the average grade of the online mastery exercises). The prediction itself for the 2015 cohort is made by applying these prediction models to the learning analytics data from the 2015 cohort.

The Java Programming course has a study load of 3 ECTS (*European Credit Transfer System*) which corresponds to 84 h of study attainment. There are 30 class hours, and the rest is pre-class preparation. Unlike previous years, the online assignments were not mandatory for students (i.e. the online activities had a formative nature this year), due to an organisational change at the institute. The class time is reserved for acquiring programming skills by programming the practical assignments. These practical assignments are uploaded to Moodle and graded by the lecturer. At the end of the course, the students are tasked with programming a small Java program and are then graded (1–100, 100 being the highest grade) on the results of their efforts. Students who failed the first exam can take a resit 5 weeks later. Note that for this study we only use the first time each student wrote the final exam, (first sit, for almost all students) for fair comparison of grades, but also because the dashboard is only generated during the first 8 weeks and not during the resit period.

The expectation is that the learning analytics dashboard and corresponding weekly email will have a signalling function, as the latter reminds them of the dashboard which in turn, when they click on the link, informs students about their performance so far in the course (in absolute numbers, but also relative to fellow students), and what that means for their current chances of passing the course. Previous literature has shown that results might be mostly expected due to the information on relative performance, although earlier studies have also showed positive effects of providing feedback on absolute performance (see, e.g. Azmat and Iriberry 2010, and references therein). The signal that is given through the learning analytics in the dashboard environment is then expected to stimulate the student to practice in the online environment, and work on the weekly (digital) assignment. This is expected to directly influence the chances of receiving a higher grade in the final exam of the programme course, and hence passing the course. The latter expectation is based on previous cohorts where a large correlation was found between online practice behaviour (and particularly the assignments in the second half of the course) and the exam performance and course outcome. Note that although it is possible that students communicate with each other, possibly leading to control students also being reminded to work on their online assignments (which in turn might lead to our effect being a lower bound), control students do not have access to a dashboard and therefore do not get an indication of both their absolute and relative performance.

Dutch higher education and the ICT-curriculum

Dutch secondary education, which precedes higher education, has a tracking system from 7th grade on with 3 different tracks. Preparatory secondary vocational education (VMBO) takes 4

years until graduation and prepares for senior secondary vocational education and training (MBO). Senior general secondary education (HAVO) takes 5 years and university preparatory education (VWO) takes 6 years until graduation. Students are admitted to a University of Applied Sciences (HBO) if they have a diploma from senior general secondary education (HAVO), senior secondary vocational education (MBO) or pre-university education (VWO) (Fig. 2) (Nuffic 2015). The study at hand takes place at the computer science (ICT) bachelor programme at a University of Applied Sciences (HBO), abbreviated to HBO-ICT in this paper. Within the HAVO and the VWO track, there are two specialisations, economics and science, where the science specialisation is more related to HBO-ICT. In MBO, there are related and non-related specialisations. Specialisations like MBO-software engineering (SE) or system and network engineering (SNE) are related to the HBO-ICT. For interpretation purposes later in the paper, it is important to mention that in the Dutch educational system, grades range from 0 to 10, with 10 being the highest grade. A grade of 5.5 is considered sufficient to pass the course.

Students registering for HBO-ICT immediately choose a specialisation when they sign up for the study. They are put into classes based on the specialisation. The computer science study at hand has five specialisations: game development (GD), business IT management (BIM), system and network engineering (SNE), software engineering (SE) and technical computing (TC). The study consists of several specialisation specific courses and several joint courses that are compulsory for all students. The java programming course that we analyse in this study is one of the compulsory courses, as it is a crucial course for later courses in the second and third year of the bachelor program, regardless of the specialisation.

