Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS

Rianne Conijn, Chris Snijders, Ad Kleingeld, and Uwe Matzat

Abstract—With the adoption of Learning Management Systems (LMSs) in educational institutions, a lot of data has become available describing students' online behavior. Many researchers have used these data to predict student performance. This has led to a rather diverse set of findings, possibly related to the diversity in courses and predictor variables extracted from the LMS, which makes it hard to draw general conclusions about the mechanisms underlying student performance. We first provide an overview of the theoretical arguments used in learning analytics research and the typical predictors that have been used in recent studies. We then analyze 17 blended courses with 4,989 students in a single institution using Moodle LMS, in which we predict student performance from LMS predictor variables as used in the literature and from in-between assessment grades, using both multi-level and standard regressions. Our analyses show that the results of predictive modeling, notwithstanding the fact that they are collected within a single institution, strongly vary across courses. Thus, the portability of the prediction models across courses is low. In addition, we show that for the purpose of early intervention or when in-between assessment grades are taken into account, LMS data are of little (additional) value. We outline the implications of our findings and emphasize the need to include more specific theoretical argumentation and additional data sources other than just the LMS data.

Index Terms—Learning analytics, learning management systems, portability, predictive modeling, student performance

1 Introduction

The emergence of information and communications technologies (ICT) into higher education has significantly changed the way in which teachers teach and students learn. Using the internet to provide content has opened up the possibility to transform face-to-face courses into courses in which a significant amount of (blended courses) or all information (online courses) is delivered and accessible online [1]. A vast majority of institutions use the internet in teaching, often through Learning Management Systems (LMSs), also known as Virtual Learning Environments (VLEs) [2]. LMSs can support student learning by providing content online, and by allowing for additional components such as quizzes, presentations and screencasts, assignments, and forums [3]. Additionally, LMSs allow teachers to provide and manage these resources in a relatively easy and integrated way.

As every action in an LMS is monitored and stored, insight can be gained into students' online behavior, which in turn can be used to improve learning and teaching. The analysis of LMS data is often referred to as learning analytics [4], defined as "the measurement, collection, analysis and reporting of data about learners and their context, for purposes of understanding and optimizing learning and the environments in which it occurs" [5]. Much research in the field of learning analytics has used LMS data for predictive modeling of student performance to predict students' grades and to predict which students are at risk of failing a course [6], [7], [8]. This is an important step in learning analytics, as it informs the implementation of interventions, such as personalized feedback.

Studies predicting student success in offline education have typically collected measurements using validated questionnaires, interviews, and observational techniques, with relevant theoretical concepts in mind so that the measurement can be geared towards the concepts that the researcher thinks need to be measured. The use of LMSs allows for tracing and analyzing students' online behavior without the necessity of time-consuming data-collection. However, LMSs provide raw log data that are not concrete measurements of previously outlined theoretical concepts. It is therefore important to understand whether and how these data can be used for learning analytics. Recent studies show a wide variety in the analytical usage of LMS data: different kinds of analytical methods and predictors are used, often without explicit mention of the theoretical argumentation behind them [9]. Moreover, many studies analyze LMS data of one or a few institutions, for one or only a few courses, or describe special cases (e.g., courses using tailor-made e-tutorial packages). This makes it hard to compare

- *R. Conijn, C. Snijders, and U. Matzat are with the Department of Industrial Engineering & Innovation Sciences, Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, MB NL-5600, Netherlands. E-mail: {m.a.conijn, c.c.p.snijders, u.matzat}@tue.nl.*
- *A. Kleingeld is with the Industrial Engineering & Innovation Sciences, Human Performance Management Group, Eindhoven University of Technology, Eindhoven, MB NL-5600, Netherlands. E-mail: p.a.m.kleingeld@tue.nl.*

Manuscript received 1 Feb. 2016; revised 16 Sept. 2016; accepted 27 Sept. 2016. Date of publication 12 Oct. 2016; date of current version 16 Mar. 2017. Recommended for acceptance by D. Gasevic, C. Rose, and G. Siemens. 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.2016.2616312

the different studies and draw general conclusions about the ways in which to use LMS data for predictive modeling.

Moreover, the question is whether there actually is a single best way to predict student performance across a diverse set of courses. Studies that have used similar methods and predictors have nonetheless found different results in the correlational analyses and prediction models. Even within one institution using the same LMS, differences have been found in the prediction models of nine blended courses [10]. Thus, the effects of LMS behavior on student performance might differ per institution or even per course. Indeed, a study using 29 courses (204 offerings, 352 unique students), has found that the variance in students' performance (final grade), was accounted for by individual differences (18 percent) as well as course offerings (22 percent) [11]. Hence, the so-called "portability" of prediction models across courses might not be that high, even though it might still be that prediction models can be successfully used in single courses.

In addition, most studies focus on predicting student performance after a course has finished, establishing how well student performance could have been predicted with LMS usage data, but at a point in time where the findings cannot be used for timely intervention anymore [12]. As LMS data provide information during the whole course, it seems useful to determine whether data from only the first weeks of a course are enough for accurate prediction of student performance [13].

In the current study, we add to the analysis of the portability of prediction models and the accuracy of timely prediction. First, we provide an overview of the theoretical arguments used in learning analytics and the predictors that have been used in recent studies. The predictive value of these predictors will be examined in 17 blended, undergraduate courses taught at the same institution (Eindhoven University of Technology). This allows us to establish effects of different types and degrees of LMS usage while controlling at least to some extent for contextual effects. Furthermore, the portability of the prediction models across the 17 courses, i.e., the effect of course, is analyzed. For this we replicate the study of Gašević et al. [10] within another institution with a larger sample of more similar courses. Moreover, to ensure comparability of findings, we only use predictors that are available for all courses (cf. [10]). In addition, we analyze whether it is possible to identify students at-risk early on in a course, and to what extent these models can be used to generate targeted interventions.

2 Grounding Learning Analytics in Theory

Most studies on learning analytics are largely data-driven and not explicitly based on theory [14]. However, the extensive literature on offline and online learning can be used to better ground learning analytics in theoretical arguments [15]. This would for instance provide better motivation for methodological choices, such as which predictor variables should be extracted and created from the raw LMS log data, or how analytical results could be interpreted. Thus far, few studies on learning analytics have explicitly connected theoretical arguments to the selection of prediction variables [9]. For example, the interaction theory of Moore [16] has been used to extract variables from the LMS in e.g., [11], [17]. Others have used the theory of self-regulated learning [10]. Recently, this issue has received more attention, partly due to a special section in the Journal of Learning Analytics in 2015 [18], with several papers focusing on the grounding of learning analytics in theoretical argumentation. Others grounded their study in the constructivist theory of learning and self-regulated learning [10].

Constructivist theorists argue that learning is a process of actively constructing rather than acquiring knowledge [19]. Additionally, instruction is a process of supporting this construction rather than just communicating it. LMSs can enhance knowledge construction, as they facilitate flexible course delivery and integrate multiple learning resources. This results in greater flexibility and control over the learning process [20], [21], which can also support self-regulated learning. Self-regulated learners first prepare their learning by clarifying a task, generating goals, and adopting a plan for reaching those goals. They then carry out their plan and construct new information [22]. These two stages are affected by internal and external factors, also called task conditions and cognitive conditions, respectively [23]. Task conditions include resources, instructional cues, available time, and social context. Cognitive conditions consist of beliefs, dispositions, motivational factors, domain knowledge, task knowledge, and knowledge about study tactics and strategies.

These theories are useful to explain differences in students' behavior between individuals and tasks (or courses), which lead to different performances across students and courses. However, the match between these theories and the measurements used for learning analytics is not optimal and other theories such as situated learning [24] or connectivism may be necessary as well. Therefore, in the current study we also consider past research on predicting student performance in traditional offline, blended, and fully online courses to guide our design.

3 Previous Research

3.1 Studies Predicting Student Performance

Most studies on learning analytics focus on predicting student performance, often quantified by final grade or by whether the student has passed a course or not. Data used for predictive modeling can come from different sources such as student characteristics, including their dispositions and demographics, but in recent years most often data from LMSs have been used [13], [25]. Studies analyzing LMS data show a wide variety in types of LMS used, courses examined (blended or fully online), and predictive analytical techniques that have been used. Most studies analyze few courses and the choice of predictor variables varies considerably across studies, which makes it hard to compare the different studies and draw general conclusions about the best and most stable predictors of student performance.

