MASTER

How to succeed?

the effect of different data sources on predicting student performance

Conijn, M.A.

Award date:
2015

Eindhoven, December 2015

How to succeed?

The effect of different data sources on predicting student performance

By M.A. (Rianne) Conijn

Identity number 0740635

In partial fulfilment of the requirements for the degree of

Master of Science
in Human Technology Interaction

Department of Industrial Engineering and Innovation Sciences
Eindhoven University of Technology (TU/e)

Supervisors:

Prof. dr. C.C.P. (Chris) Snijders	TU/e, IEIS, Human Technology Interaction
Dr. ir. P.A.M. (Ad) Kleingeld	TU/e, IEIS, Human Performance Management

Keywords:
Learning analytics, Learning Management System, predicting student performance, learner data

Preface

Eindhoven, December 2015

During my career as a student I saw how technology changed learning and teaching. I think this shift offers a lot of opportunities for improving learning and teaching, which keeps grabbing my attention. With this thesis I hope to have contributed a tiny bit to the improvement of learning and teaching. Moreover, I hope these opportunities of technology in education grab your attention as a reader as well.

The upcoming thesis consists of three self-contained parts. The first part includes a literature review on learning analytics. Secondly, a conference paper is included which I wrote in collaboration with my supervisor Prof. dr. C.C.P (Chris) Snijders. This paper has been submitted as a full paper to the sixth International Learning Analytics and Knowledge (LAK) Conference, which will be held in Edinburgh, United Kingdom on April 25-29. Two days after my graduation we will hear whether it is accepted! Lastly, my final report is included, where all parts are brought together.

This thesis would not have been as it is right now without the support of many. First, I would like to thank Chris Snijders, my first supervisor, for his thoughts on (learning) analytics, but mostly for consistently trying to get the best out of me. Without his help I would have given up on the submission of the conference paper and I would probably still rack my brains about multi-variate analyses. Secondly, I would like to thank my second supervisor Ad Kleingeld, who kept being enthusiastic about new results. Thirdly, I would like my mentor Martijn Willemsen, who helped with arranging my international semester and this master thesis, and who started an insightful discussion at my interim presentation.

Next to my supervisors and mentor, I would like to thank other academic staff who helped me, including dr. Dirk Tempelaar from Maastricht University for his thoughts on combining LMS data with learner data; dr. Mykola Pechenizkiy for his view on learning analytics using more advanced data mining and clustering techniques; ir. Jan-Willem Knopper and dr. Hans Cuypers for their explanations of the Moodle logs (all three from the department of Mathematics and Computer Science at TU/e).

I would like to thank my peers, especially Renée, Marlies, Gemma, and Wouter, for proofreading my thesis, discussing the topic, and the de-stressing dances in the office and drinks at Intermate. Lastly, I would like to thank my friends and family for supporting me from the beginning to and through the last bits. Rob, a special thanks to you for endlessly discussing my thoughts, unconditional support, and love.

Enjoy reading,

Rianne

Abstract

The following thesis consists of three self-contained parts:

1. **Literature Review**: Opportunities and challenges in the emerging field of learning analytics – A literature review

2. **Conference paper**: The value of LMS data for predictive modeling of student performance – A study on 17 blended courses using Moodle LMS *(As submitted to the sixth International Learning Analytics and Knowledge (LAK) Conference, written in collaboration with Prof. dr. C.C.P Snijders)*

3. **Final report**: How to succeed? – The effect of different data sources on predicting student performance

Literature Review

Although the field of learning analytics is relatively new, a variety of research can be found, scattered over multiple journals and conference proceedings. This paper will provide a literature review to get a better overview of the interdisciplinary field of learning analytics. Research in three main topics in learning analytics will be discussed: predicting student success, analytics and visualization tools and the implementation of learning analytics. The main focus will be on predicting student success, also including studies from educational data mining. Next to an overview of variables from learning management systems used for predicting student success, other data sources will be included, such as learner data and course data. The literature review shows that there are a lot of challenges in the field of learning analytics and numerous questions are still unanswered, resulting in a high amount of interesting opportunities for future work.

Conference paper

With the adoption of Learning Management Systems (LMSs) in a large number of educational institutions, a lot of data has become available on students' online behavior. Many researchers have used students' LMS data to predict student performance. Unfortunately, this has led to a rather diverse set of findings, possibly because the courses under study are diverse in many ways. The educational context in which the courses take place varies, just as the kind of use that is being made of the LMS, to name just a few. After providing a brief overview of recent findings, we analyze a larger sample of 17 blended courses with 4,989 students in Moodle LMS, in which we predict student performance from LMS and intermittent assignment grades based on the same kind of variable constructions as encountered in the literature. Our analyses show that, irrespective of the fact that all courses were taught at the same institution, the results of predictive modeling depend heavily on course characteristics. Our analyses also show that when in-between assessment grades are taken into account, LMS data are of small additional value. For early intervention, when such grades are not yet available, our LMS data are shown to be a rather weak predictor. To improve the prediction of student performance, especially for early intervention, more data need to be included than can be easily inferred from LMS logs.

Final report

Much research in the field of learning analytics focusses on the predictive modelling of student performance with data from Learning Management Systems (LMSs). Unfortunately, these studies are often exploratory and use different methodologies. This results in different outcomes which are hard to compare. Even when the same method is used, differences are found. Moreover, Gašević and colleagues (2016) found that the prediction models using LMS data even differ per course within one institution. Hence, the portability of the prediction models across courses might be low. Additionally, the studies mostly focus on LMS data only, while ignoring learner data such as ability, personality, and motivation. These variables have been found significant and robust predictors. However, the prediction models using learner data and LMS data have rarely been combined or compared.

In the current study, first, the portability of seventeen blended courses at Eindhoven University of Technology are studied. Contrary to Gašević and colleagues (2016), predictors are used which are available in all courses. It is again found that the portability across courses is low. Thus, although prediction models might be useful for a specific course, data of multiple courses cannot be aggregated. When learner data were added to the LMS data, the prediction models still differed. Thus, adding student characteristics alone is not sufficient for increasing the portability. Exploratory analysis showed that course characteristics do have some effect on the portability, but still differences remain. Hence, one should still be careful when aggregating data from courses with the same characteristics.

Secondly, the predictability of student performance using learner data is compared with LMS data. It is found that learner data outperform LMS data. However, LMS data outperform learner data when in-between assessment data are added to LMS data. The addition of learner data to LMS data is especially useful when in-between assessments are not (yet) available, as is the case with early intervention. However, both sources combined still are not accurate for early prediction of final exam grade and pass/fail probabilities. To improve the early intervention, more data needs to be included.

Contents

1 Literature Review 1

2 Conference Paper 33

3 Final Report 43

Appendix: Code of scientific conduct 85

Opportunities and challenges in the emerging field of Learning Analytics

a literature review

Rianne Conijn – 0740635

August 2015

Supervisors:
Prof. dr. C.C.P. (Chris) Snijders
Dr. ir. P.A.M. (Ad) Kleingeld

Abstract
Although the field of learning analytics is relatively new, a variety of research can be found, scattered over multiple journals and conference proceedings. This paper will provide a literature review to get a better overview of the interdisciplinary field of learning analytics. Research in three main topics in learning analytics will be discussed: predicting student success, analytics and visualization tools and the implementation of learning analytics. The main focus will be on predicting student success, also including studies from educational data mining. Next to an overview of variables from learning management systems used for predicting student success, other data sources will be included, such as learner data and course data. The literature review shows that there are a lot of challenges in the field of learning analytics and numerous questions are still unanswered, resulting in a high amount of interesting opportunities for future work.

Contents

1 Introduction 5
2 Defining the field 5
2.1 Academic Analytics 7
2.2 Educational Data Mining (EDM) 7
2.3 Goals of learning analytics 8
3 Predicting student success 9
3.1 Learner data 9
3.2 Course data 10
3.3 LMS data 11
3.4 Combining learner data with LMS data 14
3.5 Predicting LMS behaviour 15
3.6 Conclusion 16
4 Learning analytics tools 17
4.1 Analytics tools 17
4.2 Visualization tools 18
4.3 Conclusion 19
5 Implementing learning analytics 19
6 Conclusion and future work 20
7 References 24
Appendix A 29

1 Introduction

In the last few decades, ICT has emerged more and more in learning and teaching. This resulted in the adoption of learning management systems (LMSs), also known as Virtual Learning Environments (VLEs), in a vast majority of educational institutions (Retalis, Papasalouros, Psaromiligkos, Siscos, & Kargidis, 2006). LMSs have the goal to support student learning by providing course content, and by allowing for additional benefits such as quizzes, presentations, assignments and forums (Piña, 2012). These developments do not only change the way courses are taught and learned, but also provide opportunities to improve learning and teaching. As all clicks are monitored and stored in LMSs, this gives a lot of information about the behaviour of users in these systems. Interpreting and contextualizing this information to improve learning and teaching, increasing student success, and detecting at-risk students, i.e. students who have a high chance of failure, is also known as learning analytics (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014).

Although the term learning analytics is rather new, already a wide variety of research can be found within the field. This is mostly due to the wide variety of backgrounds of the researchers, including computer science, statistics, (educational) psychology, psychometrics, and several other fields (Clow, 2013). These backgrounds result in different goals of learning analytics, different methods used, and publications scattered amongst multiple journals. This variety and spread makes it hard to get a good overview of the field and which questions still need to be answered. Therefore, this literature review will discuss current research and topics in the field of learning analytics, based on more than 50 papers scattered over more than 30 journals and conference proceedings, starting from 1997. This review does not aim to provide a complete overview of all research conducted in the field, but it focusses on giving an overview of the range of literature available.

First, the field of learning analytics will be defined and compared to the adjacent fields of educational data mining and academic analytics. Thereafter, three central topics in learning analytics will be discussed: predicting student success, analytics and visualization tools, and implementing learning analytics. The main focus will be on predicting student success, as this is the most common subject in the literature. These studies will be categorized on the different types of predictors used: learner data, course data, and data from learning management systems. Finally, a comprehensive overview for future work and research directions will be given.