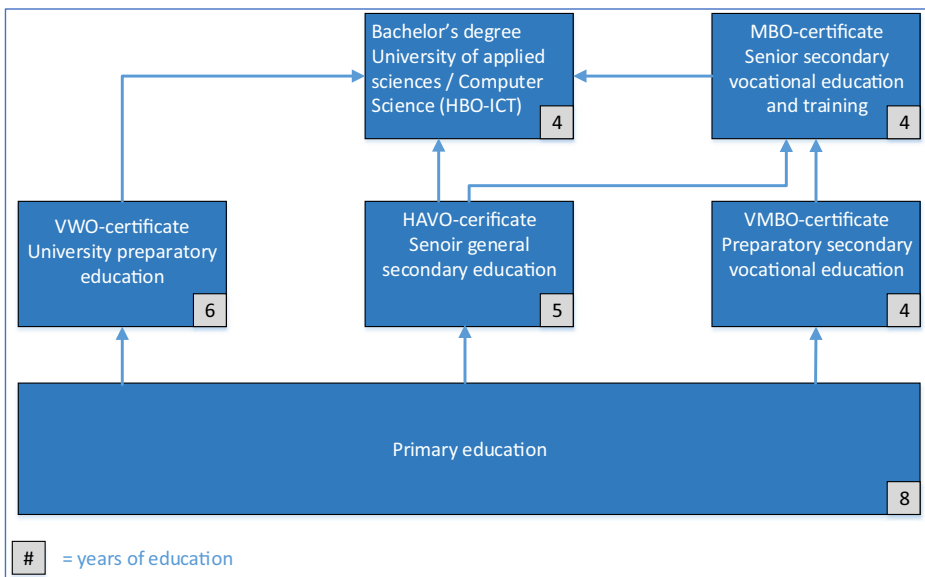


Fig. 2 Dutch education system until the bachelor's degree with the possible routes

Data and methodology

Data

This experiment is conducted at a University of Applied Sciences among 556 freshman students that started the study programme computer sciences in 2015. Both before the start of the study and during the Java programming course, a lot of data is collected on these students, their study behaviour and study performance. For the analysis of this paper, we use three types of data: student characteristics, student online practice behaviour and student performance.

The student characteristics that we include are gender, age, ethnicity (Dutch or other), a dummy whether the student was born in the Netherlands, order of registration (i.e. the sequence number of the student with regard to when he/she subscribed for the study of computer science, as a proxy for motivation), the specialisation and the class of the student within the study computer sciences and the previous educational level that the student attended. The latter is used as an indication of ability level of the student, since we do not have prior grades (because this is the first course in higher education for the students). The combination of previous educational level and specialisation/major at the previous level is an indication of prior knowledge of the topic.

The student online practice behaviour contains information on 3 indicators that are registered each week: online practical assignment (dummy whether student has done the assignment, for each of the 6 weeks of the study), online quizzes (grades between 0 and 10) and the average mastery score of a set of online exercises (score ranging from 0 to 100, number of exercises varying per week). Table 1 presents the descriptive statistics of the online behaviour indicators. An important observation from Table 1 is that we do not have all these indicators for all 556 students. There are different numbers of observations for each indicator, which we will explicitly take into account in our later analyses.

The student performance indicators that we include are both the grade for the programming course (ranging from 1 to 10, using the Dutch grading system) and an indicator for whether the student passed the course. Table 1 also includes the descriptive statistics of the student performance indicators. Note that we also have some missings in the grade for the programming course, as some students did not write an exam at all for this course (and thereby automatically failed the course).

Identification strategy

The study is organised as a randomised controlled trial (RCT) in which individuals are randomly assigned to either the treatment group (email and dashboard) or the control group.

Table 1 Descriptive statistics student online practice behaviour and student performance

Variable	<i>N</i>	Mean	SD	Min	Max
Number of sufficient practical assignments	465	3.02	1.80	0	6
Average grade of online quizzes	433	6.56	1.69	1.93	10
Average score mastery level exercises	478	52.23	29.33	1.83	100
Grade in programming course	456	57.43	28.07	0	100
Passing the programming course (1 = yes)	556	0.43	0.49	0	1

Individual randomisation with this sample size gives high confidence that the treatment and control group will be similar in both observable and unobservable characteristics.

Figure 3 presents the experimental setup of the study. Out of the initial 558 freshman computer science in the cohort, 2 students decided not to take part in the experiment and chose to opt out. The remaining 556 students are individually randomised into the treatment group ($n = 276$) or the control group ($n = 280$). We applied stratified (conditional) randomisation at the specialisation level, as student performance at the programming course is likely to differ by specialisation (based on analyses of the grades of the previous two cohorts).

Our analysis shows that the mean differences of individual categories are not statistically significant at the 5% level. Chi-squared statistics of the specialisation variable, the previous educational level variable and class distribution also do not show any significance. A joint F -test on all the characteristics also does not show a significant difference ($F(38, 517) = 0.87, p = 0.70$). Given these results, we expect the random assignment process to have functioned well, producing a treatment and control group that were ‘equal in expectation’ at the start of the experiment.