Rafaeli and Ravid [26] were among the first to use LMS data for learning analytics. They evaluated the implementation of an LMS, based on the usage of the online environment and performance in the course. Data from 178 students in three blended classes were analyzed. Students who were inexperienced in using online systems tended to stick to a page-by-page reading order, whereas more experienced students adopted a much more non-linear style. Linear regression analysis showed that 22 percent of the variance in final grades could be explained by the amount of pages read and the grades for online questions posed during the course. Likewise, Morris, Finnegan, and Wu [27] found that the number of content pages viewed was a significant predictor in three fully online courses in eCore with 354 students. Contrary to Rafaeli and Ravid [26] they used a total of eight duration and frequency variables, and no in-between measurements of performance. Multiple regression analyses with these predictors on final grade of the 284 completers showed that 31 percent of the variance in final grade was accounted for by the number of discussion posts and content pages viewed, and the time spent on viewing discussion posts. Moreover, they found that withdrawers had a significantly lower frequency of activities and less time spent online, compared to completers. Macfadyen and Dawson [28] also found that the amount of links and files viewed had a positive correlation with final grade. However, these variables did not turn out to be significant predictors in their final model. As in [27], a fully online course was analyzed, but using another LMS: Blackboard. Multiple regression analyses showed that 33 percent of the variance in final grade of 118 completers could be explained by the total number of discussion messages posted, mail messages sent, and assessments completed. Classification resulted in an accuracy of 74 percent, where 38 out of 63 students who failed were accurately predicted as at risk, and 49 out of 65 successful students could be accurately predicted as not at risk.

Discussion forum posts was the only predictor in both final prediction models of [27] and [28]. The usage of the discussion forum was important for predicting student performance in several other studies as well. In an analysis of discussion forum usage in Blackboard in a course of 1,026 students, Dawson et al. [29] found a significant effect of discussion forum usage on final grade. Students who made at least one post in the forum scored 8 percent higher on average than students who had posted nothing at all. However, Nandi et al. [30] did not obtain a significant effect of forum usage on student performance, with data from 645 students using Blackboard in two courses. They did find a trend that high-achieving students participated more in the forum than other students. However, only 40 percent of the students participated, indicating that the forum may be a more useful predictor when it is used by a high proportion of the students.

Other researchers using Moodle LMS in blended courses found that discussion posts and the amount of communication between peers were also significantly correlated with final grade. Yu and Jo [31] analyzed data of 84 students and tested six variables: total log in frequency, total time online, regularity of study interval, number of downloads, amount of communication with peers, and amount of communication with the instructor. Total time online and interaction with peers correlated significantly with final grade, and all predictor variables combined accounted for 34 percent of the variance in final grade. Using the same LMS with 134 students in one course, Zacharis [32] analyzed 29 variables, 14 of which correlated significantly with final grade. Total time online and the amount of files and links viewed significantly correlated with final grade, but as in Macfadyen and Dawson [28], these were not retained in the final model for predicting student performance. Only three predictors were included, explaining 52 percent of the variance in the final grades: number of files viewed, and two broader variables measuring various interactions and contributions to content. Classification resulted in an overall accuracy of 81 percent: 30 out of 43 students who failed were predicted correctly, and 79 out of 91 students who passed.

The overview above reveals a wide variety in the studies on LMS data. Especially the predictor variables that are being used show a great diversity, which can be explained by the fact that not all researchers have access to all variables in the LMS. Also, different courses and institutions may use different tools in the LMS. This incomplete availability and access to data may also explain why these studies are largely data-driven [14]. One exception is the usage of the classification of communication or interactions in learning processes to extract variables from the LMS log data. Moore [16] distinguished three different types of interactions: student-student, student-teacher, and student-content interactions.

Agudo-Peregrina et al. [17] have used these types of interaction in their model for analyzing Moodle data. The LMS data were classified based on types of interaction, frequency of use, and passive or active participation. The classification was tested on data of two blended and six fully online courses, with 20 to 30 students per course. Student-student and student-teacher interactions were found to be significant positive predictors of final course grades in the fully online courses, but not in the blended courses. In both the blended and fully online courses, student-system and student-content interactions were not found to be significant predictors.

Joksimović et al. [11] found that the amount of student-student interactions had a significant effect in core and elective courses, but not in prerequisite courses for entering a masters' program. Time spent on student-course interactions was in fact negatively related with final course grade in the core courses. Thus, even when theoretical arguments provide guidance and lead to similar kinds of predictors across studies, inconsistencies in terms of the effect sizes and the direction of the effects have nevertheless been found. In fact, it was shown in [33] that merely employing different calculations of the predictors may have strong effects on the findings. Using ten courses in Moodle and 15 different methods of estimating time on task, resulted in substantial differences in the prediction models of the same courses.

3.1.1 The Inconsistency of Findings

Although it is tempting to argue that differences in the findings can be explained by the different predictors being used, even when similar predictor variables have been used, the results are not always robust. Despite the different predictor variables used, a reasonable amount of variance in final grade could be explained, ranging from 22 percent [26], 31 percent [27], 33 percent [28], 34 percent [31], to even 52 percent [32]. However, recently Tempelaar et al. [13] found that the amount of clicks in Blackboard of 873 students in two blended courses could only explain a marginal 4 percent of the variance in final exam performance.

Another explanation for the different outcomes could be that studies describe specific cases (such as courses using tailor-made e-tutorial packages), so that it is unclear whether the outcomes apply in general, or to a specific institution, course, or group of students only. One exception is a large-scale study on 2,674 Blackboard courses with 91,284 students and 40 Moodle courses with 1,515 students [34]. A positive correlation was found between number of clicks and final grade in both LMSs. Only this single correlational analysis was conducted, so no general conclusions could be drawn on other prediction variables.

It therefore remains unclear how, in general, LMS data can be used for predictive modeling. Actually, the results bring up the question whether there is one general set of variables for predicting student performance: can the same models be used in multiple courses and institutions, or are online courses (and perhaps also students and institutions) so diverse that they each need their own prediction model? This issue is often referred to as the portability of the prediction models [10], [35]. Given the current results in the literature, we feel that research into the portability of prediction models is a crucial next step.

3.1.2 Portability of Models Predicting Student Performance

The issue of the portability of the prediction models has been recognized at least from 2011, when the Open Academic Analytics Initiative (OAAI) was initiated. The objective of the OAAI is to advance the field of learning analytics by exploring the challenges in scaling learning analytics across all higher education institutions [35]. The first two sub goals of this initiative specifically focus on the scaling of prediction models and on developing an open source model for predicting student success [36]. Lauría et al. [36] have tested the portability of a prediction model for final grade between two institutions: Purdue University and Marist College ($n = 18{,}968$ and $n = 27{,}276$, respectively). Although these institutions differ in institutional type, size, approaches, and type of LMS (Blackboard versus Sakai), correlational analysis and prediction models for final grade revealed similarities [36]. All variables analyzed (number of sessions, number of content views, number of discussions read, number of discussions posted, number of assignments submitted, and number of assessments submitted) correlated significantly with final grade in both institutions and had a similar effect size. A follow-up study found that the prediction model used at Marist College for classifying students as pass or fail, had a 10 percent lower accuracy when applied to data from three partner institutions [35]. The authors argued that the portability of prediction models for student performance might be higher than expected.

However, Gašević et al. [10] found that the portability across courses in an Australian university was not that high at all. They compared prediction models of nine first-year courses with a total of 4,134 students. The predictor variables consisted of the number of actions in the different modules in Moodle, with courses differing in the modules that they used. The analysis controlled for student and program characteristics: age, gender, domestic versus international student, full versus part time program, and first course versus later course. The multiple linear regression models for all courses separately differed from each other and from the generalized model that included all courses. The authors argued that analyzing the whole sample may underestimate or overestimate the effects of the predictors, and that it may not be a good idea to use a single model for multiple courses, thus questioning the portability of the models between courses.

These contradicting results show that there is a need for further studies that enlarge the empirical base of the issue of portability of prediction models. The differences in the prediction models of [10] may be explained by the fact that the courses differed in type and learning design. Different modules were used in the courses, which resulted in different predictors. Moreover, different learning designs associated with different available activities in an LMS have been found to result in a difference in LMS usage [37]. For example, courses focusing primarily on assimilative activities, such as reading course specific information or watching lectures, were associated with lower course completion rates [37]. Hence, in the current study we compare the portability of the prediction models of 17 courses in Moodle LMS with a similar learning design. As the courses are all from a technical university, they are quite similar in type. Moreover, similar modules are used in Moodle and a set of predictor variables is used that is available in all courses.

3.2 Studies Investigating Early Predictors of Study Performance

Most studies that have tried to predict student performance analyzed the behavior of students in the LMS during the entire course. This allows for inferring study success from LMS data, but at a point in time where interventions are no longer meaningful [12]. Some researchers have acknowledged this issue and decided to analyze potentially predictive data from early stages in a course.

For instance, Milne et al. [38] have analyzed LMS data of the first week of a course for 658 students in 9 blended courses. They found that LMS usage in the first week of the course was significantly higher for successful students than for students who failed the course. Hu et al. [39] predicted student performance of 300 students at three points in time during a course. In total 14 LMS variables were grouped for the first four, eight, and thirteen weeks of the course. They found that prediction accuracy increased as the course progressed.