2 Defining the field

Learning analytics is defined as “the measurement, collection, analysis and reporting of data about learners and their context, for purposes of understanding and optimising learning and the environments in which it occurs” (Siemens, 2011). Learning analytics has emerged as an interdisciplinary research field over the last decade. The growing research community initiated the annual Learning Analytics and Knowledge (LAK) conference in 2010 and in 2011 the Society of Learning Analytics Research (SoLAR) was formed (Clow, 2013; Siemens & Baker, 2012). Although the field and the term learning analytics is rather new, analysing student data to understand how students learn, and to improve learning and teaching has been a topic of research over decades. For example, for 80 years class attendance has been found to predict performance (Dollinger, Matyja, & Huber, 2008).

Formerly, analyses student of data was mostly done by researchers from the fields of social sciences, educational psychology and pedagogy. These studies were often based on earlier research or frameworks, and were tested with validated questionnaires. For example, Jenson (1953) used standardized tests and grade point average (GPA) to predict student achievement. One often used theory in these studies is the constructivists learning theory, which proposes that learning is based on an active process of constructing knowledge rather than just acquiring it. Based on this theory Vermunt (1998) found that for realising constructive, high-quality learning, the control of the learning process should be transferred from the teacher to the students.

With the advancement of computers and internet, the field entered a whole new era. The adoption of learning management systems (LMSs) to assist courses resulted in new and more data available. This changed the field from theory driven analyses to more data driven analyses, with the main focus on using LMS data. LMSs can be used for online content creation, communication, assessment, and administration (Piña, 2012). A variety of commercial academic learning management systems are available, including Blackboard, Angel, Desire2Learn, and Pearson eCollege as well as open source LMSs, including Moodle and Sakai. All these systems record every click, resulting in a rich pool of (raw) data. LMSs are used for fully online as well as blended learning courses. Blended learning is a combination of a face-to-face course with e-learning, where a significant amount of the course is presented online (Hoic-Bozic, Mornar, & Boticki, 2009). Thus, with blended courses, not all behaviour is monitored in the LMS, as there is also offline behaviour, for example in lectures. However, even with fully online courses not all behaviour is monitored, as students can for example download materials and read them offline, or use other offline or online communication platforms to contact their peers.

Thus, data from LMSs cannot give a complete overview of all behaviour, but it can provide a significant amount of information about students and their learning processes without intervention. Additionally, LMS data show actual (online) behaviour of all students, compared to questionnaires which only consist of self-reports on behaviour, learner dispositions, and abilities, and of only students who participated in the questionnaire. Because of these advantages, more and more researchers started using LMS data. As LMSs became more in use, the amount of data available increased extensively, which made it harder and more time consuming to analyse the data. Improvements in data mining techniques in other fields made it possible to deal with those large amounts of data and to conduct more advanced analyses (Clow, 2013). Both the adoption of LMS data and advancements in data mining techniques led to an increased interest in the field, the advent of the term ‘learning analytics’, and the development of the subarea academic analytics and the adjacent field educational data mining.

2.1 Academic Analytics

Learning analytics is a subarea of academic analytics. Academic analytics not only focusses on the usage of LMS and student administration data for improving teaching and learning, but for improving all decision-making processes in educational institutions. Goldstein (2005) was the first to use the term academic analytics "to describe the intersection of technology, information, management culture, and the application of information to manage the academic enterprise" (p. 2). In the beginning most institutions used these analytics for recruitment strategies to improve the enrolment processes. Nowadays more and more institutions also use it to improve teaching, learning and student success, i.e. for learning analytics (Agudo-Peregrina et al., 2014; Campbell, DeBlois, & Oblinger, 2007). Some current challenges in academic analytics are: possibility of oversimplification of the data, issues with privacy, obligation to act, and more skilled staff is needed (Campbell et al., 2007; Goldstein, 2005).

2.2 Educational Data Mining (EDM)

The field of learning analytics shows quite some overlap with educational data mining. The goal of educational data mining (EDM) is to better understand how students learn and identify the settings in which they learn, to improve educational outcomes and gain insight into and explain educational phenomena (Romero & Ventura, 2013). The first EDM workshop was held in 2005, followed by its first international conference in 2008 and the Journal of Educational Data Mining in 2009 (Siemens & Baker, 2012). The current topics of interest in EDM include the development of generic frameworks and methods for mining the data, to be able to obtain more general results across studies; educational process mining, based on the processes in LMSs; data-driven adaptation and personalization; and replication studies (Romero & Ventura, 2013).

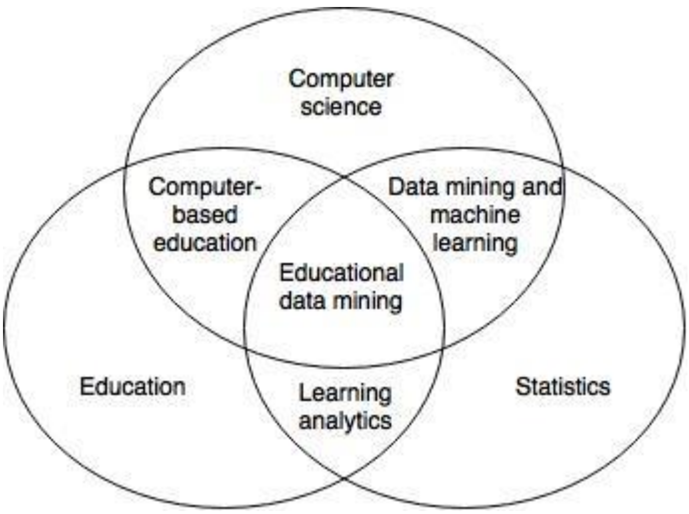

Figure 1: Overview of fields related to Educational Data Mining and Learning Analytics (Romero & Ventura, 2013, p. 13).

Hence, both learning analytics and educational data mining focus on improving learning and teaching. Romero and Ventura (2013) describe EDM as a combination of the fields of computer science, education and statistics, with the subareas computer-based education, data mining and machine learning, and learning analytics (Figure 1). However, other literature shows that a distinction between learning analytics and educational data mining is not made that easy. Multiple papers try to compare and contrast the fields and come up with different comparisons (Romero & Ventura, 2013; Siemens & Baker, 2012). The most notable difference between researchers using the term 'educational data mining' compared to researchers who use the term 'learning analytics' is that EDM is mostly used by computer scientists. This results in more advanced data mining techniques used, and a focus on comparing these techniques, automated discovery, and automated adaption as intelligent tutoring systems in EDM (cf. Siemens & Baker, 2012). On the other hand, learning analytics is mostly conducted by researchers from social sciences, resulting in models which are primarily used for informing educators about their decision-making processes and improve their teaching, rather than about automated adaptations in teaching and student feedback (cf. Siemens & Baker, 2012).

Thus, both learning analytics and educational data mining have similar goals, but somewhat different methods to achieve these goals. Even though the fields show quite some overlap, there is little communication and collaboration between the fields. Sharing findings and collaboration could benefit both fields, and therefore Siemens and Baker (2012) argue that more communication between the fields is necessary. Accordingly, in the current literature review, we will not only discuss current challenges and research goals in learning analytics, but also include some relevant studies in educational data mining.

2.3 Goals of learning analytics

Research in the field of learning analytics can be categorized by a large amount of goals and tasks, but little consensus has been reached on these categorizations. Romero and Ventura (2007) analysed papers in educational data mining from 1995-2005 and grouped these papers based on task in two categories: 1) statistics and visualization, and 2) web mining. Three year later the same authors conducted a literature review on 304 papers and categorized these in eleven task categories: 1) analysis and visualization of data, 2) providing feedback for supporting instructors, 3) recommendations for students, 4) predicting student's performance, 5) student modelling, 6) detecting undesirable student behaviour, 7) grouping students, 8) social network analysis, 9) developing concept maps, 10) constructing courseware, and 11) planning and scheduling (Romero & Ventura, 2010). Baker (as cited in Baker & Yacef, 2009) classified the trends in EDM into five categories: 1) prediction, 2) clustering, 3) relationship mining, 4) distillation of data for human judgement, and 5) discovery with models. Clow (2013) distinguished five methods in learning analytics: 1) predictive modelling, 2), social network analysis, 3) usage tracking, 4) content/semantic analysis, and 5) recommendation engines.

Thus, most categorizations mention predictive models, visualization and some way of actually using the findings for recommendations or human judgement. These are also the categories of studies we found in the consulted literature. Currently most research is focussed on predictive modelling of student success (Baker & Yacef, 2009; Romero & Ventura, 2010; Shum & Ferguson, 2012). A significant amount of literature focusses on learning analytics tools, to facilitate the analyzation and visualization of the data. Lastly, an upcoming topic is based on successfully implementing learning analytics to improve learning and teaching.

3 Predicting student success

Most learning analytics focus on predicting student success. Student success is often quantified by final grade, or whether the student passed a course or not. Data used for predictive modelling can come from the learner itself, such as student characteristics, demographics and dispositions, the course, and the learning management system used.

3.1 Learner data

Most analytics on learner data fall into the field of learning analytics, and only a small number of studies can be categorized as educational data mining. This is because data mining techniques are new, compared to the use of learner data for predicting student success. Moreover, learner data does often not offer enough data for these complex data mining techniques in comparison to the vast amount of data available in learning management systems.

Studies on learner data influencing student success have resulted in a stable set of variables found influencing academic performance. The most important and robust predictors of student success are ability, measured by tests such as SAT and ACT, and past performance, quantified with past GPA (Bipp, Kleingeld, & Schinkel, 2013; Conard, 2006; Dollinger et al., 2008; Superby, Vandamme, & Meskens, 2006). However, ability and GPA cannot account for all variability in student success. Especially in higher education they have less predictive power, as the range of intelligence scores get restricted. Therefore, researchers also tested other predictors, cited in literature as 'non-cognitive predictors' (O'Connor & Paunonen, 2007). These non-cognitive predictors have been grouped into trait and state variables. Trait variables are non-controllable and relatively stable in a person over time. State variables are under control of the student; they can change over time due to practice, training or different contexts. Even though trait variables can often explain large amount of variance in students' results, researchers emphasize the importance of the state variables as these can actually be changed by students to improve their success.