Compliance with assignment to treatment group

As it happens, not all students complied with the random assignment in the experiment. Although we technically disabled the dashboard for control students, treatment students can still decide not to open the dashboard, and some indeed did not use the dashboard at all. Among the students that did use the dashboard, there is large variation in the usage. This means that we have three groups of students: (1) students that were assigned to the treatment group ($A = 1$) and also opened the dashboard at least once ($D = 1$) ($n = 205$), (2) students that were assigned to the control group ($A = 0$) and therefore could not use the dashboard ($D = 0$) ($n = 280$) and (3) students that were assigned to the treatment group ($A = 1$), but who did not open the dashboard at all ($D = 0$) ($n = 71$). Comparison of these three groups of students shows that this compliance is not a random process, and this will explicitly be taken into account in the used empirical strategy, as will be explained in more detail in the “Empirical strategy” section.

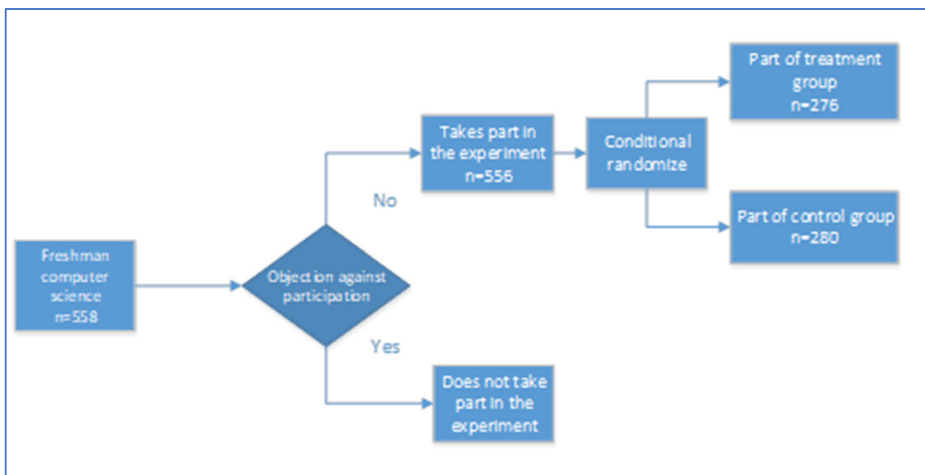


Fig. 3 Experimental setup of the study

Dashboard use

As described above, the setup of the experiment is such that students have to actively click on the email to open the dashboard and see its content. Therefore, in addition to studying the effect of giving the student access to the dashboard, it is also interesting to look into the intensity of the treatment.

In Fig. 4, we show the descriptive statistics of the dashboard usage of student per week (there were 2 weeks in between the end of the experiment and the exam, in which students could also open the dashboard). Figure 4 shows the spread of number of students that used the dashboard at least once and the number of times per week the dashboard was used by these same students. In total, 205 students used the dashboard at least once during the whole period of 8 weeks, for an average of 18 times in total. However, the range is large, as the number of times the dashboard was opened varies from 1 to 203 times in total.

Figure 4 furthermore shows that the average number of times the student opened the dashboard each week declines during the last 6 weeks of the course, to an average of a little less than 3 times per week in week 8. The number of students that used the dashboard was highest in the first 3 weeks (up to almost 200 in the second week) but declined to a relatively stable number of around 75 from week 4 on.

Empirical strategy

As the selective non-compliance that we showed in the “Compliance with assignment to treatment group” section is likely to create (a small) bias in the estimated effect, that is, a simple comparison of the control and experiment groups reveals the intent-to-treat effect (ITT) and the effect of sending a weekly email, but not the average treatment effect (ATE) of actually using the learning analytics dashboard, because a specific group seems to have self-selected away from treatment. Therefore, we control for the non-compliance using a two-stage least