Schell et al. [40] have also found that prediction accuracy increases over time. Their results showed that 29 percent of the variance in final grades of 89 students in a blended course was explained by the entry test. Explained variance increased to 34 percent when self-efficacy was included. The addition of midterm grades over time led to a substantial increase in prediction (partly because midterm scores were a significant part of students' final grades), and to a decrease in the predictive power of self-efficacy. Tempelaar et al. [13] also found that the prediction accuracy increases over time and that performance data are especially important. The number of clicks in the week before the course had started was found to have the highest predictive power. As the course progressed, the prediction of student performance gradually improved. Data from interim assessments were shown to be the best predictor. Indeed, a notable improvement in predictive power was found in the week where the first assessment data became available. The authors therefore argued that the best time to predict student performance is as soon as possible after the first assessment, as this would be the best compromise between early feedback and sufficient predictive power.

TABLE 1
Course Characteristics

Course name	Quarter	Level (year)	Type	F2F hours per week	Clicks per student	Online activities (% of clicks): Content	Forum	Quiz	Assignment	Peer-review assignment	Wiki	Assessment weights: Entry test	Homework online	Homework offline	Midterm	Final exam	N
1 Calculus A	1	1	Basic	4.5	889	2.9%	.4%	80%				10%	10%		10%	70%	438
2 Calculus B	1	1	Basic	5.3	1164	.6%	.5%	85%				10%	10%		10%	70%	1121
3 Calculus C	1	1	Basic	5.3	742	.9%	.5%	75%				10%	10%		10%	70%	227
4 Calculus pre M Architecture	1	Pre M	Basic	3.0	815	1.8%	.0%	94%				10%	10%			80%	135
5 Set theory and Algebra	1	1	Mathematics	6.0	587	8.3%	.1%	71%						15%	15%	70%	73
6 Linear Algebra and Vector Calculus	2	2	Mathematics	6.0	673	1.0%	.1%	90%						10%	30%	60%	120
7 Linear Algebra	1	Pre M	Mathematics	4.5	279		.2%	89%								100%	76
8 Experimental Physics 1	1	1	Physics	5.3	302	4.1%	.2%	77%							40%	60%	168
9 Experimental Physics 2	2	1	Physics	6.0	94	4.7%	.0%	75%							40%	60%	155
10 Behavioural Research Methods	2	2	Psychology	4.5	620	14.1%	3.1%	58%			5%			30%		70%	136
11 Applied Physical Sciences formal	2	1	Basic	6.0	234	1.4%	.1%	79%					10%		20%	70%	836
12 Applied Physical Sciences conceptual	2	1	Basic	6.0	227	1.1%	.1%	81%					10%		20%	70%	822
13 Condensed Matter	2	3	Physics	3.0	189	4.1%	.1%	78%							30%	70%	74
14 Intro to Psychology & Technology	1	1	Psychology	4.5	189	13.2%	.2%	47%	6%				10%	20%	20%	50%	154
15 Linear Algebra 1	1	1	Mathematics	6.0	61		.5%	29%		30%				15%	15%	70%	66
16 Statistics	2	2	Mathematics	6.0	164		.0%	89%						15%	15%	70%	326
17 The Effectiveness of Mathematics	2	1	Mathematics	6.0	198	18.5%	.1%		37%					50%		50%	62

In line with these studies, we also analyze how the prediction accuracy of student performance improves as the course progresses. Contrary to Tempelaar et al. [13], we consider a less specific online environment (largely using standard modules in Moodle LMS) with a more homogeneous set of students, with the hope to be able to draw more general conclusions this way. LMS data and assessment data are used to examine how the prediction changes over time, whether using only LMS data may be of use for timely intervention, and how the effectiveness of predictions changes after the assessment data has become available.

4 Method

4.1 Participants

Data about students' online behavior were collected from blended courses using Moodle LMS taught in the first two quarters (fall and winter) of cohort 2014-2015 at Eindhoven University of Technology (TU/e). In this period, 42 courses used Moodle LMS. These were mostly courses for first year students. Data were used from courses in which at least 75 students had participated, resulting in a sample of 17 courses. Students who did not take the final exam, or who did not take the final exam for the first time directly after the lecture period, were excluded from the analyses. In total 1,072 students were excluded ($M = 63$, $SD = 103$ per course). The final sample included 4,989 students in 17 courses, ranging from 62 to 1,121 students per course ($M = 293$, $SD = 324$). As students could enroll for multiple courses, the sample consisted of 2,913 unique students; 1,445 students who were enrolled in one course, 1,121 students in two courses, 143 students in three courses, 147 in four courses, and 57 in five courses. An overview of the courses used and the instructional conditions and course designs can be found in Table 1.

Of the 17 courses, nine courses were taught in the fall quarter from September 1st to November 9th, 2014, and eight courses were taught in the winter quarter from the 10th of November, 2014 to the 1st of February, 2015. All courses consisted of eight weeks of course work and two weeks of final exams. LMS log data was collected over these ten weeks, as well as the week before the start of the lectures (week 0), and was grouped per week. The winter quarter included data from the two-week Christmas break, as this was part of the lecture period, bringing the total to 13 weeks of LMS data (week 0, 8 lecture weeks, 2 break weeks, and 2 exam weeks).

Fifteen courses were undergraduate courses, and two were prerequisite courses for entering the graduate programs (pre M). Most of the undergraduate courses were first-year courses, but also three second-year and one third-year course were included. The courses varied from basic courses that every undergraduate student at TU/e has to take, to specific courses in the fields of mathematics, physics, statistics, and psychology. The courses were blended, as next to three to six hours of face-to-face lectures and instructions a significant amount of information was accessible online via Moodle. According to the classification of blended courses made by [41], most of the courses can be classified as sharing or submission courses as Moodle was mostly used for sharing resources and providing quizzes and assignments. One course, course 15, can also be considered a communication or collaboration course, as the course included peer-reviewed assignments. Course 10 could also be considered a delivery or discussion course, as in this course used a wiki and used the discussion forum more extensively. However, apart from these two courses, the courses were quite similar in how they implemented blended learning.

Next to LMS data, assessment data were collected, which consisted of in-between assessment grades, the final exam grade, and the overall course grade. All grades are on a 0 to 10 scale, where grades < 5.5 imply a student does not pass a course and grades ≥ 5.5 represent a pass. The weight and the type of the assessments differed per course. In-between assessments included midterms, quizzes, reports,

TABLE 2
Predictor Variables Used for Prediction

Predictor	Used in	N	M	SD
Total number of clicks	[13], [17], [34], [37], [42]	4,989	605	630
Number of online sessions	[10], [18], [29], [31], [32], [33], [35], [36], [39], [43], [44]	4,989	30.3	21.2
Total time online (min)	[13], [28], [31], [32], [37], [39]	4,989	815	678
Number of course page views	[32], [33]	4,989	208	144
Irregularity of study time	[31], [42], [43]	4,989	1,926	993
Irregularity of study interval		4,989	309,000	252,000
Largest period of inactivity (min)		4,989	20,500	13,100
Time until first activity (min)		4,989	17,167	11,250
Average time per session (min)	[29], [39]	4,989	27.2	910
Number of resources viewed	[10], [32], [33]	2,277	7.38	14.2
Numbers of links viewed	[28], [32]	4,037	7.81	17.0
Number of content page views	[10], [26], [27], [33], [35], [36], [42]	4,989	84.6	80.5
Number of discussion posts views	[27], [28], [30], [32], [33], [36], [45]	4,989	1.92	7.36
Total number of discussion posts	[18], [28], [30], [32], [36], [43]	2,831	0.06	0.46
Number of quizzes started	[28], [32], [33], [36]	2,256	22.4	11.9
Number of attempts per quiz		4,927	1.02	0.28
Number of quizzes passed	[28], [32]	4,927	9.43	6.91
Number of quiz views	[32], [33]	4,927	110	80.3
Number of assignments submitted	[28], [32], [33], [36], [39]	774	1.95	1.36
Number of assign. (submission) views	[28], [32], [33]	2,665	3.79	7.76
Number of wiki edits	[32]	136	68.0	61.7
Number of wiki views	[32]	4,989	0.38	1.04
Average assessment grade	[13], [26], [35], [40]	4,913	6.78	1.96

assignments, and homework. Some of these assessments were online and logged in the Moodle LMS, while other assessments were offline and handed-in on paper or via other systems. Contrary to most previous work, final exam grade was used as the outcome variable instead of final course grade, as in-between assessments are part of the final course grade in 16 of the 17 courses. A binary outcome variable was computed with grade ≥ 5.5 coded as pass (1), and grade < 5.5 as fail (0).