3.1.1 Trait variables

Several trait factors have been found to be important in predicting student success, although not all of them are stable predictors. Personality is stable variable known to be useful in predicting student success. Personality is frequently tested with the Big Five traits of personality: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism. Especially conscientiousness is found to be a stable predictor. A meta-analysis of papers using the Big Five traits as predictors showed that the mean correlation between performance and conscientiousness and performance was $r = 0.24$ (O'Connor & Paunonen, 2007). The mean correlations of the other factors were considerably lower: openness to experience $r = 0.06$, extraversion $r = -0.05$, agreeableness $r = 0.06$, and neuroticism $r = -0.03$.

Contrary, sex is found to be an unstable predictor of academic success, with women being more successful than men in some cases (Bipp et al., 2013; Kotsiantis, Pierrakeas, & Pinteas, 2004; Van den Berg & Hofman, 2005), and no significant effects in other (Bipp et al., 2013; Superby et al., 2006). Secondly, being older than average as well as having children is found to have a negative effect on performance (Kotsiantis et al., 2004; Superby et al., 2006). Finally, educational level of one's parents is found relevant for study success, but only for immigrants (Kaufman, Agars, & Lopez-Wagner, 2008). For natives, educational level of parents does not have a significant effect (Superby et al., 2006; Van den Berg & Hofman, 2005).

3.1.2 State variables

Various state variables are found to have a significant effect on student success. Dollinger et al. (2008) showed that class attendance is significantly correlated with exam scores ($r = 0.38$). Class attendance also increased the amount of variance explained in exam scores when added as an explanatory variable next to next to the uncontrollable factors GPA and verbal ability. Superby et al. (2006) also found that class attendance was positively related to academic success ($r = 0.25$). Conard (2006) found that attendance, partly mediated by conscientiousness, was positively related to GPA.

Motivation and time management are also positively correlated with student success (Bipp et al., 2013; Britton & Tesser, 1991; Kaufman et al., 2008). Kaufman et al. (2008) found that intrinsic and extrinsic motivation could explain an additional 6% of variance in student success next to GPA and parental educational level. Britton and Tesser (1991) found that time attitudes, i.e. the feeling that you are in charge of how your time is spent, could account for 15% of the variance in GPA, and short-range planning for an additional 6%. Long-range planning was not found to have a significant influence. Superby et al. (2006) found that perceptions of the environment and the academic context did not have a significant influence on academic success. However, students who felt they had made a thorough decision for what university they wanted to go to did receive a higher average grade ($r = 0.18$).

Overall, research showed state and trait variables combined could account for 16% (Kaufman et al., 2008), 20-30% (Bipp et al., 2013), 36% (Britton & Tesser, 1991) and 43% (Dollinger et al., 2008) of the variance in student success. Classification techniques could accurately classify up to 57% of high risk students (Superby et al., 2006).

3.2 Course data

Next to learner data, course design and scheduling characteristics have also been pointed as possible predictors for student success. These variables are not controllable at the student level, but only at the institutional level and might be easier to change to improve overall performance. However, Van den Berg and Hofman (2005) found that most variance is explained at the student level, and that course and scheduling characteristics had almost no influence. Only passive education (e.g. lectures) resulted in a significantly lower study progress compared to active education (e.g. seminars, workshops). Rienties, Toetenel and Bryan (2015) tested data of 87 courses and showed that the course itself does not have a big impact on performance. Constructivists modules, with a high proportion of assimilative learning activities, such as reading content, had a negative influence on completion and pass rates. All other learning activities did not have a significant relation with performance.

3.3 LMS data

LMS data is analysed in a variety of studies, using different types of LMSs, blended or fully online courses, and different techniques. These analyses range from relatively simple multiple linear and logistic regression models to more complex and advanced data mining techniques. An overview of the studies, techniques and LMSs used to predict student success can be found in Appendix A.1.

3.3.1 Regression and classification

One of the first predictive modelling studies using LMS data focussed on the evaluation of an online environment which was used to support an offline course (Rafaeli & Ravid, 1997). The online environment included an online textbook, review questions, a search engine, homework assignments, and a discussion board. Rafaeli and Ravid (1997) analysed online behaviour of three classes, with a total of 178 undergraduate and graduate students. Final grade, which was based on paper tests and homework assignments, was regressed on the amount of pages read and online review questions grade. They found that 22.2% of the variance in final grade could be explained by online usage. This is quite far away from an accurate prediction, but is still a high amount when you take into account that a large proportion of students read the materials offline, about one third of the students used usernames and passwords from other students on occasion, and about half of the students did not use internet prior to the course. This all restricts the reliability of the predictor variables.

Yu and Jo (2014) analysed data from of 84 students in the open source LMS Moodle, who participated in the blended course ‘Understanding of science and public administration’. A multiple linear regression was on final grade with the independent variables: total login frequency, total time online, regularity of study interval, number of downloads, interactions with peers and with instructor. Total time online and interaction with peers correlated significantly with final grade. In total all six predictor variables accounted for 33.5% of variance in final grade.

The same amount of variance explained is found in the prediction of final grade in fully online courses in both eCore and Blackboard LMS. Morris, Finnegan and Wu (2005) analysed data in eCore for 354 students in three courses: ‘English Composition II’, ‘Introduction to Geology’, and ‘U.S. history to 1865’. In total eight variables were used for analysis, using both frequency and duration variables; time spent on and number of: content page views, discussion posts read, original posts, and follow-up posts. *t*-Tests showed for all these variables that withdrawers had significant lower frequencies and less time spent online than completers. Thus, as could be expected, withdrawers showed less participation than completers. Multiple regression with the 284 completers showed that 31% of the variability in final grade was accounted for by the number of discussion posts and content pages viewed, and time spent on viewing discussion posts.

Macfadyen and Dawson (2010) used data from 118 students who completed a fully online 'Biology' course in Blackboard. Of the 22 variables tested, 13 variables had a significant positive correlation with final grade. With backwards stepwise regression, a final model was constructed which could account for 33% of variance in student grade using the total number of discussion messages posted, mail messages sent, and assessments completed. A binary logistic regression was conducted to show the reliability of the model for predicting students 'at risk'. The authors showed that 38 out of 63 students with high chance of failure were accurately predicted as 'at risk', and 49 out of 65 successful students could be accurately predicted as 'not at risk', resulting in an overall classification accuracy of 74%.

Zacharis (2015) could even predict 52% of the variance in student success, using Moodle data from a blended course of Computer Science using Java with 134 students. Out of the 29 variables tested, 14 had a significant correlation with course grade. Stepwise regression resulted in the final model consisting of: number of messages read and posted via email, chat or on the discussion forum (combined in one variable named *REPO*); content creation contribution (*CCC*), including the creation of content in the class wiki and site blog; quiz engagement, including the number of quiz attempts, quiz continue attempts and quiz close attempts; and files viewed. Zacharis (2015) proved the robustness of the model with a 10-fold cross-validation. Binary logistic regression showed that out of 43 students who failed, 30 were correctly predicted as 'fail', and out of 91 students who passed, 79 were correctly predicted as 'pass', resulting in an overall classification accuracy of 81%.

Other studies used more complex classification methods to classify students, based on whether they are likely to pass the course or not. Minaei-Bidgoli and Punch (2003) compared six classification techniques for classifying students in 2, 3 or 9 classes based on their final grade. Data was collected from 227 completers in one course in LON-CAPA LMS and included the variables: number of correct answers, getting problem right on the first try, number of tries, time spent on problem until solved, time spent on the problem regardless of it was solved, and participating in communication. In total 72% of the students passed the course. The maximum prediction accuracy was 82% for 2-class prediction, so 10% higher than when they would just predict that everyone would pass. The accuracy was 60% for 3-class prediction and 43% for 9-class prediction. A combination of the classification techniques yielded a better prediction accuracy (87%, 71% and 51%, for 2, 3 and 9 classes, respectively).

Romero, Espejo, Zafra, Romero and Ventura (2013) compared 21 classification techniques and 10-fold cross-validation with 438 students in seven courses. To classify whether a student passed the course or not, they used nine variables: course, number of assignments submitted, passed quizzes, failed quizzes, discussions posted, discussions read, time spent on assignments, quizzes and forum. The highest percentage of correctly classified students as pass or fail found was 65%.

Thus, more complex classification techniques will not always result in higher accuracy in the classification. Also, most complex techniques are not easy interpretable, which makes it harder to use the results for improving learning and teaching. When more complex techniques are desired, Romero et al. (2013) suggest to use decision trees, rule indication and fuzzy rule algorithms, as these provide the best interpretable results.

Although all these studies report how well the regression or classification model performs, this is not always an useful metric. It might be more insightful to know how far away the predictions is from the true value, on average, or how the classification accuracy is away from a baseline, such as just predicting that everyone will pass. This could for example give more insight in whether the model could be used for automated assessment. It would therefore be useful if future work would include more metrics to get a better understanding of the outcomes.

3.3.2 Other analyses

Next to regression and classification, some studies only described correlations between LMS usage and final grade or *t*-tests. A large scale study on 2,674 Blackboard courses with 91,284 students and 40 Moodle courses with 1,515 students showed that there was a positive correlation between the number of clicks and final grade in both learning management systems (Beer, Clark, & Jones, 2010). Dawson, McWilliam and Tan (2008) analyzed the quantity of the online sessions and the time per session of 1,026 students in one course. They showed with *t*-tests that low and high performing students did not differ in time per session, but low performing students attended fewer online sessions than high performing students. Milne, Jeffrey, Suddaby and Higgins (2012) analysed LMS data of 658 students in 9 blended courses. Only data from the first week of the course was analysed. Students were grouped in no LMS usage, 1-5 page views, 6-20 page views and more than 20 page views. Milne et al. (2012) found that usage of the LMS in the first week of the course was significantly higher for successful students than for students who failed the course.