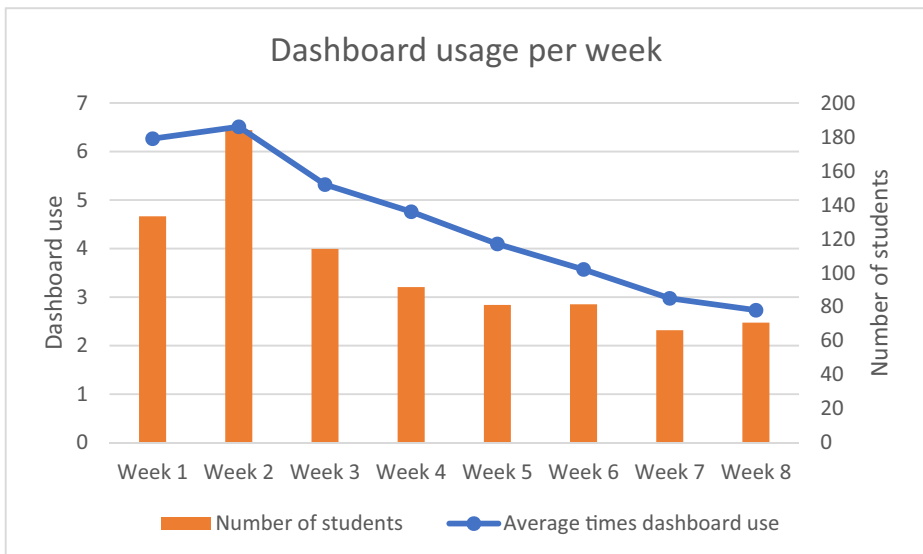


Fig. 4 Statistics on dashboard use (only students that used the dashboard at least once)

squares (2SLS) instrumental variable approach in order to control for the actual use of the tool (see Murnane and Willett 2011 Chapter 11 for an extensive description of an example of this method applied to education). Here, we use the dummy that indicates the random assignment to the treatment as an instrument for the actual use of the dashboard. By doing this, we ensure that we can still use the randomisation to analyse causal effects, but at the same time, we analyse the more interesting question whether actual use of the dashboard has a positive effect on performance. For this analysis, it is important that the assignment to treatment or control group is (highly) correlated with the use of the tool, which is represented in the statistically significant and large coefficient of the treatment group indicator (access to the tool) in the first stage regression. This first stage regression estimates the probability that students that were randomised into the treatment group actually use the dashboard. In the second stage of the regression, we use the outcome of the first stage (predicted probability) to estimate the effect of using the dashboard. Rather than using the observed use of the dashboard, we now gauge the effect of predicted use, being an indicator that does not entail unobserved reasons for the use of the tool but strictly reflects the effect of the offer of the dashboard. In the regressions, we add the following control variables: age, gender, ethnicity, order of registration, specialisation and former education level. In our analyses, we standardise our outcome variables such that all the variables have a mean of zero and a standard deviation of one. This implies that differences between treatment and control group in the t tests and the regression coefficient of the treatment or usage dummy can be interpreted as standardised effects (i.e. Cohen's d), where 0.2 is a small effect, 0.5 is a medium effect and 0.8 is a large effect.

Results

In this section, we present the results of the dashboard: first, on online behaviour of students throughout the course and next, on student performance on the final outcomes. Lastly, we present heterogenous analyses. In the “Effects on online behaviour” section and “Effects on student performance” section, we present 4 types of analyses in each table. First of all, we show the intent-to-treat (ITT), which presents the effect of sending the weekly email with the link to the dashboard (providing access to the dashboard). The second type of analysis is the simple OLS-analysis, in which we look at the relation between actual dashboard use and the outcome, without controlling for selection effects. In the third analysis, we present the first stage results, which analyse the take-up rate (dashboard use) of students that were in the treatment group. The final analysis shows the two-stage least squares (2SLS) analysis in which we analyse the effect of actual dashboard use while controlling for the randomisation into treatment and control group. In the “Heterogenous effects” section, where we check heterogenous effects, we only show the 2SLS results.

Effects on online behaviour

Before we analyse the effects on student performance, it is important to analyse whether the dashboard has made a difference in students' online behaviour. We do this because we assume the effect of the dashboard to work via the online behaviour of the student, as in previous cohorts, the online behaviour was a large predictor of student performance on the final test. As explained above, we only include students that have shown online behaviour for each of the indicators at least once (regardless of whether these students are in the treatment or control

group). This implies that we have different numbers of observations for each online behaviour outcome. Note that we have checked the comparison of treatment and control groups for each of the subselections, which did not give us reason for concern for any of the online behaviour outcomes.

First, we analyse the effect on the number of practical assignments that the student has completed that were graded as sufficient. Table 2 shows the results of this analysis, with standardised coefficients, where we only include students that wrote at least one out of the 6 practical assignments over the 8 weeks of the course. Table 2 shows that there is no effect of the dashboard use or access on the number of sufficient assignments written by the students. Although the simple OLS shows positive significant results, there is clearly selection, as these results disappear as soon as we control for the original randomisation.