4.2 Data Pre-Processing

The raw Moodle log data were pre-processed using R to create predictor variables. We used predictor variables that were available in the current data set from the Moodle LMS and that have been found to be relatively robust predictors in previous (also offline) research, or have been used in previous studies analyzing LMS data. An overview of all predictor variables and some descriptive statistics are shown in Table 2.

Past performance, entry tests and in-between assessments have been shown to be among the most robust and important predictors of student success [15], [42]. Therefore, we used in-between assessment data (available for 16 courses) as a predictor variable, and expect that it has a high predictive value for the final exam grade. The amount, weight, and type of in-between assessments differed among the sixteen courses (see Table 1); hence for tractability the average grade of all in-between assessments per course was used. As most in-between assessments took place in week 4 or 5, we have analyzed the data assuming that grades would be available at the end of week 5. Because past performance does not explain all variability in student performance [47] and is not always available, we also included 22 variables extracted from the LMS log data.

Four basic predictors per course were extracted from the log data, based on the total online participation. When a student spends more time online or clicks more this could be associated with a higher grade, as this is expected to be an indicator of higher motivation (i.e., effort, persistence). Motivation has been shown to be a robust predictor for student performance [48], [49]. Moreover, face-to-face class attendance has been shown to be significantly positively correlated with exam scores [42], [50], [51]. Hence, we propose that 'online attendance' may also be related to higher grades. We therefore used the following basic predictors: the total amount of clicks, the number of online sessions, the total time online, and the total amount of views. These are some of the most frequently used predictors in the literature. A session was defined similarly as in Zacharis [32], as the sequence of behavior from the first click after the login to Moodle until the last click before logging out, or the last click before staying inactive for at least 40 minutes. Each session consisted of at least two clicks. The time between the first and the last click of a session was used to compute the total time online. As the strategy used to calculate the time online can influence the final prediction model [33], also count measures (such as the total amount of clicks) were calculated to limit the effect of the current strategy on the final prediction models.

Next to these basic predictors, more complex predictors related to study patterns were included, as previous studies have shown that the regularity of studying also influences the final grade. For example, a meta-analysis has shown that procrastination has a moderate negative effect on performance, and time management has a moderate positive effect [15]. Similarly, a small case study ($n = 46$) found regularly and consistently accessing online resources resulted in a higher performance compared with frequent access at a late stage [52]. Despite this, previous studies only included regularity of study time (*SD* of time per session) and average time per session to predict student performance. Therefore, we included more study pattern predictors: the irregularity of study interval (*SD* of time between sessions), the largest period of inactivity, and the time until first activity.

Finally, several other variables were computed that relate to the usage of different modules available in Moodle. These variables were not available for all courses, as not all courses used the same modules in Moodle. If a module was not used in a course, the values were coded as missings to exclude them from analyses. Predictors were included from the Moodle modules: resource, URL, page, forum, quiz, scorm, assignment, workshop, and wiki. Some modules that were used in previous literature, such as mail [28], [32], chat [10], [28], [32], [33], or blog [32] were not used in any of the courses and therefore excluded. Some features within modules that were used in previous literature, such as upload photo to profile [43], use of 'map' tool [10], [33], or announcements [28] were not or only rarely used in our courses and therefore excluded as well.

First, predictors were included about the use of content. This includes the amount of additional resources viewed (available in eight courses), the amount of external links clicked (available in ten courses), and the amount of content pages viewed (available in all courses).

Second, predictors from the forum and wiki (available in all courses) were used. As the forum was only rarely used (see Tables 1 and 2), only the amounts of forum posts and forum views were included and we did not further categorize the posts, e.g., into interactions as done in [11], [17], [53]. For the wiki module the amount of wiki views and edits were used. Participation in the forum or wiki, especially active participation such as editing and posting, is thought to be related to more active learning, which requires more effort and results in higher performance [54].

Third, the amount of quizzes started, the average amount of attempts per quiz, the amount of quizzes passed, and the amount of quizzes viewed were extracted (available in 16 courses) as these generate feedback, with potentially strong positive effects on performance [55], [56]. For the quizzes, data from the Moodle modules "quiz" and "scorm" were combined, as both can provide quizzes; the former in Moodle itself, the latter through an external source whose output has been integrated into Moodle.

Lastly, the amount of assignments submitted and assignment submission views were extracted (available in eight courses). For the assignments, the Moodle modules "assignment" and "workshop" were combined, as both provide the ability to upload an assignment, with the workshop module having the extra option of peer review. Participation in the assignments and the quizzes is also expected to reflect more active learning.

4.3 Data Analysis

After data pre-processing, all analyses were run with Stata 14. First, a correlation analysis was conducted for all predictor variables with final exam grade, per course and for all courses combined. To measure the portability of the prediction models, both multi-variate analyses and ordinary least squares regressions were used. As the variables are measured at different levels of hierarchy, i.e., both course level (courseid) and student level (all predictor variables), multi-level analyses were used with crossed random effects (course and individual student). Ordinary least squares regressions were run per course to determine to what extent the effects of the predictors differ per course.

To determine the predictability of the models for each course, multiple linear regressions were conducted for each course separately, using stepwise backward elimination. The criterion for exclusion in each step was $p > 0.2$ for the courses separately and $p > 0.05$ for all courses combined. As the assumption of homoscedasticity was often not met, robust standard errors were used.

To determine the predictability of the models over time, regressions were run using the data available at the end of every week during the course, to analyze to what extent the (accuracy of the) prediction changed over time. For brevity, we do not report the separate regression for every course for every week. As only a few predictors are included, we report standard linear and logistic regressions, using student clustered standard errors, with courses coded as dummies and interactions effects for each course with the mean and deviances from the mean within courses instead of multi-level analysis, as this makes the results easier to interpret. Robustness of all models was checked with 10-fold cross-validation, using the function "crossfold" [57].

Although most previous studies report how well the regression or classification model performed in terms of (pseudo) R-squared values, this is not always a very insightful metric. When grades are predicted, it is useful to know how much the predictions deviate from the true value, on average, or how much better the classification accuracy is than a baseline model (such as a model just predicting that everyone will pass). Such accuracy measures give more insight into whether the model could be used for automated assessment. For this reason, we decided to calculate such fit statistics as well.

5 Results

First, we discuss the results regarding the portability of the prediction models. We then report the findings on the accuracy of the prediction of student performance using LMS data, by discussing the regressions on all courses separately. Finally, we show regression analyses on the LMS data as they become available on a week-by-week basis, to determine whether early intervention may be a reasonable possibility.

5.1 Portability of Prediction Models

Pearson correlation analyses for all courses combined showed that 21 of the 23 predictor variables had a statistically significant correlation with final exam grade (see Table 3). Exceptions were irregularity of study time and number of discussion posts. In-between assessment grade had a large effect size ($r = .54$, $p < .001$), and the number of wiki views a moderate effect size ($r = .43$, $p < .001$). All other variables had an effect size below .30. The correlations between the predictor variables and final exam grade for all courses separately showed mixed results. Only midterm grade correlated significantly with final exam grade for all courses (in which it was available). The number of online sessions, the number of resources viewed, the number of quizzes started, and the number of quizzes passed represented significant correlations for at least 75 percent of the

TABLE 3
Correlations between Dependent Variable Final Exam Grade and Independent Variables for All Courses