As most LMSs offer a discussion forum, a substantial amount of research can be found on the quantitative analyses forum usage. Davies and Graff (2005) collected discussion forum usage data in Blackboard from 122 students in six courses. Final grade was grouped into fail, low, medium, and high passing grade. Kruskal-Wallis test showed the trend that more activity increased the likelihood of better performance and students who failed showed a consistently lower usage of the forum. However, for only one of the six courses a significant difference in discussion forum usage was found between students who passed the course and students who failed. Nandi, Hamilton, Harland and Warburton (2011) analysed discussion forum usage in Blackboard of 645 students in two courses. They found a trend that high achieving students participated more in the forum than other students. However, only 40% of the students participated in the forum, indicating that the forum might be a more useful predictor when it is used by a high proportion of the students. Dawson and colleagues

(2008) found a significant effect of discussion forum usage on final grade. They analysed discussion forum usage in Blackboard in a course of 1,026 students. A *t*-test showed that students who made at least one post in the forum scored 8% higher on average than students who posted nothing at all. Network analysis in Netdraw on forum relationships of 118 students showed that low-performing students had a small student interaction network, mainly consisting of low-performing peers, while high-performing students had a dense network, comprised of more high-performing peers (Macfadyen & Dawson, 2010).

3.4 Combining learner data with LMS data

Thus, both LMS data and learner data have shown to predict at least some of the variance in student success. Data in learning management systems is collected real-time and each click is recorded, and might be more extensive and accurate than learner data (Campbell & Oblinger, 2007). On the other hand, self-reports provide a higher order of information about someone's state or intentions, compared to raw LMS event logs (Shum & Crick, 2012). Interestingly, after the introduction of LMSs most researchers started from scratch and only focused on online behavioural data, while ignoring the previous findings from learner data. Comparing LMS data with learner data might give insight about the usefulness of both predictors. Additionally, combining research on LMS data with social sciences might result in more accurate predictions, as learner data can give more detailed and timely information (Shum & Crick, 2012).

However, the combination of behavioural data with learner data to predict student success is rare. Some studies did supplement LMS data with basic background information, such as age, gender and prior education or prior GPA (Tempelaar, Heck, Cuypers, van der Kooij, & van de Vrie, 2013), but most do not state any statistics about the influence of these variables (Arnold & Pistilli, 2012; Beer et al., 2010). Hence, it cannot be determined whether background variables are of any added value next to the LMS data. It could be reasoned that the background variables, prior academic data and demographics, cannot be controlled by the student and are therefore not useful as indicators for students and teachers on how to improve student success (Yu & Jo, 2014). Other learner data, such as motivation and time management might be of more value, as these could be influenced. To stimulate research in this area, Shum and Crick (2012) proposed a theoretical framework for combining learning dispositions, values and attitudes with online behavioural data.

Based on this framework, Tempelaar et al. (2013) combined online data from two test-directed environments with demographics, entry test data, self-reports on culture, learning style and emotions. Regressions were done on data of 1832 students in two courses. Tempelaar et al. (2013) showed that prior education and entry test were significant predictors, and therefore these variables were controlled for in later analyses. The most important predictor of academic performance was the level reached in the online test environment. Culture was found to have an impact on the amount of practice: masculinity and hedonism had a stronger influence on the intensity of practicing, than on the outcomes of practising. Students with the stepwise learning style practiced more often and had a better performance than other students. External regulated students benefitted most from practising while students with behavioural lacking regulation practiced longer and more, but achieved less. Lastly, positive emotions had a positive influence on performance, while negative emotions had a negative impact. Unfortunately no statistics were mentioned, so the effects cannot be compared.

Tempelaar and colleagues (2015a) replicated the study from Tempelaar et al. (2013), and added data from Blackboard LMS and motivation and engagement questionnaires. The predictive power of all sources was analysed and compared. Linear hierarchic regression on 873 Mathematics and Statistics students showed that behaviour in the two test-directed environments could best predict performance (R = 0.51-0.66). Especially behaviour in the week before the course started had the highest predictive power, and week 3 seemed to be the best compromise between early feedback and high predictive power. Furthermore, entry test could predict performance (R = 0.41-0.45), followed by motivation and engagement (R = 0.27-0.34), and learning styles (R = 0.21-0.25). Interestingly, LMS data played just a minor role: only the number of clicks was a significant predictor, and all LMS data combined could explain a marginal 4% of the variance in performance. This low percentage of variance explained could be due the fact that most behaviour occurred in the e-tutorials, while the LMS was marginally used. Thus, which such rich data available, LMS data might not be of an added value. However, more studies need to compare the predictability of learner data versus LMS data to be able to draw conclusions about whether LMS data or learner data is more useful for predicting success, and whether the added value of combining the data is significant enough to warrant the extra time needed to collect and combine the data.

3.5 Predicting LMS behaviour

Both LMS behaviour and learner data are shown to have predictive power in modelling student success. It can be argued that LMS behaviour mediates the relation between learner data and student success. For example, students who have a higher motivation might make more use of the LMS and therefore receive higher grades. Because of this learner data is also used to predict LMS behaviour (Iglesias-Pradas, Ruiz-de-Azcárate, & Agudo-Peregrina, 2015). Iglesias-Pradas and colleagues (2015) used a questionnaire to measure the competencies commitment and teamwork of 39 students. LMS data was collected from Moodle. They showed that commitment and teamwork could not significantly predict LMS usage. This could be due to the low variety in the scores on these competencies, as commitment and teamwork were measured with six questions on a four point scale. As all participants also worked as teachers, they probably already had acquired these skills. It is thus likely that (almost) all scored on the upper half of the scale. Future research with a larger sample size and a less experienced group needs to be conducted to find out whether learner data can predict LMS behaviour.

Course design and participation of the teacher have been shown to influence LMS behaviour. Rienties et al. (2015) used correlation and three different clustering techniques to compare the learning design and its impact on LMS behaviour and performance of 87 courses in Moodle. Four learning design clusters were distinguished: constructivist, assessment-driven, balanced-variety and social constructivist. Of 32 courses with a total of 19,322 students, the number of visits to the LMS and the average time spent in the LMS were measured and aggregated per week. Rienties et al. (2015) found that communication activities had a positive effect on LMS visits and time spent in the LMS, while assessment activities had a negative effect. Beer et al. (2010) found that participation of the instructor in the discussion forum increased the amount of clicks in the LMS.

3.6 Conclusion

The studies above show that there is a wide variety in the studies on LMS data. Especially the predictor variables used show a great diversity (see Appendix A.2). Also, the regression models show a high variety in explained variance in final grade, including 4% (Tempelaar et al., 2015a), 22% (Rafaeli & Ravid, 1997), 31% (Morris et al., 2005), 33% (Macfadyen & Dawson, 2010), 33.5% (Yu & Jo, 2014), and even 52% (Zacharis, 2015), see Appendix A.1. Thus, there is no consistency in the methods and predictors used, but also the findings show a vast diversity.

The differences in predictor variables used can be explained by the fact that not all researchers have access to all variables in the LMS. Also, different courses and institutions can use different tools in the LMS. The differences in the findings can be explained by the multiple predictors used, but even when similar predictor variables are used, they are not always robust. For example, Morris et al. (2005), and Macfadyen and Dawson (2010) found a significant positive correlation between discussion forum posts and final grade, while Zacharis (2015) did not find a significant correlation. Another explanation for the different outcomes is that most studies only describe special cases, where the outcomes only apply to a specific institution, course, or group of students. For example when in a specific course the discussion forum is rarely used, it will be highly likely that it is a bad predictor, as there will be low variance in this predictor. On the other hand, in courses which regularly use the forum, it can be a very good predictor.

The case studies are useful for the institution or course itself, but of less value for other institutions and no general conclusions for the field of learning analytics can be drawn. Therefore, several researchers have tried to create frameworks for analysing LMS data to make comparison of the results between different studies easier. As research and theories suggest that interaction with instructors and peers has an important influence in education, most frameworks are based on interaction in the learning management system. Petropoulou, Retalis, Siassiakos, Karamouzis and Kargidis (2008) propose a framework including 1) outcomes: quantitative and qualitative, 2) types of interaction: learner-content, instructor-learner, and learner-learner, and 3) the effectiveness of the applied pedagogical model for building and maintaining a collaborating community. Agudo-Peregrina and colleagues (2014) used these types of interaction in their model for analysing Moodle data. The LMS data was classified on types of interaction, frequency of use: most, moderately and rarely used, and participation mode: passive and active. The classification was tested on data of eight courses, of which six fully online, with 20-30 students per course. The classification did not result in any significant predictive model for student success.

Other researchers based their framework on the different components in LMSs (Rankine, Stevenson, Malfroy, & Ashford-Rowe, 2009). The different components identified were content, communication, collaboration, assessment, and explicit learner support. This classification was shown useful for benchmarking activity in LMSs across two universities in Australia, using different versions of Blackboard LMS, with a sample of 10% of the courses. With the framework, the authors were able to find that in each component a similar amount of activity was found across both universities. The framework is however not (yet) used for grouping predictor variables for predicting student success.

Thus, a lot of different predictor variables are used and found significant. Using frameworks in these studies would structure the analyses, but the few studies using a framework did not find any significant results yet. Further development and testing of the frameworks, and replication studies using the same predictor variables in different courses and institutions, are needed to gain a better understanding about which factors in online learning influence academic performance. This is also why Clow (2013) and Romero and Ventura (2013) claim that learning analytics is not a mature discipline yet, which does facilitate rapid development, but lacks coherency.

4 Learning analytics tools

As log data can be large, relatively information poor, can have a lot of irrelevant entries, and as most educators lack extensive statistical background, learning analytics tools are made to help educators with processing the raw LMS log data (Zaïane & Luo, 2001). Additionally, visualization tools are developed to help instructors with interpreting this data. A selection of the available tools is described below.

4.1 Analytics tools

The Academic Analytics Tool (AAT) performs complex analytical queries with the use of an SQL editor on data of any LMS (Graf, Ives, Rahman, & Ferri, 2011). The tool focusses mainly on behaviour of students in relation to a learning object, such as the discussion forum, quizzes, or learning material. Teachers can specify what information they want from which courses (or group of courses), learning objects and time span. In this manner educators can more easily extract useful information out of the log data and analyse the relation between students' behaviour and learning objects.