In Table 3, again with standardised coefficients, we show the effects on the average grade of the online quizzes, again, only for students who wrote at least one quiz during those 6 weeks. The average is based only on the quizzes they wrote. Table 3 shows that only having access to the dashboard already increases the average grade in the quizzes by 0.16 of a standard deviation. The first stage shows that the randomisation to the dashboard group is a strong predictor for actual dashboard use. The results in the OLS are much higher than in the ITT, but they are an overestimation, as the 2SLS shows. The effect of dashboard use on average quiz grade is about 0.2 of a standard deviation. However, we also see that the coefficients of the ITT and the 2SLS are almost equal, implying that there is no separate nudging effect from sending the email with the link to the dashboard. The potentially signalling function of the email to remind students of their coursework does not have a separate effect, regardless of the dashboard use. Note that the results are very stable between the models with and without covariates.

In Table 4, we show the final indicator of online usage, namely the average grade of the online mastery exercises, where we again only include students that wrote the mastery exercise at least once. Table 4 shows similar results to Table 3, with respect to significance and the OLS overestimating the effect of actual dashboard use. Note that the coefficients are similar in magnitude as before. Table 4 shows that access to the dashboard gives a significant higher performance on the online mastery exercises, which works through the actual use of the dashboard (although the intent-to-treat coefficients are also significant, but smaller).

Table 2 Regression results for effect on the number of sufficient practical assignments over all weeks (standardised coefficients)

Outcome = number of pract. sufficient	ITT		OLS		First stage		2SLS	
Dashboard group	0.024	0.096			0.848***	0.851***		
	(0.090)	(0.080)			(0.020)	(0.020)		
Dashboard use			0.204**	0.223***			0.028	0.112
			(0.090)	(0.080)			(0.110)	(0.090)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
<i>N</i>	465	465	465	465	465	465	465	465
<i>F</i> -stat	0.067	5.436	4.773	5.662	1304.764	38.819	0.067	5.487
<i>R</i> ²	0.002	0.261	0.008	0.271	0.738	0.751	0.000	0.268

Standard errors between brackets. Covariates are specialisation, previous educational level and class

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3 Regression results for effect on the average grade of the online quizzes over all weeks (standardised coefficients)

Outcome = average grade quizzes	ITT		OLS		First stage		2SLS	
Dashboard group	0.158*	0.160*			0.845***	0.845***		
	(0.100)	(0.090)			(0.020)	(0.030)		
Dashboard use			0.269***	0.242***			0.187*	0.189*
			(0.100)	(0.090)			(0.110)	(0.100)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
<i>N</i>	433	433	433	433	433	433	433	433
<i>F</i> -stat	2.719	3.041	7.800	3.169	1159.853	33.880	2.747	3.067
<i>R</i> ²	0.004	0.149	0.015	0.157	0.728	0.738	0.014	0.156

Standard errors between brackets. Covariates are specialisation, previous educational level and class

p* < 0.10, *p* < 0.05, ****p* < 0.01

Effects on student performance

Now that we have established that students that have access to the dashboard, and actually use it, have a higher performance in the online exercises, the next question is whether this accumulates to a higher performance on the final exam of the course as well. First, we look at the grade in the first chance of the final exam, for every student that wrote the exam, and next at the chance of passing the course (for all students, regardless of whether they wrote the exam, if they did not write the exam, they automatically did not pass the course).

Table 5 shows that there is no effect of dashboard access nor dashboard use on the grade of the final exam of the programming course, not even in the OLS analysis. The effect size is 0.05 of a standard deviation (in the analysis with covariates), and the standard errors are large. Given our sample size, we should be able to detect an effect of 0.23 of a standard deviation, which strengthens our suspicion that overall there is no effect on grades.