	Course																	
	All	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Total number of clicks	.04**	.16***	.01	.03	.11	.05	.10	−.16	.15	−.07	.41***	.41***	.32***	.29*	.17*	.08	.15**	.36**
Number of online sessions	.21***	.37***	.32***	.29***	.31***	.20	.22*	−.04	.20**	.04	.53***	.41***	.30***	.26*	.26***	.36**	.16**	.44***
Total time online (min)	.12***	.24***	.18***	.09	.33***	−.04	.22*	−.04	.12	−.06	.49***	.37***	.29***	.40***	.11	−.04	.04	.20
Number of course page views	.19***	.32***	.23***	.20**	.18*	.22	.15	−.03	.09	−.09	.39***	.41***	.31***	.25*	.15	.27*	.14*	.37**
Irregularity of study time	.03	−.03	−.04	−.06	.06	−.19	.18*	−.10	−.01	−.09	.05	.31***	.20***	.30*	−.21**	−.17	−.08	−.15
Irregularity of study interval	−.11***	−.33***	−.29***	−.28***	−.19*	−.17	.00	.09	.07	.01	−.33***	−.05	−.02	−.12	−.13	−.27*	−.07	−.35**
Largest period of inactivity (min)	−.06***	−.16***	−.17***	−.32***	−.12	−.12	.06	−.01	.13	−.04	−.31***	.10**	.06	.02	.02	−.25*	.00	−.17
Time until first activity (min)	−.13***	−.15**	−.16***	−.08	−.32***	−.13	−.19*	−.36**	−.29***	−.20*	−.05	−.13***	−.13***	−.25*	−.04	−.06	−.18**	.04
Average time per session (min)	−.05***	−.06	−.05	−.14*	.02	−.17	−.05	.07	−.04	−.07	.05	.16***	.15***	.06	−.20*	−.22	−.10	−.27*
Number of resources viewed	.13***					.28*	.00				.40***	.15***	.09**	.21	.23**			.26*
Number of links viewed	.09***	.18***	.24***	.21**	.16	−.10			.05	.03		.06	.01					.21
Number of content page views	.17***	.34***	.28***	.24***	.21*	.25*	.16	.08	.19*	.01	.28***	.37***	.26***	.13	.16*	.23	.09	.40**
Number of discussion post views	.04**	−.02	.13***	.18**	.18*	.16	−.02	−.02	−.06	−.12	.24**	.07*	.05	.11	−.13	.12		.01
Number of discussion posts	.03	−.02	.04	.07							.24**							
Number of quizzes started	.11***	.19***	.13***	.03	.42***	.28*	.28**	.23*								.04		
Mean number of attempts per quiz	.08***		.06*		.08	.06	.05	.05	.21**	−.06	−.02	.22***	.15***	.18	−.08	.13	−.06	
Number of quizzes passed	.20***	.26***	.25***	.12	.46***	.31**	.28**	.25*	.31***	.14	.15	.41***	.30***	.30**	.22**	.03	.19***	
Number of quiz views	.14***	.19***	.12***	.05	.11	.08	.12	−.11	.04	−.12	.23**	.39***	.30***	.27*	.04	.02	.12*	
Number of assignments submitted	.21***										−.03				.22**	.14	.17**	.18
Number of assignment subm. views	.04*		.05	.03	.11						.11				.16	.29*	.16**	.23
Number of wiki views	.43***										.43***							
Number of wiki edits	.04**	−.02	.04	.07					.05		.27**				.23**	.24	.18***	.25
Average assessment grade	.54***	.54***	.48***	.64***	.38***	.27*	.58***		.69***	.30***	.52***	.71***	.47***	.76***	.59***	.74***	.59***	.60***
N	4,989	438	1121	227	135	73	120	76	168	155	136	836	822	74	154	66	326	62

* $p < .05$, ** $p < .01$, *** $p < .001$.

courses. These variables are the most stable predictors in our sample. Most other predictor variables correlated significantly for 30 percent to 60 percent of the courses. Discussion forum and wiki usage had the lowest percentage of significant correlations with final exam grade.

Thus, the predictor variables did not correlate significantly for all courses, and some of the variables showed significant and substantial differences in effect sizes and even the direction of the correlation across courses. This suggests that the effects of these variables as predictors in a multivariate analysis may also differ across courses.

To determine to what extent the effects of the predictors differed across courses, ordinary least squares regression were run on all courses with courses coded as dummies and interaction effects for each course with the other predictors, using student clustered standard errors. All nine basic and study pattern predictors varied significantly and substantially with the course (all $p'\mathrm{s} < .01$). However, running standard regressions is an obvious simplification of the true structure of the data. Because some students followed multiple courses, the cases do not represent unique students. Moreover, the data are obviously clustered by course. To take this hierarchical structure into account, we ran multilevel regressions with final exam grade as our target variable and crossed random effects for course and student. The analysis showed that 8 percent of the variance resided at the course level and 48 percent at the student level. This indicates that we cannot simply ignore the clustering at the student level and the course level, but also that the larger part of the variance can be found at the student level, which may make it easier to find portable results across courses.

The hierarchical structure of the data allows for two types of effects at the student level, i.e., an effect within a course between students or within students across courses. For example, a student may get a higher grade in courses where he or she shows more online activity (a within-student effect), or higher grades may be achieved in courses in which students show more online activity (a between-students effect). Therefore, multi-level analyses for each predictor were run with the mean and deviance from the mean for each predictor (allowing distinguishing within and between student effects), dummies for all courses, and random intercepts for students. We found that the total number of clicks, the number of online sessions, the total time online, the total number of views, the irregularity of study time, and the in-between assessment grade had a significant positive effect on final exam grade both within and between students (all $p'\mathrm{s} < .001$). The time until first activity had a significant negative effect on final exam grade within and between students (all $p's < .05$). The irregularity of study interval had a significant negative effect on final grade between students ($p < .001$). Largest period of inactivity and average time per session did not have significant effects both between and within students. In general, these results show that it may be useful to not only compare students with their peers, but also with their own behavior in other courses.

Combined, these results show that aggregating the LMS data of all courses into a single analysis without using a large number of interaction effects would be a gross oversimplification: only a small subset of the variables has consistent effects across courses. Hence, the portability of the prediction models for final exam grade using LMS variables across these courses is low. On the other hand, the results also show that final grade is to a large extent an individual characteristic, with effects within and between students. This suggests that it may be possible to capture this individual variance through LMS characteristics, for example by including LMS data of the same student in other courses. In addition, it may still be useful to analyze the LMS data for courses separately, to get additional insight into this specific course.

TABLE 4
Final Models Multiple Linear Regression on All Courses

	Course																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Total number of clicks	$-.20^{**}$	$-.39^{***}$	$-.36^{***}$		$-.31$	$-.14$	$-.30^{*}$	1.30^{**}			$.14^{*}$	$.21^{*}$	1.91^{**}				
Number of online sessions		$.24^{**}$	$.12$						$.14$	$.55^{***}$	$.44^{***}$	$.24^{*}$		$.60^{***}$	$.46^{**}$		$.27$
Total time online (min)		$-.19^{*}$		$.17$	$-.23$	$.25^{*}$	$.29$			$.34^{**}$	$-.34^{***}$	$-.18^{*}$	$.76^{***}$	$.16$	$-.23^{*}$	$.19$	
Number of course page views	$.31^{***}$	$.33^{***}$	$.35^{**}$		$.63^{***}$			-1.38^{**}	$-.17$	$-.42^{**}$	-2.28^{**}	$-.52^{**}$					
Irregularity of study time		$-.09$				$.26^{*}$	$-.57^{**}$				$.17^{***}$			$-.40^{*}$	$-.16$		
Irregularity of study interval	$-.37^{***}$	$-.32^{***}$		$-.19^{**}$			$.38$	$-.52^{*}$	$.27$	$-.10$	$-.22^{**}$	$-.12$		$-.34$		$-.53^{***}$	$-.77$
Largest period of inactivity (min)	$.16^{*}$	$.15^{**}$	$-.25^{**}$				$-.56^{*}$	$.56^{*}$	$-.24$		$.32^{***}$	$.19^{**}$		$.33^{*}$		$.50^{**}$	$.72^{*}$
Time until first activity (min)	$-.08$	$-.10^{***}$		$-.27^{**}$		$-.17$	$-.58^{***}$	$-.27^{***}$	$-.21$							$-.12^{*}$	
Average time per session (min)		$.22^{**}$				$-.22$	$.35^{**}$				$.14$	$.13^{**}$	$-.15$	$.20$		$-.18^{*}$	$-.20$
R^2	.17	.19	.18	.18	.13	.13	.37	.18	.08	.32	.23	.12	.29	.19	.17	.10	.32
N	438	1121	227	135	73	120	76	168	155	136	836	822	74	154	66	326	62

Constants omitted from the table, standardized betas reported for all variables.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

5.2 Predicting Student Performance Per Course

We now investigate whether LMS data can be used to predict student performance and explain the variance at the student level. To determine this, a multi-level analysis on final exam grade was conducted with LMS data and crossed-random effects for course and student. It was found that after adding the LMS data, the amount of variance that could be explained at the student level dropped from 48 percent to 38 percent, and at the course level raised from 8 percent to 18 percent. Thus, LMS data can indeed be used to explain part of the variance in final exam grade.

To investigate the differences between the prediction models per course, separate multiple linear regressions were run per course, with final exam grade as outcome variable and all basic and pattern variables as predictors, as these were available in all courses. All predictors with a significance level above .2 were removed from the models. The results in Table 4 show that LMS data do explain some of the variance in final exam grade in each course, but that the amount of explained variance differs strongly: from 8 percent for course 9 (no significant predictors), to 37 percent in course 7. Table 4 shows the regression coefficients and p-values of the variables included in the final models per course (containing only the variables that were included in the final models of at least six courses). The irregularity of study time per session is the least present (6 out of 17 courses), whereas the total time online had a significant regression coefficient most often (12 out of 17). However, the direction of the coefficient of time online varied across courses: in some courses time online has an unexpected negative influence on final exam grade, while in other courses it has a positive influence. This contradiction also holds for the total number of clicks and the total number of views, but for a different set of courses. The amount of sessions is positively related to the final exam grade in all six final models, while the regularity of study interval and time until the first activity have a negative influence in all eight and six final models, respectively. These findings imply that more general conclusions based on these data must be restricted to the following variables: more online sessions, lower standard deviation of the time between the sessions, and less time until the first session (i.e., starting early) all go with a higher grade.