AnalyticsTool also helps educators to extract useful information out of the log data, but is especially focussed on interaction patterns in Moodle (Petropoulou et al., 2008). The tools stores the interaction patterns in case-by-case matrices. The interaction patterns are based on the interaction framework described in section 3.6. The tool can report the following indicators: actor's degree centrality, work amount, argumentation, collaboration, average number of contributions, participation, and number of messages. With these indicators instructors can easily analyse interaction patterns in statistical programs.

CosyLMSanalytics focusses on learning paths of students in Moodle (Retalis et al., 2006). The tool uses input from web analytics tools, automatically gathers this data and analyses the learning patterns. It provides correlations among students' learning paths and the data can be used to cluster the learners using SPSS. The tool also provides ways to analyse discussion forum usage qualitatively, as the teacher can annotate the messages based on the content and use these annotations in their analyses.

Zaïane and Luo (2001) made a tool which implemented several data mining algorithms. They used association rule mining to discover correlations between online activities; sequential pattern mining for analysing the sequences of activities; and clustering for grouping learners with similar behaviour. With the tool educators can set constraints and use the algorithms, without knowledge of the algorithms needed. Zaïane and Luo (2001) tested the tools with association rule mining algorithm in two experiments using data from 100 students in two courses. It was shown that the tool was useful to extract which pages are often visited together, which can provide useful insight for the educators in terms of recommending activities or structuring the content.

4.2 Visualization tools

Next to analytics tools, visualization tools have been used to support educators with interpreting the data and results. An often cited visualization tool is Netdraw, which is used for analysing the social network and relationships in discussion fora (Dawson et al., 2008; Macfadyen & Dawson, 2010; Retalis et al., 2006). CourseVis is a graphical student monitoring tool, used in web-based courses (Mazza & Dimitrova, 2007). The tool provides graphical representations of data in three aspects: social (interactions), cognitive (performance) and behavioural (attendance, progress). The effectiveness of the tool was tested with a focus group (N=5), experiments (N=6), and a questionnaire (N=6). It was found that teachers could gain information faster and with a higher accuracy using the CourseVis tool than with textual explanations only. However, there was some confusion when graphs were rotated, variables were missing or too many variables were displayed. In these cases visualizations were difficult to understand and not really useful.

Visualization tools are also made for students to inform them about their study progress. These tools are often referred to as dashboards. According to Clow (2013), one of the most prominent dashboards used is Course Signals. Course Signals is a plugin on Blackboard which provides feedback to the students (and educators) in the form of a traffic light on the homepage, which indicates whether students are at risk (Arnold & Pistilli, 2012). The feedback is calculated by student success algorithm, which is based on performance (available to date), interaction compared to peers, prior academic history and student characteristics. Next to the feedback, educators can send personalized emails to encourage students. Although it is not reported whether the algorithm accurately predicts student success, the tool is shown to be successful in providing feedback. An evaluation of this application on nearly 24,000 students showed that students who used Course Signals got higher grades, and the earlier students used the application in their academic career, the better their performance. Moreover, it was shown that their motivation was positively influenced, but the students were negative about the number of messages received and would like to have more specific information about their progress.

4.3 Conclusion

Although these tools are a good start, they are often still too complex to use for educators and non-experts (Romero & Ventura, 2007). For faculty it is important that data is presented in contextualized ways (Macfadyen & Dawson, 2012). This could be done by including the tool into the e-learning environment, and generating dashboards such as Course Signals to better inform educators and students about their progress (Arnold & Pistilli, 2012; Graf et al., 2011; Romero et al., 2013). It would also be useful to empirically test the tools with educators from different departments and institutions and with data from different LMSs (Macfadyen & Dawson, 2010; Retalis et al., 2006). These experiments can help improving the user interface and further development of the tool. If tools are currently evaluated, the evaluations are usually small and not generalizable (Dyckhoff, Lukarov, Muslim, Chatti, & Schroeder, 2013). Often, the evaluations are focussed on the outcomes of the tools and analyses, instead of the effectiveness in informing the teacher and students, which is done by Course Signals (Arnold & Pistilli, 2012). Thus, more empirical tests especially focussing on the user experience are needed. To stimulate the development of the tools, the tools must move outside the universities and become open source and freely available (Romero et al., 2013). Eventually, tools can be used to automatically intervene to enhance student retention, motivation and learning success (Graf et al., 2011).

5 Implementing learning analytics

Increasingly, adaptive hypermedia systems, adaptive LMSs, and recommendation systems are used to improve the learning environment based on the data of the student (Hoic-Bozic et al., 2009; Romero & Ventura, 2010). However, to successfully implement learning analytics, some challenges should be taken into account. Successful implementation of learning analytics should lead to improvement in learning and teaching. According to Greller and Drachsler (2012) to successfully implement learning analytics six dimensions need to be covered in the design. First of all, stakeholders need to be identified. Campbell and Oblinger (2007) identified five stakeholder: faculty, students, executive officers, student affairs and IT. Secondly, it is important to identify their objectives, as these could differ between the stakeholders. Next to that, educational data is needed in useful data formats and instruments for analysing this data. Lastly, there are external constraints such as ethics and privacy, and internal limitations such as competences and acceptance. Additionally, it is important to be open about learning analytics being conducted, without giving students the feeling that they are monitored all the time (Clow, 2013). Also, resources should not only be directed to students who have a high chance of failure.

Next to coping with these challenges, the implementations need to be evaluated, which has rarely been done until now (see also section 4.3). The reflection of the instructor on improvement of pedagogical practice is often omitted as well (Dawson et al., 2008). Action research is focussed on the reflective process of testing whether learning analytics is actually successfully implemented and indeed improved learning and teaching (Dyckhoff et al., 2013).

Dyckhoff et al. (2013) conducted meta-analysis on case studies of the German eLearning conference. They argued that the questions from teachers need to be included in learning analytics, as this can inform whether these questions were answered and if they had an impact on learning and teaching. Teachers questions were grouped into qualitative evaluation, quantitative measures of use and attendance, differentiation between groups or course offerings, data correlations, and effects on performance. The meta-analysis showed that many of these questions still remain unanswered, especially the qualitative questions. To answer all questions, more data sources are needed to identify the whole learning process. Especially teacher data can be useful to identify whether teacher activities had an influence on learning and teaching.

Macfadyen and Dawson (2012) showed that learning analytics will not always be able to lead to pedagogical changes. They analysed LMS data of 18,909 courses and found that LMSs had a positive value in supporting student learning, as the use of the discussion boards and the grade and content and grade views were positively related to final grade. However, this did not lead to more discussion in order to extend the usage of technology in the educational institution. Thus, although learning analytics can be very informative and provides a lot of opportunities, implementation will not always improve learning and teaching.

6 Conclusion and future work

In the current paper we provided an extensive literature review on learning analytics. Learning analytics mainly focusses on predicting student success, the development of analytics and visualization tools, and the integration of learning analytics. Within these topics, a wide variety can be found in the tools, techniques and data used. Standardization of data and methods is needed to be able to compare the results more easily (Romero & Ventura, 2007). The emergence of public directories is a step into the right direction, as this makes it easier to externally validate data (Baker & Yacef, 2009). However, replication studies and further developments of frameworks are still needed to draw more general conclusion about improving learning and teaching. Moreover there are still a lot of open questions in the field. Thus, there are enough interesting venues and opportunities for extending research in the field of learning analytics and get a better understanding of how to improve learning and teaching.

Structure studies predicting student success. First of all, replication studies are necessary which use the same predictors of student success over a large amount of courses in multiple educational institutions, including different course designs and different groups of students. This can give more insight in which variables are robust predictors, and which are only significant in special occasions. This could also explain the different amount of explained variances found in different studies (see Appendix A.1). To get more structure in the variables used, it would also be useful to further test the frameworks proposed by Rankine et al. (2009) and Petropoulou et al. (2008). In this manner, an argumentation can be given of which variables should be tested and combined, instead of just testing all variables available.

Compare methods used for predicting student success. Researchers in learning analytics mostly use multiple regression to predict student success, while educational data mining researchers use more complex data mining techniques. Although different data mining techniques are often compared (Romero et al., 2013), they are almost never compared to the more simpler regressions. Future work should compare these simpler techniques with the more advanced data mining techniques, in multiple cases. In this way, it could be determined whether complex data mining techniques actually result in better predictions, and in which cases it is useful to take this extra effort.

Predicting student success with qualitative analyses. Most current studies focus on quantitative analyses, while qualitative analyses are often omitted (Davies & Graff, 2005; Dyckhoff et al., 2013; Nandi et al., 2011). Future work should include quantitative analysis, especially in the usage of the discussion forum. This can give more insight in the type of participation of students and might therefore be more useful for predicting student success (Davies & Graff, 2005; Nandi et al., 2011). This could also give more insight in students who show high participation but receive low grades (Morris et al., 2005).

Predicting student success with behaviour outside LMS. The behaviour in LMS cannot explain all variance in final grade. This is possible due to the fact that not all behaviour occurs in the LMS, especially with blended courses, there is also a lot of offline behaviour. Therefore, it would be useful to also analyse behaviour outside the LMS, for example class attendance (Agudo-Peregrina et al., 2014) or communication in other online tools. As behaviour inside the LMS and outside the LMS behaviour might correlate highly, it should also be tested whether the extra effort needed for collecting data outside the LMS is worth it, i.e. does lead to a significant better prediction.

Predicting student success with LMS data and learner data. Although learner data and LMS data are both useful predictors for students success, both are rarely combined. Tempelaar et al. (2015a) did compare learner data with LMS data and found LMS data could only explain 4% of the variance in academic success. This is however not a reason to discard online behaviour data in future studies, as others have shown that online behaviour can explain up to 52% of the variance in academic success (Zacharis, 2015). Replication studies, using a more homogenous student sample and courses with different designs should be conducted to verify whether LMS data has indeed low predictive power compared to learner data, and whether LMS data has additional (unique) predictive power next to learner data.

Predicting student success over time. It can be expected that the predictions using LMS data improve over time, since more data becomes available in the LMS. It can however be considered more useful to be able to predict student success and especially student failure early in the semester, in order to be able to intervene in a timely fashion (Campbell & Oblinger, 2007). Learner data may be more useful at this time. Therefore, future work should investigate how the prediction of student success using LMS data evolves over time, and which factors are especially useful in the beginning of the course.