Table 6 shows a similar picture as in Table 5, except that here we do see a significant coefficient in the OLS analysis, which disappears once we properly control for the randomisation (again indicating selection). Based on the above found effects on online behaviour, we might have expected to find a positive effect here as well. However, further analyses including interaction terms (moderator effects) between the two variables of online

Table 4 Regression results for effect on the average grade of the online mastery exercises over all weeks (standardised coefficients)

Outcome = average score mastery level	ITT		OLS		First stage		2SLS	
Dashboard group	0.162*	0.192**			0.795***	0.793***		
	(0.090)	(0.080)			(0.030)	(0.030)		
Dashboard use			0.487***	0.436***			0.204*	0.242***
			(0.090)	(0.080)			(0.110)	(0.100)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
<i>N</i>	478	478	478	478	478	478	478	478
<i>F</i> -stat	3.143	5.477	28.547	6.339	941.544	27.494	3.243	5.682
<i>R</i> ²	0.004	0.263	0.055	0.298	0.7664	0.679	0.035	0.290

Standard errors between brackets. Covariates are specialisation, previous educational level and class

p* < 0.10, *p* < 0.05, ****p* < 0.01

Table 5 Regression results for effect on the grade of the course (standardised coefficients)

Outcome = grade	ITT		OLS		First stage		2SLS	
Dashboard group	- 0.006	0.034			0.835***	0.835***		
	(0.090)	(0.080)			(0.020)	(0.020)		
Dashboard use			0.153	0.120			- 0.007	0.041
			(0.100)	(0.090)			(0.110)	(0.100)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
<i>N</i>	456	456	456	456	456	456	456	456
<i>F</i> -stat	0.004	5.004	2.584	5.073	1167.398	34.802	0.004	5.015
<i>R</i> ²	0.002	0.246	0.003	0.249	0.719	0.733	0.003	0.247

Standard errors between brackets. Covariates are specialisation, previous educational level and class

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

behaviour that we found a significant effect on with assignment to and compliance with the treatment were not significant and had a coefficient close to zero in both analyses (results available upon request from the corresponding author), implying that the relation between online behaviour and student performance is not significantly different between the treated and the control students. It is also possible that the one indicator of online behaviour for which we did not find an effect (online assignments) is the most important one in predicting the course outcome. Another possible explanation is that all students realised that they have not used the online environment to practice (enough) and all put such a large amount of study hours in preparing for the exam that the small effects that were found on online behaviour fade out because of the large amount of preparation. A last possibility is that there are heterogenous effects, which may explain the almost zero result on grades, which we will explore in the next section.

Heterogenous effects

Table 7 shows the heterogenous effects of the effect analysis on grade by specialisation. Because some specialisations are more directly related to the programming course, it is likely that students from the more related specialisations are more motivated and/or see the relevance of the course for their future studies more. These separate analyses might give an indication

Table 6 Regression results for effect on the passing rate of the course (standardised coefficients)

Outcome = passed	ITT		OLS		First stage		2SLS	
Dashboard group	0.013	0.062			0.743***	0.747***		
	(0.080)	(0.080)			(0.030)	(0.030)		
Dashboard use			0.335***	0.231***			0.017	0.084
			(0.090)	(0.080)			(0.110)	(0.100)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
<i>N</i>	556	556	556	556	556	556	556	556
<i>F</i> -stat	0.022	4.947	14.912	5.206	805.543	26.957	0.022	4.984
<i>R</i> ²	0.002	0.217	0.024	0.228	0.592	0.646	0.001	0.223

Standard errors between brackets. Covariates are specialisation, previous educational level and class

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

why the overall effect is around zero (and not significant). In Table 7, we see that for students from game development, the treatment has a negative effect on their grade (at a significance level of 10%), whereas for students from software engineering, there is a positive and significant effect on their grade. For the other specialisations, we do not find anything. These heterogenous findings seem a plausible explanation as to why we find a coefficient of almost zero in the overall effect analysis on grade.

Further heterogenous analyses based on this variable, on passing the course, provide a similar picture as shown in Table 7. Another analysis, where we split the sample by whether the previous educational level is related to programming (regardless of the level of previous education) or not (potentially signalling a better preparation for a course such as programming), does not reveal any significant results (results for further heterogenous analyses are available upon request). Splitting up by previous education or gender leads to too small subsamples and thereby unreliable results. However, if we split the sample by whether the student was born in the Netherlands, ($n = 52$), which might indicate a higher motivation, we find highly significant positive results for the students that were not born in the Netherlands (both for grades and for whether the student passed the exam the first time) and no significant results for students that were born in the Netherlands ($n = 404$). Students that were born in other countries also more often choose the software engineering specialisation and less often game development. Unfortunately, due to too small subsample size, we cannot check which one of the two characteristics is driving our results.