The large variety in these prediction models again reveals low portability of the prediction models. The models do give some insight into the factors influencing student performance within a single course. However, with an average mean residual (averaged across all courses) of 1.56, where 70 percent of the grades are between 4.0 and 9.0, the models are quite far away from an accurate prediction. Moreover, these prediction models are based on the complete LMS data available at the end of a course, a point in time at which it is not possible to intervene anymore.

5.3 Predicting Student Performance Over Time

To assess whether LMS data could offer a basis for interventions during a course, we reran our analyses with the Moodle LMS usage variables grouped per week. Only basic predictors (total amount of clicks, the number of online sessions, the total time online, and the total amount of views) were used, as study patterns (e.g., the regularity of study time) were often not available (for example SD of study interval for two sessions and hence one interval) or not yet meaningful (for example SD of study time for two sessions). In addition, the variables per module were also excluded from the analyses, as they did not provide enough variability in the first few weeks. For brevity, we do not report separate regressions for all courses for all weeks, but we report on the standard regressions per week with courses coded as dummies and interaction effects for each course with the mean and deviances from the mean within courses for each predictor, using student clustered standard errors.

5.3.1 Predicting Final Exam Grade Over Time

In the week before the course starts (week 0), all predictor variables had to be excluded from the stepwise backwards regression, so no valid prediction model could be produced from the LMS data. From week 1 until the end of the course more LMS data became available each week, which resulted in a slightly improved prediction, with adjusted R^2 increasing from .14 to .22 (see Fig. 1). When in-between assessment grades, available after week 5, were taken into account, a larger improvement in prediction was found (week 5: $F\,(177,\ 4{,}811) = 22.32$, adjusted $R^2 = 0.43$). With the LMS data of all weeks included, LMS data could only explain an additional 2 percent of the variance in final grade, over and above the average midterm grades. Thus, LMS data, at least as implemented here, are of substantially smaller predictive value than the in-between assessment grades. However,

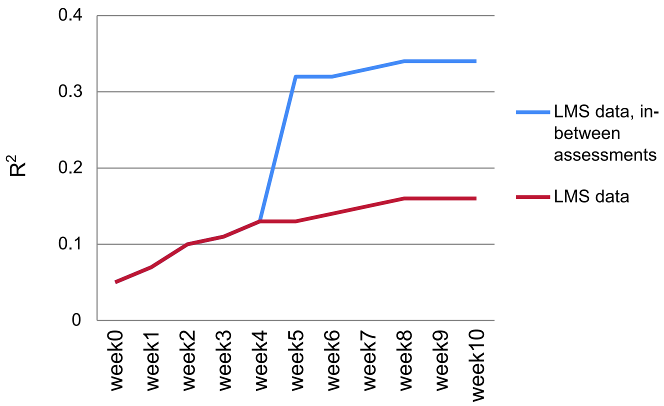

Fig. 1. R^2 of multiple linear regressions on final exam grade over weeks, using LMS data with and without in-between assessment grade.

these midterm grades are often not available until the course is halfway and sometimes not available at all. For earlier intervention (that is, before assessment grades become available), LMS data can perhaps be used, but in our data they were of limited value in predicting the final exam grade.

5.3.2 *Predicting Pass-Fail Probabilities*

Actually, we might not need to know the exact final exam grade to be able to improve learning and teaching. Knowing whether a student is at risk of failure might be enough to determine whether an intervention is needed. Therefore, binary logistic regressions were conducted as in [28], [32] with exam grade > 5.5 (out of 10) coded as 1 (pass). Again, courses were coded as dummies and interaction effects for each course with the mean and difference from the mean were included. However, these binary logistic regressions did not lead to a high accuracy either. After week 0, 1,548 out of 2,704 students who passed were correctly predicted as a pass, and 1,423 out of the 2,266 students who failed where correctly classified as 'at risk'. This represented a total classification accuracy of 60 percent. The classification accuracy increased to 69 percent when all LMS data and in-between assessment grades were used, with 1,950 out of 2,720 students correctly classified as pass, and 1,479 out of 2,268 students correctly classified as 'at risk'. The best compromise between early feedback and classification accuracy seems to be after week 3, where 1,922 out of 2,720 students were correctly classified as pass, and 1,397 of the 2,268 failing students were correctly classified as 'at risk', resulting in an overall accuracy of 67 percent.

If we intervened with students at risk based on this information, still 871 students would eventually fail without having received an intervention. To improve learning and teaching, it may be more useful to intervene with as many students 'at risk' as possible, at the cost of intervening with some students who do not need it. To consider this, we set the specificity (true negative rate) at 95 percent. This resulted in 656 out of 2,720 students correctly classified as successful, and 2,158 out of 2,268 students correctly classified as 'at risk'. Thus, to be able to intervene with 95 percent of the students who would fail without it, we need to intervene with 85 percent of the students, of which 49 percent would not need the intervention.

Thus, using LMS data does not lead to a very accurate prediction of whether a student would pass or fail the final exam. As the final exam counted toward 68 percent of the final grade on average for all courses, the findings are expected to be similar for predicting whether a student would pass or fail the whole course. Indeed, binary logistic regressions of final course grade showed a similar increase in the prediction and only a somewhat higher accuracy (2-3 percent) for the predictions per week compared to the binary logistic regressions on final exam grade.

6 Conclusion and Discussion

In the current study we analyzed and compared prediction models of student performance using LMS data of 17 blended courses, to determine the portability of these prediction models across courses. Moreover, we assessed the accuracy of these prediction models for early prediction of student performance.

6.1 Portability of Prediction Models

For the prediction models we extracted variables similar to those used in previous research: basic predictors such as total time online and number of clicks, and predictors found in the different modules of the LMS, such as discussion posts and quizzes passed. In addition, we included more complex variables related to study patterns, such as the irregularity of study time and the time until the first activity. Our results show that there is no comprehensive set of variables that can consistently predict student performance across multiple courses. Correlational analyses as well as linear regressions showed differences in predictive power of the variables between the courses. Only the in-between assessment grade correlated significantly with the final exam grade in all courses, confirming that the in-between measurement of performance is in line with the measurement of performance at the end of the course. Discussion forums and wiki usage showed significant correlations in the lowest number of courses, indicating that these variables are not very useful (and not very stable) predictors of final exam grade across courses, at least in our data.

Our results add to the empirical base of learning analytics findings, and corroborate previous studies on predicting student success, which have also shown different results in correlations and prediction models, albeit for contexts that were more varied than ours. We tried to account for potential differences by focusing on a larger set of courses at a single institution, and by including most if not all of the predictors that have been used in previous research. However, in spite of the fact that this is likely to keep the contextual effects more constant, we still found substantial differences in the sign and size of the effects of different predictors. These findings are in line with [10] and show that even within one institution, using one LMS, and controlling for a large set of predictor variables, the portability of the prediction models across courses is low.

It is possible that the differences in the prediction models, both here and in [10], could be explained by the learning designs, as the content of the activities available in an LMS have been shown to influence the number of LMS visits [37]. What goes against this argument is that, compared to [10], the courses in our data used fewer course-specific modules in the LMS and are hence more similar in the activities provided in the LMS (even when one assumes some variation in learning design within given modules). Moreover, we only included predictors from modules that were available in all courses. Our analyses, hence based on a more generic set of variables than in previous research, nevertheless confirm that different predictors are useful in different courses. It is possible that a more fine-grained disentanglement of the content of modules might change this finding, but this would call for a more elaborate set of courses. As far as we can see here, the data of multiple courses can thus neither be easily combined for a common analysis, nor to construct general models that are likely to predict well outside the boundaries of the actual data.

A limitation of this study is that data was available for only a part of the courses provided by the university, resulting in a skewed representation of courses: mostly first year courses with similar use of blended learning. Thus, our results might not carry over to other types of courses or blended learning. However, given the similarity of our courses, one would argue that portability of the results would be more likely, which we nevertheless do not find.

We suspect that the differences that we found may well be due to both the extent to which a given module is used and the way in which a module is supposed to be used. For example, the small effects of forum and wiki usage could be due to the fact that variability of use happened to be low for these modules (cf. [30]). Post-hoc multi-level analyses and regressions with interaction effects for each course confirm that the effects of the predictors strongly differ between the courses. It may be useful to further investigate the effect of specific course and module characteristics, for example based on the course syllabi, on the use of the LMS (cf. [11]). In our case, we could only investigate 17 courses, which does not allow for a lot of variation within course modules. Another potential issue with this explanation is that these post-hoc analyses suggest that the larger amount of the variation seemed to reside at the student level.