Predicting student success with midterm data. A lot of courses also use assessments during the course. These grades might also be good predictors of student success, as it is a direct measurement of what students are supposed to learn during a course. Therefore, it should be analysed whether LMS data and learner data have any remaining value in predicting student success as soon as in-between assessments have taken place.

Predicting LMS usage with learner data. Another interesting venue for future research is predicting LMS usage from learner data. Iglesias-Pradas and colleagues (2015) tested the influence of the competencies commitment and teamwork on LMS usage, but used a small sample with not much variation in the predictor values. Therefore, replication studies are needed with larger sample sizes, showing more variation in the predictor values. Also, other learner data can be used to predict LMS usage in general, and usage of specific modules in the LMS. For example, future work could include the influence of the personality type extraversion on the amount of discussion forum posts, or the personality type conscientiousness on the regularity of the study interval.

Predicting LMS usage with teacher data. Dyckhoff et al. (2013) argues to include more teacher data in analysis. This data could also be used to predict LMS usage based on the behaviour of the teacher. For example, Beer et al. (2010) found that the participation of the teacher in the discussion forum increased the amount of clicks in the LMS. Future work should analyse whether other types of teacher activities have an influence on LMS usage as well.

Improve visualization and analytics tools. Analytics and visualization tools provide a lot of room for development and improvement. For teachers, the tools need to be more flexible and user friendly (Romero, Ventura, & García, 2008; Zaïane & Luo, 2001). It is useful to integrate the tools into the e-learning environment (Romero & Ventura, 2010), and thereby contextualize the data to help interpretations (Macfadyen & Dawson, 2012). The integration should also support decisions or give recommendations (Retalis et al., 2006; Romero et al., 2013). The dashboards for students must have better features to monitor progress (Hoic-Bozic et al., 2009), be more customizable (Macfadyen & Dawson, 2010), and motivate behavioural change (Macfadyen & Dawson, 2012). The tools should dynamically update navigation (Rafaeli & Ravid, 1997) and generate automated interventions to improve student retention and motivation (Graf et al., 2011).

Test visualization and analytics tools. For all the improvements of the tools it is important to empirically test the tools with multiple educators and different LMSs in realistic scenarios, see also section 4.3 (Petropoulou et al., 2008; Retalis et al., 2006). Different formats and types of learner feedback should be tested to determine the preferences and sensitivities of the students to these types of feedback. (Tempelaar, Rienties, & Giesbers, 2015b).

Evaluating implementation of learning analytics. Lastly, future work should include action research, the evaluation of implementations of learning analytics. In this way, it could be analysed if learning analytics lead to pedagogical changes and whether it indeed improved learning and teaching.

7 References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). 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. *Computers in Human Behavior*, *31*, 542–550. http://doi.org/10.1016/j.chb.2013.05.031

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). New York, NY, USA: ACM. http://doi.org/10.1145/2330601.2330666

Baker, R. S. J. d, & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM - Journal of Educational Data Mining*, *1*(1), 3–17.

Beer, C., Clark, K., & Jones, D. (2010). Indicators of engagement. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference*. Retrieved from http://cms.ascilite.org.au/conferences/sydney10/procs/Beer-full.pdf

Bipp, T., Kleingeld, A., & Schinkel, S. (2013). Bachelor Entrance Study (BEST), Onderzoek naar studiesucces en drop-out binnen de Bachelor opleidingen aan de faculteit Industrial Engineering & Innovation Sciences aan de TU/e.

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, *83*(3), 405–410. http://doi.org/10.1037/0022-0663.83.3.405

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic Analytics: A new tool for a new era. *Educause Review*, *42*(4), 40–57.

Campbell, J. P., & Oblinger, D. G. (2007). Academic Analytics. Educause.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality*, *40*(3), 339–346. http://doi.org/10.1016/j.jrp.2004.10.003

Davies, J., & Graff, M. (2005). Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, *36*(4), 657–663.

Dawson, S., McWilliam, E., & Tan, J. P.-L. (2008). Teaching smarter: how mining ICT data can inform and improve learning and teaching practice. *Graduate School of Medicine - Papers (Archive)*, 221–230.

Dollinger, S. J., Matyja, A. M., & Huber, J. L. (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality*, *42*(4), 872–885. http://doi.org/10.1016/j.jrp.2007.11.007

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting action research with learning analytics. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 220–229). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2460340

Goldstein, P. J. (2005). Academic Analytics: The uses of management information and technology in higher education. *Educause*, 1–12.

Graf, S., Ives, C., Rahman, N., & Ferri, A. (2011). AAT: a tool for accessing and analysing students' behaviour data in learning systems. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 174–179). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2090145

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*(3), 42–57.

Hoic-Bozic, N., Mornar, V., & Boticki, I. (2009). A Blended Learning Approach to Course Design and Implementation. *IEEE Transactions on Education*, *52*(1), 19–30. http://doi.org/10.1109/TE.2007.914945

Iglesias-Pradas, S., Ruiz-de-Azcárate, C., & Agudo-Peregrina, Á. F. (2015). Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies. *Computers in Human Behavior*, *47*, 81–89.

Jenson, R., E. (1953). Predicting scholastic achievement of first-year graduate students. *Educational and Psychological Measurement*, *13*, 322–329.

Kaufman, J. C., Agars, M. D., & Lopez-Wagner, M. C. (2008). The role of personality and motivation in predicting early college academic success in non-traditional students at a Hispanic-serving institution. *Learning and Individual Differences*, *18*(4), 492–496. http://doi.org/10.1016/j.lindif.2007.11.004

Kotsiantis, S., Pierrakeas, C., & Pinteas, P. (2004). Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence*, *18*(5), 411–426. http://doi.org/10.1080/08839510490442058

Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. http://doi.org/10.1016/j.compedu.2009.09.008

Macfadyen, L., & Dawson, S. (2012). Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan. *Educational Technology & Society*, *15*(3), 149–163.

Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, *65*(2), 125–139. http://doi.org/10.1016/j.ijhcs.2006.08.008

Milne, J., Jeffrey, L. M., Suddaby, G., & Higgins, A. (2012). Early identification of students at risk of failing. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference* (Vol. 2012). Retrieved from http://www.ascilite.org.au/conferences/wellington12/2012/images/custom/milne,_john_-_early_identification.pdf

Minaei-Bidgoli, B., & Punch, W. F. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In *Genetic and Evolutionary Computation—GECCO 2003* (pp. 2252–2263). Springer. Retrieved from http://link.springer.com/chapter/10.1007/3-540-45110-2_119

Morris, L. V., Finnegan, C., & Wu, S.-S. (2005). Tracking Student Behavior, Persistence, and Achievement in Online Courses. *Internet and Higher Education*, *8*(3), 221–231. http://doi.org/10.1016/j.iheduc.2005.06.009

Nandi, D., Hamilton, M., Harland, J., & Warburton, G. (2011). How active are students in online discussion forums? In *Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114* (pp. 125–134). Australian Computer Society, Inc. Retrieved from http://dl.acm.org/citation.cfm?id=2459952

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990. http://doi.org/10.1016/j.paid.2007.03.017

Petropoulou, O., Retalis, S., Siassiakos, K., Karamouzis, S., & Kargidis, T. (2008). Helping educators analyse interactions within networked learning communities: a framework and the AnalyticsTool system. In *Proceedings of 6th International Conference on Networked Learning* (pp. 5–7). Retrieved from http://nlc.ell.aau.dk/past/nlc2008/abstracts/PDFs/Petropoulou_317-324.pdf

Piña, A. A. (2012). An overview of learning management systems. In *Virtual Learning Environments: Concepts, methodologies, tools and applications* (First edition, pp. 33–51). Sullivan University System, USA: IGI Global.

Rafaeli, S., & Ravid, G. (1997). Online, web-based learning environment for an information systems course: Access logs, linearity and performance. In *Information Systems Education Conference (ISECON)* (Vol. 97, pp. 92–99). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.9119&rep=rep1&type=pdf

Rankine, L., Stevenson, L., Malfroy, J., & Ashford-Rowe, K. (2009). Benchmarking across universities: A framework for LMS analysis. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference* (pp. 815–819). Retrieved from http://www.ascilite.org.au/conferences/auckland09/procs/rankine.pdf

Retalis, S., Papasalouros, A., Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). Towards Networked Learning Analytics–A concept and a tool. In *Proceedings of the 5th International Conference on Networked Learning*. Retrieved from http://nlc.ell.aau.dk/past/nlc2006/abstracts/pdfs/P41%20Retalis.pdf

Rienties, B., Toetenel, L., & Bryan, A. (2015). "Scaling Up" Learning Design: Impact of Learning Design Activities on LMS Behavior and Performance. In *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (pp. 315–319). New York, NY, USA: ACM. http://doi.org/10.1145/2723576.2723600

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. http://doi.org/10.1002/cae.20456

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. http://doi.org/10.1016/j.eswa.2006.04.005

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *40*(6), 601–618. http://doi.org/10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. http://doi.org/10.1002/widm.1075

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. http://doi.org/10.1016/j.compedu.2007.05.016

Shum, S. B., & Crick, R. D. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92–101). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2330629

Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society*, *15*(3), 3–26.

Siemens, G. (2011). Announcing Open Course: Learning and Knowledge Analytics. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. Retrieved from https://tekri.athabascau.ca/analytics/

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2330661

Superby, J. F., Vandamme, J.-P., & Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 37–44). Taiwan.

Tempelaar, D. T., Heck, A., Cuypers, H., van der Kooij, H., & van de Vrie, E. (2013). Formative Assessment and Learning Analytics. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 205–209). New York, NY, USA: ACM. http://doi.org/10.1145/2460296.2460337

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015a). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, *47*, 157–167. http://doi.org/10.1016/j.chb.2014.05.038

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015b). Verifying the Stability and Sensitivity of Learning Analytics Based Prediction Models, an Extended Case Study. In *Proceedings of the 7th International Conference on Computer Supported Education* (pp. 1–18). Lisbon, Portugal.