Conclusion and discussion

In this study, we analyse the effects of a randomised controlled experiment on providing freshman students in a Java programming course with a Learning Analytics Dashboard, with weekly emails with a personalised link, aimed to signal students about their study progress as well as about their chances of success, by showing visual performance indicators of the student and the average of the cohort. The results show that the weekly email and corresponding dashboard access (ITT), as well as dashboard use (2SLS), have positive effects on student performance in the online environment during the course (standardised effect sizes of 0.2 and 0.25), but not on student behaviour in the online environment or on student performance in the final exam or their chances of passing the programming course. However, heterogenous

Table 7 Heterogenous effects by specialisation (only 2SLS shown) (standardised coefficients)

Outcome = grade	Business IT management (BIM)	Game development (GD)	Software engineering (SE)	System and network engineering (SNE)	Technical computing (TC)
Dashboard use	0.25 (0.34)	- 0.42* (0.22)	0.24* (0.14)	- 0.16 (0.18)	- 0.08 (0.28)
Covariates	Yes				
N	74	110	162	63	47
F-stat	1.59	2.07	2.73	1.80	0.55
R ²	0.30	0.30	0.30	0.42	0.22

Standard errors between brackets. Covariates are previous educational level and class

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

analysis shows that final exam grades are significantly higher for treatment students in the specialisation software engineering and are significantly lower for treatment students in game development (the two largest specialisations). We also discuss key challenges and possible solutions when using learning analytics dashboards in combination with prediction models that are linked to online student behaviour

The fact that we do not find an overall result on student performance is in line with findings from previous studies (e.g. Dodge et al. 2015; Park and Jo 2015). However, we do find heterogenous effects by specialisation, and we also find positive effects for intermediate results such as the quizzes and the online mastery exercises. Possible explanations from the literature why a dashboard would not work, such as the competition element that could lead to lower motivation or the ‘one size fits all’ approach that could lead to ineffective use (Teasley 2017), therefore do not seem fit for our study. In our case, the lack of effects on overall student performance is more likely to be due to too low online activity by the students and due to differences between specialisations.

An important aspect of the dashboard treatment is the effect of the LAD on the use of the online environments by the students. And although our analysis shows that there is a positive effect of the dashboard of the use of and performance in the online environments by students, the online activity strongly declined over time and dropped from about 70% in week 1 to less than 50% in week 3. This was most likely due to the summative nature of the online activities. After 2 weeks, (some) students probably figured out that there were no sanctions against poor online performance and they stopped using Moodle and MyProgramming Lab or only used them infrequently. Apparently, also the dashboard could not encourage students (enough) to do their online exercises. Maybe this is due to the business tone of the dashboard or to the inaccurate predictions of the dashboard. The dashboard should have stimulated the students more to do their online exercises, and maybe this can be better achieved by only showing the real facts like the online progress and not present the predicted values of their grade anymore, which would probably increase their faith in the dashboard. It is important to realise that the prediction models would most likely have been much more accurate if the online assignments had a summative nature, which could have led to a different effect of the dashboard.

It is very likely that students realised that they had spent too little time in the online environment (where the dashboard group spent a little more time online than the treatment group, but still not much) and therefore overcompensated with additional study hours as preparation for the exam. This may explain why we do not see an effect on the overall performance, because the effect would have to run through online activity, which was too low, and for which students may have compensated with additional study hours for the exam.

The heterogenous effects that we find may point towards something like a motivation effect. The general content of software engineering is much more related to programming than the general content of game development, so SE-students may be more motivated for the programming course, either because it very much relates to the study of their choice or because they see the benefit of this course more for their future student life (although it is beneficial for all students, as we argue and describe in the introduction, it is unlikely that freshman students of more unrelated specialisations also see it that way). The same might hold for country of birth: students that were not born in the Netherlands might be more motivated to succeed in their studies.

Future work should focus more on these heterogenous effects, as well as seriously consider the formative nature of these online exercises, as both these aspects have most likely influenced the results we present in this paper.