Next to the potential inclusion of more detailed course characteristics, it may also be useful to consider that students differ in how they use LMSs while studying. Our multi-level analyses indeed show that a high proportion of variance could be explained at the student level. This is promising, in the sense that it does appear to be variation in student characteristics or usage level that can accurately predict the final exam grade. On the other hand, none of the usage characteristics that have been used in the literature before (most of which we included here) seemed to pick up this variance.

6.2 Predicting Student Performance Per Course

As we could not simply aggregate all data, separate regressions were run for each course. The results showed that the accuracy of the prediction models differed to a large extent between the courses, from 8 percent to 37 percent explained variance in final grade. With an average R^2 of 0.20 our LMS data turned out to be a weak predictor. Most other studies explained more of the variance in final grade [27], [28], [31], [32]. Only one study explained less of the variance using LMS data [13], probably because the students made more use of the other online systems, including e-tutorials. Thus, as in previous research, the LMS variables do show some relation with final exam grade. However, for the prediction of student performance in the courses under study the variables are of limited value. Even though we have generally used the same variables and methods as previous research, and several variables show significant effects (which is insightful in itself), the effects are not strong enough to generate a precise enough prediction.

To determine if the prediction was accurate enough for early intervention, linear and binary logistic regressions were run across weeks. The prediction of final grade with LMS data increased only slightly across the weeks, with a serious improvement after week 5, when in-between assessment grades became available. For early intervention, before the assessment data becomes available, LMS data resulted in only a weak prediction. For early prediction of the pass-fail probabilities LMS data also showed to be a weak predictor, with a total accuracy of 67 percent in week 3. To be able to intervene with at least 95 percent of the students at risk at the end of week 3, 85 percent of the students should receive an intervention, which half of them would not actually need. Thus for early intervention at the level of the aggregate set of courses, when midterm grades are not available yet, our LMS data are of limited value.

A final remark is that the limited value for (early) prediction of the LMS data might be due to the (lack of) relation between the activities in the LMS and the final exam for at least some courses. Final exams were often individual, written on paper, while the activities and teaching methods in the course were online and made use of different online tools typically not available at the exam.

6.3 What Are We Measuring? The Need for Theory

Moreover, the low predictability and also the low portability of the LMS variables may be due to the fact that we do not really know what our measurements are actually measuring. LMSs provide us with raw log data, but these are not concrete measurements of a previously defined theoretical concept. LMS data is for example at best an indirect measurement of motivation. To improve the predictions with LMS data, we need to get a better insight into what the LMS data represents, what the effects are, and how they can be converted into concrete measurement of concepts. Several attempts at creating more general theoretical frameworks have been made ([58], [59]) but none of them have led to concrete suggestions on how to process LMS data in such a way that they more accurately reflect underlying concepts.

Taken together, our findings underscore the notion that more elaborate theoretical reasoning is needed in learning analytics to achieve generalizable results [9]. Still, learning analytics can be used to analyze a single course. For example, basic variables such as the amount of clicks can be useful to predict student performance or to evaluate the learning design for a specific course [60]. More advanced analyses such as checkpoint and process analyses can be used to relate students' behavior over time with the learning design [60]. However, an increased emphasis on theory, including how the theoretical arguments are converted into appropriate measurements, is needed to guide inclusion of predictors and interpretation of the results. For example, this could result in more complex variables related to patterns of the activities in the LMS, such as the order of the events or the type of the event after an action of the teacher.

This would also be useful for early prediction, as these types of patterns are available early in the course.

Next to additional theory and accordingly more appropriately created measurements, it may be useful to add other types of data such as qualitative data and other data sources as predictors as well. Qualitative data, for instance from the discussion forum, can give more insight into the type of participation of students and may therefore be more useful for predicting student performance [30], [46]. Additionally, this could allow for better identification of students who show high participation but nevertheless receive low grades [27].

Other sources of data could be useful as well, especially to improve early prediction when they are available at the beginning of the course. Shum and Crick [25], among others, have already argued that variables that are traditionally used in the social sciences, such as learning dispositions or other personality characteristics, can provide more detailed and timely information about the performance of students. While LMS data are a by-product of learner activity, self-disclosure data about dispositions might give higher-order information about students' states that is harder to infer from the raw LMS logs [25]. Accordingly, Tempelaar et al. [13] analyzed demographics, entry test results, learning dispositions, motivation and engagement, LMS data, e-tutorials, and assessment data in two courses with 873 students. They found that entry test results, learning styles, and motivation and engagement had a significant correlation with final grade. Assessment data was found to be the best predictor, but until this data is available, learning dispositions would be the best and proper alternative, as these were found to be most complementary to LMS data. As their study was conducted on a heterogeneous set of students from only two courses, and as previous studies have shown to be quite diverse, future work is needed to draw conclusions about the usage of learning dispositions combined with LMS data for early feedback. Currently, we are supplementing our data with such other data sources.

To conclude, the emergence of ICT into learning and teaching has supplied us with a rich information source of raw logs of behavior in LMSs. Unfortunately, inconsistencies across course findings make it difficult to draw general conclusions about the online behavior of potential students at risk. Additional theoretical argumentation and data sources need to be included to predict student performance and improve learning and teaching.

Acknowledgments

This work was supported in part by a grant from the 4TU. Centre for Engineering Education.

References

[1] N. Hoic-Bozic, V. Mornar, and I. Boticki, "A blended learning approach to course design and implementation," *IEEE Trans. Educ.*, vol. 52, no. 1, pp. 19–30, Feb. 2009.

[2] S. Retalis, A. Papasalouros, Y. Psaromiligkos, S. Siscos, and T. Kargidis, "Towards networked learning analytics–A concept and a tool," in *Proc. 5th Int. Conf. Netw. Learn.*, 2006, pp. 1–8.

[3] A. A. Piña, "An overview of learning management systems," in *Virtual Learning Environments: Concepts, Methodologies, Tools and Applications*, 1st ed. Louisville, KY, USA: Sullivan Univ. Syst., 2012, pp. 33–51.

[4] G. Siemens and R. S. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in *Proc. 2nd Int. Conf. Learn. Analytics Knowl.*, 2012, pp. 252–254.

[5] P. Long, G. Siemens, G. Conole, and D. Gasevic, (Eds.) "Announcing open course: Learning and knowledge analytics," in *Proc. 1st Int. Conf. Learn. Analytics Knowl.*, Banff, AB, Canada, 2011.

[6] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[7] S. B. Shum and R. Ferguson, "Social learning analytics," *Educ. Technol. Soc.*, vol. 15, no. 3, pp. 3–26, 2012.

[8] R. S. J. d Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Educ. Data Min.*, vol. 1, no. 1, pp. 3–17, Oct. 2009.

[9] D. W. Shaffer, et al., "Epistemic network analysis: A prototype for 21st-century assessment of learning," *Int. J. Learn. Media*, vol. 1, no. 2, pp. 33–53, 2009.

[10] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *Internet High. Educ.*, vol. 28, pp. 68–84, Jan. 2016.

[11] S. Joksimović, D. Gašević, T. M. Loughin, V. Kovanović, and M. Hatala, "Learning at distance: Effects of interaction traces on academic achievement," *Comput. Educ.*, vol. 87, pp. 204–217, Sep. 2015.

[12] J. P. Campbell and D. G. Oblinger, "Academic analytics," *Educause*, vol. 42, pp. 40–42, 2007.

[13] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," *Comput. Human Behavior*, vol. 47, pp. 157–167, Jun. 2015.

[14] D. Clow, "An overview of learning analytics," *Teach. High. Educ.*, vol. 18, no. 6, pp. 683–695, 2013.

[15] M. Richardson, C. Abraham, and R. Bond, "Psychological correlates of university students' academic performance: A systematic review and meta-analysis," *Psychol. Bull.*, vol. 138, no. 2, pp. 353–387, 2012.

[16] M. G. Moore, "Editorial: Three types of interaction," *Amer. J. Distance Educ.*, vol. 3, no. 3, pp. 1–6, Jan. 1989.

[17] Á. F. Agudo-Peregrina, Š. Iglesias-Pradas, M. Á. Conde-González, and Á. Hernández-García, "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning," *Comput. Hum. Behav.*, vol. 31, pp. 542–550, Feb. 2014.

[18] S. Dawson, N. Mirriahi, and D. Gasevic, "Importance of theory in learning analytics in formal and workplace settings," *J. Learn. Anal.*, vol. 2, no. 2, pp. 1–4, Dec. 2015.

[19] D. Cunningham and T. Duffy, "Constructivism: Implications for the design and delivery of instruction," *Handbook Res. Educ. Commun. Technol.*, New York: McMillan, pp. 170–198, 1996.

[20] M. Ally, "Foundations of educational theory for online learning," *Theory and Practice of Online Learning*. Athabasca, AB: Athabasca Univ., pp. 15–44, 2004.

[21] H. Coates, R. James, and G. Baldwin, "A critical examination of the effects of learning management systems on university teaching and learning," *Tertiary Educ. Manage.*, vol. 11, no. 1, pp. 19–36, Mar. 2005.