Van den Berg, M. N., & Hofman, W. H. A. (2005). Student Success in University Education: A Multi-Measurement Study of the Impact of Student and Faculty Factors on Study Progress. *Higher Education*, *50*(3), 413–446.

Vermunt, J. D. (1998). The regulation of constructive learning processes. *British Journal of Educational Psychology*, *68*(2), 149–171. http://doi.org/10.1111/j.2044-8279.1998.tb01281.x

Yu, T., & Jo, I.-H. (2014). Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education. In *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (pp. 269–270). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2567594

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, *27*, 44–53. http://doi.org/10.1016/j.iheduc.2015.05.002

Zaïane, O. R., & Luo, J. (2001). Towards evaluating learners’ behaviour in a Web-based distance learning environment. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies* (pp. 357–360). http://doi.org/10.1109/ICALT.2001.943944

Appendix A

Table 1: Overview of studies using LMS data to predict student success

	LMS	Number of students	Number of courses	Techniques used	Regression Accuracy (R^2)	Classification Accuracy (pass/fail)
(Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014)	Moodle	356	8	Classification of predictors, Multiple Regression		
(Beer, Clark, & Jones, 2010)	Moodle + Blackboard	91,284 + 1,515	2,674 + 40	Correlation		
(Davies & Graff, 2005)	Blackboard	122	6	Kruskal-Wallis test		
(Dawson, McWilliam, & Tan, 2008)	Blackboard	1,026	1	*t*-Test		
(Macfadyen & Dawson, 2010)	Blackboard	118	1	Correlation, Multiple Regression, Logistic Regression, Network Analysis	33%	74%
(Milne, Jeffrey, Suddaby, & Higgins, 2012)	N/A	658	9	Chi-squared test		
(Minaei-Bidgoli & Punch, 2003)	LON-CAPA	227	1	Classification (6 techniques)		87%
(Morris, Finnegan, & Wu, 2005)	eCore	284	3	Correlation, Multiple Regression	31%	
(Nandi, Hamilton, Harland, & Warburton, 2011)	Blackboard	645	2	Graph (visualization)		
(Rafaeli & Ravid, 1997)	OnLine	178	3	Multiple Regression	22%	
(Romero, Espejo, Zafra, Romero, & Ventura, 2013)	Moodle	438	7	Classification (21 techniques)		66%
(Tempelaar, Rienties, & Giesbers, 2015)	Blackboard	873	2	Correlation, Hierarchical Multiple Regression	4%	
(Yu & Jo, 2014)	Moodle	84	1	Multiple Regression	34%	
(Zacharis, 2015)	Moodle	134	1	Correlation, Multiple Regression, Logistic Regression	52%	81%

Table 2: Predictor variables used in LMSs

		(Agudo-Peregrina et al., 2014)	(Beer et al., 2010)	(Davies & Graff, 2005)	(Dawson et al., 2008)	(Macfadyen & Dawson, 2010)	(Milne et al., 2012)	(Minaei-Bidgoli & Punch, 2003)	(Morris et al., 2005)	(Nandi et al., 2011)	(Rafaeli & Ravid, 1997)	(Romero et al., 2013)	(Tempelaar et al., 2015)	(Yu & Jo, 2014)	(Zacharis, 2015)
Total	Number of online sessions				X	X								X	X
	Regularity of study interval													X	
	Number of LMS hits	X[1)]	X										X[2)]		X
	Number of LMS hits first week						X						X		
	Time per session				X										
	Total time online					X							X[2)]	X	X
General	Number of course page views														X
	Number of announcement views					X									
Content	Number of downloads													X	
	Number of files viewed					X									X
	Number of resources viewed														X
	Numbers of links viewed					X									X
	Number of content page views								X		X				
	Time spent on viewing content pages								X						
Forum	Number of discussion posts views			X		X			X	X		X			X
	Total number of discussion posts				X	X		X		X		X			X
	Number of original posts					X			X						
	Number of follow-up posts					X			X						
	Number of follow-up posts to peers													X	
	Number of follow-up posts to instructor													X	
	REPO messages (email, chat, discussion forum)														X
	Time spent on forum								X			X			
	Time spent on creating original posts								X						
	Time spent on creating follow-up posts								X						
Mail	Mail messages read					X									X
	Mail messages sent					X									

Chat	Number of chat posts														X
	Number of chat views					X									X
Quiz / Assessment	Quiz engagement (start, continue and close attempt)														X
	Number of quizzes started					X		X							X
	Number of quizzes continued														X
	Number of quizzes passed					X		X				X			X
	Number of quizzes failed											X			
	Number of quizzes right at first try							X							
	Number of quiz views														X
	Number of quiz reviews														X
	Quiz grades										X				
	Time spent on quizzes until solved							X				X			
	Time spent on (un)solved quizzes					X		X							
Assignment	Number of assignments read					X									X
	Number of assignments submitted					X						X			X
	Time spent on assignments					X						X			
LMS tool	Uses of 'compile' tool					X									
	Uses of 'search' function					X									
	Visits to MyGrades tool					X									X
	Visits to MyProgress tool					X									
	Uses of the 'who is online' viewer					X									
Wiki	Number of wiki edits														X
	Number of wiki views														X
	Number of wiki add pages														X
Blogs	Number of blog updates														X
	Number of blog views														X
	Content creation contribution (Number of wiki and blog edits)														X

1) categorized on interaction type
2) grouped per week

The value of LMS data for predictive modeling of student performance

A study on 17 blended courses using Moodle LMS

M.A. Conijn
Eindhoven University of Technology
The Netherlands
m.a.conijn@student.tue.nl

C.C.P. Snijders
Eindhoven University of Technology
The Netherlands
c.c.p.snijders@tue.nl

ABSTRACT

With the adoption of Learning Management Systems (LMSs) in a large number of educational institutions, a lot of data has become available on students' online behavior. Many researchers have used students' LMS data to predict student performance. Unfortunately, this has lead to a rather diverse set of findings, possibly because the courses under study are diverse in many ways. The educational context in which the courses take place varies, just as the kind of use that is being made of the LMS, to name just a few. After providing a brief overview of recent findings, we analyze a larger sample of 17 blended courses with 4,989 students in Moodle LMS, in which we predict student performance from LMS and intermittent assignment grades based on the same kind of variable constructions as encountered in the literature. Our analyses show that, irrespective of the fact that all courses were taught at the same institution, the results of predictive modeling depend heavily on course characteristics. Our analyses also show that when in-between assessment grades are taken into account, LMS data are of small additional value. For early intervention, when such grades are not yet available, our LMS data are shown to be a rather weak predictor. To improve the prediction of student performance, especially for early intervention, more data need to be included than can be easily inferred from LMS logs.

Categories and Subject Descriptors

K.3.1 [**Computer and Education**]: Computer Uses in Education

Keywords

Learning Analytics, Learning Management Systems, Moodle, Predictive Modeling

1. INTRODUCTION

The emergence of ICT into higher education has significantly changed the way in which teachers teach and students learn. Using internet for providing content has given the possibility to transform face-to-face courses into blended courses even fully online courses. In blended courses a significant amount of information is presented online and can be accessed at any place, at any time [7], while in fully online courses all information is presented online. Nowadays, a vast majority of institutions make use of the internet in their teaching, often through Learning Management Systems (LMSs), also known as Virtual Learning Environments (VLEs) [15]. LMSs can support student learning by providing content online, and by allowing for additional benefits such as quizzes, presentations and screencasts, assignments, and forums [13]. LMSs allow teachers to provide such content in a relatively easy and integrated way.

Whereas face-to-face courses can provide teachers with direct feedback about their teaching through the direct observation of their students, this direct feedback is not available in online systems. However, as every click is monitored and stored, LMSs do provide very precise, but also abstract logs of students' online behavior. Learning analytics focuses on interpreting and contextualizing this information to improve learning and teaching, to increase student performance, and to detect at-risk students, that is, students who have a high probability of failing the course [1].

Although learning analytics with LMS data is a relatively recent research topic, a wide variety of studies can be found within this interdisciplinary field. Currently, most of the empirical studies are focused on predictive modeling of student performance [2, 16, 18]. In these studies, various analytical methods are used to establish which factors influence student performance in a specific institution, for a specific course or set of courses. Issues that make it hard to compare the substantive findings of these studies are both the variety of factors that can be (and are) extracted from the log data and the different predictive methodologies that have been used. In fact, many findings seem to contradict one another. A related issue is that 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 [4]. In the current study, we intend to overcome part of the abovementioned issues, by first providing a brief overview of recent findings in the learning analytics field, and furthermore analyzing the predictive value of LMS data in a set of 17 blended courses taught at the same institution (Eindhoven University of Technology). This allows us to establish effects of different sorts of LMS usage while controlling to some extent for contextual effects. In addition, we analyze whether it is possible to identify students at-risk early on in a course, and to what extent the LMS usage data supplements the information from in-between measurements of performance.

2. PREVIOUS RESEARCH

2.1 Predicting student performance

Most studies on learning analytics focus on predicting student performance. Student performance is often quantified by final grade, or by whether the student passed a course or not. Data used for predictive modeling can come from different sources such as student characteristics, dispositions and demographics, but in recent years most often data from LMSs are used [17, 19]. Studies analyzing LMS data show a wide variety in types of LMS used, courses examined (blended or fully online), and analytical techniques used. Most of the studies are exploratory in nature, in the sense that they use a lot of different predictor variables in only a few courses. The exploratory nature and the different methods used make it hard to compare the different studies and draw general conclusions about what the best and most stable predictors are for predicting student performance.