Data availability This paper uses confidential student data from the University of Applied Sciences that we study and can therefore not made publicly available without violating the European General Data Protection Regulations (GDPR).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Azmat, G., & Iriberry, N. (2010). The importance of relative performance feedback information : evidence from a natural experiment using high school students ☆., *94*, 435–452. <https://doi.org/10.1016/j.jpubecon.2010.04.001>.
- Bandiera, O., Larcinese, V., & Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics*, *34*, 13–25.
- Beattie, G., Laliberté, J. W. P., & Oreopoulos, P. (2018). Thrivers and divers: using non-academic measures to predict college success and failure. *Economics of Education Review*, *62*(January 2017), 170–182. <https://doi.org/10.1016/j.econedurev.2017.09.008>.
- Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, 309–318. <https://doi.org/10.1145/3027385.3027403>.
- Bos, N. (2016). *Effectiveness of blended learning*. Heerlen: Open University, Heerlen.
- Chevalier, A., Luehrmann, M., & Dolton, P. (2014). "Making it count": evidence from a field experiment on assessment rules, study incentives and student performance. *IZA Discussion Papers, No. 8582* (Vol. C07). Bonn.
- Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: a test of stereotype threat and social comparison explanations for women's math performance. *Computers and Education*, *78*, 66–77. <https://doi.org/10.1016/j.compedu.2014.05.005>.
- Dodge, B., Whitmer, J., & Frazee, J. P. (2015). Improving undergraduate student achievement in large blended courses through data-driven interventions. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 412–413. <https://doi.org/10.1145/2723576.2723657>.
- Durall, E., & Gros, B. (2014). Learning analytics as a metacognitive tool. In *CSEDU (1)* (pp. 380–384).
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers and Education*, *80*, 152–161. <https://doi.org/10.1016/j.compedu.2014.08.019>.
- Hu, Y.-H., Lo, C.-L., & Shih, S.-P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behaviour*, *36*(0), 469–478. <https://doi.org/10.1016/j.chb.2014.04.002>.
- Jivet, I., Scheffél, M., Drachler, H., & Specht, M. (2017). Awareness is not enough. Pitfalls of learning analytics dashboards in the educational practice. In *Data driven approaches in digital education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Tallinn, Estonia, September 12–15, 2017, Proceedings* (p. 14). <https://doi.org/10.1007/978-3-319-66610-5>.
- Kim, J., Jo, I. H., & Park, Y. (2015). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, 13–24. <https://doi.org/10.1007/s12564-015-9403-8>.
- Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-regulation and ability predictors of academic success during college. *Journal of Advanced Academics*, *20*(1), 42–68.
- Lam, S.-F., Yim, P.-S., Law, J. S. F., & Cheung, R. W. Y. (2004). The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology*, *74*(2), 281–296.
- Lauría, E. J. M., Moody, E. W., Jayaprakash, S. M., Jonnalagadda, N., & Baron, J. D. (2013). Open academic analytics initiative: initial research findings. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK'13)*, 150–154. <https://doi.org/10.1145/2460296.2460325>.

- Ministry of Education Culture and Science. (2016). 123studiekeuze. Retrieved June 16, 2016, from <https://student.hva.nl/hbo-ict/a-z/item/propedeuse-hbo-ict.html?f=propedeuse>
- Mumane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. Oxford: Oxford University Press.
- Nuffic. (2015). The Dutch education system described. Retrieved March 10, 2017, from <https://www.epnuffic.nl/en/publications/find-a-publication/education-system-the-netherlands.pdf>
- Oreopoulos, P., & Petronijevic, U. (2018). Student coaching: how far can Technology go? *Journal of Human Resources*, 53(2), 299–329.
- Park, Y., & Jo, I. (2015). *Journal of Universal Computer Science Support Students' Learning Performance*, 21(JANUARY 2015), 110–133.
- Rojas, R. (2009). AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boosting. *Freie Universitat Berlin*. Retrieved from http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/adaboost4.pdf. Accessed 9 Jun 2020.
- Symonds, William, C., Schwartz, R., & Ferguson, R. F. (2011). *Pathways prosperity meeting the challenge of preparing young Americans for the 21st century*. Harvard University Graduate School of Education.
- Teasley, S. D. (2017). Student facing dashboards: one size fits all? *Technology, Knowledge and Learning*, 1–8.
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: learning analytics in a data-rich context. *Computers in Human Behaviour*, 47, 157–167. <https://doi.org/10.1016/j.chb.2014.05.038>.
- Tufte, E. R. (1991). Envisioning information. *Optometry & Vision Science*, 68(4), 322–324.
- Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Journal of Educational Technology & Society*, 15, 133–148.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioural Scientist*, 57(10), 1500–1509. <https://doi.org/10.1177/0002764213479363>.
- Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6), 1499–1514. <https://doi.org/10.1007/s00779-013-0751-2>.
- Wayne, I., & Langley, P. (1992). Induction of one-level decision trees. In *9th international conference on Machine Learning*.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.