[22] P. H. Winne, "A metacognitive view of individual differences in self-regulated learning," *Learn. Individual Differences*, vol. 8, no. 4, pp. 327–353, 1996.

[23] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," *Metacognition Educ. Theory Pract.*, vol. 93, pp. 27–30, 1998.

[24] J. S. Brown, A. Collins, and P. Duguid, "Situated cognition and the culture of learning," *Educ. Res.*, vol. 18, no. 1, pp. 32–42, 1989.

[25] S. Buckingham Shum and R. D. Crick, "Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics," in *Proc. 2nd Int. Conf. Learn. Analytics Knowl.*, 2012, pp. 92–101.

[26] S. Rafaeli and G. Ravid, "Online, web-based learning environment for an information systems course: Access logs, linearity and performance," in *Proc. Inf. Syst. Educ. Conf.*, 1997, vol. 97, pp. 92–99.

[27] L. V. Morris, C. Finnegan, and S.-S. Wu, "Tracking student behavior, persistence, and achievement in online courses," *Internet High. Educ.*, vol. 8, no. 3, pp. 221–231, 2005.

[28] L. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators: A proof of concept," *Comput. Educ.*, vol. 54, no. 2, pp. 588–599, Feb. 2010.

[29] S. Dawson, E. McWilliam, and J. P.-L. Tan, "Teaching smarter: How mining ICT data can inform and improve learning and teaching practice," in *Proc. Annu. Conf. Australasian Soc. Comput. Learn. Tertiary Educ.*, Jan. 2008, pp. 221–230.

[30] D. Nandi, M. Hamilton, J. Harland, and G. Warburton, "How active are students in online discussion forums?" in *Proc. 13th Australasian Comput. Educ. Conf.*, 2011, vol. 114, pp. 125–134.

[31] T. Yu and I.-H. Jo, "Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education," in *Proc. 4th Int. Conf. Learn. Anal. Knowl.*, 2014, pp. 269–270.

[32] N. Z. Zacharis, "A multivariate approach to predicting student outcomes in web-enabled blended learning courses," *Internet High. Educ.*, vol. 27, pp. 44–53, Oct. 2015.

[33] V. Kovanović, D. Gašević, S. Dawson, S. Joksimović, R. S. Baker, and M. Hatala, "Penetrating the black box of time-on-task estimation," in *Proc. 5th Int. Conf. Learn. Analytics Knowl.*, 2015, pp. 184–193.

[34] C. Beer, K. Clark, and D. Jones, "Indicators of engagement," in *Australian Soc. Comput. Learn. Tertiary Educ. Annu. Conf.*, 2010, pp. 75–86.

[35] S. M. Jayaprakash, E. W. Moody, E. J. Lauría, J. R. Regan, and J. D. Baron, "Early alert of academically at-risk students: An open source analytics initiative," *J. Learn. Anal.*, vol. 1, no. 1, pp. 6–47, 2014.

[36] E. J. M. Lauría, J. D. Baron, M. Devireddy, V. Sundararaju, and S. M. Jayaprakash, "Mining academic data to improve college student retention: An open source perspective," in *Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, 2012, pp. 139–142.

[37] B. Rienties, L. Toetenel, and A. Bryan, "'Scaling up' learning design: Impact of learning design activities on LMS behavior and performance," in *Proc. 5th Int. Conf. Learn. Anal. Knowl.*, 2015, pp. 315–319.

[38] J. Milne, L. M. Jeffrey, G. Suddaby, and A. Higgins, "Early identification of students at risk of failing," in *Proc. Australian Soc. Comput. Learn. Tertiary Educ. Annu. Conf.*, 2012, pp. 657–661.

[39] Y.-H. Hu, C.-L. Lo, and S.-P. Shih, "Developing early warning systems to predict students' online learning performance," *Comput. Hum. Behav.*, vol. 36, pp. 469–478, Jul. 2014.

[40] J. Schell, B. Lukoff, and C. Alvarado, "Using early warning signs to predict academic risk in interactive, blended teaching environments," *Internet Learn.*, vol. 3, no. 2, 2014, Art. no. 6.

[41] Y. Park, J. H. Yu, and I.-H. Jo, "Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute," *Internet High. Educ.*, vol. 29, pp. 1–11, Apr. 2016.

[42] S. J. Dollinger, A. M. Matyja, and J. L. Huber, "Which factors best account for academic success: Those which college students can control or those they cannot?" *J. Res. Personality*, vol. 42, no. 4, pp. 872–885, Aug. 2008.

[43] M. Munoz-Organero, P. J. Munoz-Merino, and C. D. Kloos, "Student behavior and interaction patterns with an LMS as motivation predictors in e-learning settings," *IEEE Trans. Educ.*, vol. 53, no. 3, pp. 463–470, Aug. 2010.

[44] J. W. You, "Identifying significant indicators using LMS data to predict course achievement in online learning," *Internet High. Educ.*, vol. 29, pp. 23–30, Apr. 2016.

[45] J. B. Arbaugh, "System, scholar or students? Which most influences online MBA course effectiveness?" *J. Comput. Assist. Learn.*, vol. 30, no. 4, pp. 349–362, 2014.

[46] J. Davies and M. Graff, "Performance in e-learning: Online participation and student grades," *Brit. J. Educ. Technol.*, vol. 36, no. 4, pp. 657–663, 2005.

[47] M. C. O'Connor and S. V. Paunonen, "Big five personality predictors of post-secondary academic performance," *Personality Individual Differences*, vol. 43, no. 5, pp. 971–990, Oct. 2007.

[48] J. G. Nicholls, "Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance," *Psychol. Rev.*, vol. 91, no. 3, pp. 328–346, 1984.

[49] J. Hattie, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Evanston, IL, USA: Routledge, 2008.

[50] J. F. Superby, J.-P. Vandamme, and N. Meskens, "Determination of factors influencing the achievement of the first-year university students using data mining methods," in *Proc. 8th Int. Conf. Intell. Tutoring Syst.*, 2006, pp. 37–44.

[51] M. A. Conard, "Aptitude is not enough: How personality and behavior predict academic performance," *J. Res. Personality*, vol. 40, no. 3, pp. 339–346, Jun. 2006.

[52] J. Knight, "Distinguishing the learning approaches adopted by undergraduates in their use of online resources," *Active Learn. High. Educ.*, vol. 11, no. 1, pp. 67–76, Mar. 2010.

[53] S. Iglesias-Pradas, C. Ruiz-de-Azcárate, and Á. F. Agudo-Peregrina, "Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies," *Comput. Human Behavior*, vol. 47, pp. 81–89, 2015.

[54] R.-S. Shaw, "A study of the relationships among learning styles, participation types, and performance in programming language learning supported by online forums," *Comput. Educ.*, vol. 58, no. 1, pp. 111–120, Jan. 2012.

[55] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2007.

[56] A. N. Kluger and A. DeNisi, "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory," *Psychol. Bull.*, vol. 119, no. 2, 1996, pp. 254–284.

[57] B. Daniels, "CROSSFOLD: Stata module to perform k-fold cross-validation," *Statistical Software Components S457426*, Boston College Department of Economics, 2012.

[58] L. Rankine, L. Stevenson, J. Malfroy, and K. Ashford-Rowe, "Benchmarking across universities: A framework for LMS analysis," in *Proc. Australian Soc. Comput. Learn. Tertiary Educ. Annu. Conf.*, 2009, pp. 815–819.

[59] O. Petropoulou, S. Retalis, K. Siassiakos, S. Karamouzis, and T. Kargidis, "Helping educators analyse interactions within networked learning communities: A framework and the AnalyticsTool system," in *Proc. 6th Int. Conf. Netw. Learn.*, 2008, pp. 5–7.

[60] L. Lockyer, E. Heathcote, and S. Dawson, "Informing pedagogical action: Aligning learning analytics with learning design," *Amer. Behavioral Sci.*, Mar. 2013, pp. 1439–1459.

Rianne Conijn is a PhD candidate in learning analytics and educational data mining at Tilburg University. Her research interests include the analysis and interpretation of (sequences of) online and offline learning behavior to improve learning and teaching.

Chris Snijders is professor of Sociology of Technology and Innovation at Eindhoven University of Technology. His research interests include decision-making and expertise, especially model-based decision-making, and (online) trust and cooperation.

Ad Kleingeld is an assistant professor in the Human Performance Management group at Eindhoven University of Technology. His current research interests include work motivation, goal-setting and feedback interventions to enhance performance, subconscious goals, and predicting student performance.

Uwe Matzat works as an assistant professor of sociology in the HTI group at Eindhoven University of Technology. He is an editor of the *International Journal of Internet Science*. His research interests include the design, use, and the social implications of social media. He has a special interest in the use of social media for educational purposes and self-regulated learning.