Rafaeli and Ravid [14] were one of 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. They found that students who were inexperienced with using online systems tended to stick to a page-by-page reading order, while experienced students adopted a non-linear style. Linear regression using the amount of pages read and the grades for online questions as predictor for the final grade showed that 22% of the variance could be explained. Likewise, Morris, Finnegan and Wu [11] found that number of content pages viewed was also a significant predictor in three fully online courses in eCore with 354 students. Contrary to Rafaeli and Ravid [14] they used a total of eight duration and frequency variables, and no in-between measurement of performance. Multiple regression analyses with these predictors on final grade of the 284 completers showed that 31% of the variability in final grade was accounted for by the number of discussion posts and content pages viewed, and time spent on viewing discussion posts. Moreover, they found by using t-tests that withdrawers had a significant lower frequency of activities and less time spent online, compared to completers. Macfadyen and Dawson [8] also found that the amount of content viewed, in their case measured by links and files viewed, had a positive correlation with final grade. However, these variables were not included in their final model. Like Morris et al. [11], a fully online course was analyzed, but using another LMS: Blackboard. Multiple regression analyses found that 33% 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%, 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. Thus, only the number of discussion posts was found in both final prediction models of Macfadyen and Dawson, and Morris et al. [8, 11]. The usage of the discussion forum was found important for predicting student performance in several other studies as well. In analysis of discussion forum usage in Blackboard in a course of 1,026 students, a significant effect was found of discussion forum usage on final grade [6]. A t-test showed that students who made at least one post in the forum scored 8% higher on average than students who posted nothing at all. However, Nandi, Hamilton, Harland and Warburton [12] were not able to find 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% of the students participated in the forum, indicating that the forum might be a more useful predictor when it is used by a high proportion of the students.

Thus, the relation between discussion forum usage and final grade in fully online courses using Blackboard showed some different results. Other researchers using LMS Moodle in blended courses found that discussion posts and interactions with peers were significant correlated with final grade. Yu and Jo [20] analyzed data of 84 students. Six variables were tested: total log in frequency, total time online, regularity of study interval, number of downloads, interactions with peers, and interactions with instructor. Total time online and interaction with peers correlated significantly with final grade, and all predictor variables accounted for 34% of variance in final grade. Using the same LMS with 134 students in one course, Zacahris [21] could even explain 52% of variance in student performance, using three predictors. Contrary to Yu and Jo [20], 29 variables were analyzed of which 14 correlated significantly with final grade. Total time online and amount of files and links viewed were found to have a significant correlation with final grade, but as in Macfadyen and Dawson [8], these were not included in the final model for predicting student performance. Only three predictors were included: the number of messages read and posted via email, chat or on the discussion forum (combined in one variable named REPO); content creation contribution (CCC), including the creation of content in the class wiki and site blog; quiz engagement, including the number of quiz attempts, quiz continue attempts and quiz close attempts; and files viewed. Classification resulted in an overall accuracy of 81%, where 30 out of 43 students who failed were correctly predicted as fail, and out of 91 students who passed, 79 were correctly predicted as pass.

The studies above show that there is a wide variety in the studies on LMS data. Especially the predictor variables used show a great diversity. The differences in predictor variables used can be explained by the fact that not all researchers have access to all variables in the LMS. Also, different courses and institutions can use different tools in the LMS. This lack of availability and access to the data might also explain why the studies are mostly data-driven instead of theory-driven: almost no theory can be found about why certain variables are included. The differences in the findings can be explained by the multiple predictors used, but even when similar predictor variables are used, they are not always robust. Despite of the different predictor variables used, still a reasonable amount of variance in final grade could be explained, including, ranging from 22% [14], 31% [11], 33% [8], 34% [20], and even 52% [21]. However, recently Tempelaar, Rienties and Giesbers [19] found that the amount of clicks in Blackboard of 873 students in two blended courses could only explain a marginal 4% of the variance in student performance on the final exam.

Another explanation for the different outcomes is that most studies only describe special 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. Beer, Clark and Jones [3] did conduct a large scale study on 2,674 Blackboard courses with 91,284 students and 40 Moodle courses with 1,515 students. They showed that there was a positive correlation between the number of clicks and the final grade in both learning management systems [3]. Unfortunately, only correlational analysis of the amount of clicks with the final grade was conducted, so no general conclusions could be drawn on other prediction variables.

The different methods and relatively small samples that have been used make it hard to compare the different studies and draw general conclusions about what the best and most stable predictors are for predicting student performance. Moreover, it poses the question how, in general, LMS data should be used for predictive modeling. Therefore, we chose to conduct a somewhat larger scale study (17 courses using Moodle LMS at the same institution), using the different predictor variables as have been used in recent work (if available).

2.2 Predicting student performance over time

Most studies that have tried to predict student performance analyze the behavior of students in the LMS during the whole course. This allows researchers to answer the question to what extent it is possible to infer study success from LMS data but at a point in time where interventions are no longer possible [4]. Several but not many researchers have acknowledged this issue and decided to analyze potentially predictive data from early stages in a course. For instance, Milne, Jeffrey, Suddaby and Higgins [9] analyzed LMS data of the first week of a course for 658 students in 9 blended courses. Students were grouped into no LMS usage, 1-5 page views, 6-20 page views, and more than 20 page views. Milne et al. found that usage of the LMS in the first week of the course was significantly higher for successful students than for students who failed the course. Using a more elaborate setup, Tempelaar et al. [19] analyzed the prediction of student performance over time, using six types of predictors: demographics, entry test performance, learning dispositions, Blackboard LMS data, e-tutorial data, and quiz data. The amount of clicks in the week before the course started (week 0) had the highest predictive power. As the course progressed, the prediction of student performance gradually improved. Assessment data from the quizzes were shown to be the best predictor, but these data are typically only available after a couple of weeks. Indeed, a notable improvement in predictive power was found in the week where the first assessment data became available. The authors therefore argued that as soon as possible after the first assessment would the best compromise between early feedback and sufficient predictive power.

In line with these studies, we analyzed how the prediction of student performance evolves over time. Contrary to Tempelaar et al. [19], a less specific and data rich environment is used, with a more homogeneous set of students, to be able to draw more general conclusions. LMS data and assessment data are used to examine how the prediction changes over time, whether using only LMS data might be of use for timely intervention, and how the effectiveness of predictions changes after the assessment data has become available.

3. METHOD

3.1 Participants and study context

Data were used 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). Data were gathered from courses where at least 75 students had participated. 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 analysis. In total 1,072 students were excluded (M = 63, SD = 103 per course). Data remained from 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, 1,121 students who were enrolled in 2 courses, 143 students in 3 courses, 147 in 4 courses and 57 in 5 courses. An overview of the courses used and the amount of students per course can be found in Table 1. The courses varied from basic courses that every Bachelor student at TU/e has to take, to specific courses in the fields of mathematics, physics, statistics, and psychology. Most courses belonged to the first year of the undergraduate program (B), but also second year, third year, and some prerequisite courses for entering graduate programs (pre M) were included. Nine courses were taught in the fall quarter from September 1^{st} to November 9^{th}, 2014, and eight courses were taught in the winter quarter from the 10^{th} of November, 2014 to the 1^{st} of February, 2015. The courses consisted of eight weeks of lectures 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. Data from the two weeks long Christmas break were also included, as these fell into the lecture weeks of the winter quarter, making for a total of 13 weeks of LMS data (week 0, 8 lecture weeks, 2 break weeks, and 2 exam weeks). 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 scale from 0 to 10, where grades < 5.5 imply a student does not pass a course and all grades ≥ 5.5 represent a pass. In-between assessment grades included grades for midterms, quizzes, reports, assignments, and homework.

Table 1: Course characteristics

ID	Course name	Quarter	Year B	N
1	Calculus A	1	1	438
2	Calculus B	1	1	1121
3	Calculus C	1	1	227
4	Calculus pre M Architecture	1	pre M	135
5	Set Theory and Algebra	1	1	73
6	Linear Algebra and Vector Calculus	2	2	120
7	Linear Algebra	1	pre M	76
8	Experimental Physics 1	1	1	168
9	Experimental Physics 2	2	1	155
10	Behavioral Research Methods	2	2	136
11	Applied Physical Sciences formal	2	1	836
12	Applied Physical Sciences conceptual	2	1	822
13	Condensed Matter	2	3	74
14	Introduction to Psychology and Technology	1	1	154
15	Linear Algebra 1	1	1	66
16	Statistics	2	2	326
17	The Effectiveness of Mathematics	2	1	62

3.2 Data pre-processing

To align the raw Moodle log data, data were pre-processed using R. 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). The predictor variables were chosen based on the predictor variables that have been used in previous research and their availability in the current data set from the Moodle LMS. An overview of all predictor variables and their descriptive statistics is shown in Table 2. Four basic predictors per course were used: *tot1*: total amount of clicks, *tot2*: number of online sessions, *tot3*: total time online, and *tot4*: total amount of views. Session was defined similarly as in Zacharis [21], 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.
Next to these basic predictors, more complex predictors based on study patterns were included: *pat1*: regularity of study time (*SD* of time per session), *pat2*: regularity of study interval (*SD* of time between sessions), *pat3*: largest period of inactivity, *pat4*: time until first activity, *pat5*: average time per session. Finally, several other variables were computed that relate to the usage of different modules available in Moodle. These variables were adapted from previous research, and were not available for all courses, as not all courses used the same modules in Moodle. Of the 17 courses, 8 made use of additional resources (variable *res1*), 10 used links to external websites (*url1*), all used content pages (*con1*), 1 used a poll (*pol1, pol2*), all used the discussion forum (*for1, for2*), 16 used quizzes (*qui1*-*qui7*), 8 used assignments (*ass1*-*ass3*) and all used the wiki (*wik1, wik2*). For the quizzes, data from the Moodle modules quiz and scorm were combined, as both can provide quizzes; the first ones in Moodle itself, the latter ones in an external source that had been integrated into Moodle. For the assignments, the Moodle modules "assignment" and "workshop" were combined, as both provide the ability to upload an assignment, where the workshop module has the extra option of peer review. If a module was not used in a course, the values were coded as missings to exclude them from analyses.

Next to LMS data, in-between assessment data was used as a predictor variable (*gra1*), which was available for 16 courses. The amount, weight, and type of in-between assessments differed among the 16 courses, 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 week 5, we have analyzed the data assuming that the grades would be available at the end of week 5.

3.3 Data analysis

After data pre-processing, all analyses were run with Stata 14. First of all, a correlation analysis was run on all predictor variables with final exam grade, per course and for all courses combined. Thereafter, multiple linear regressions and binary logistic regressions were run, using stepwise backwards elimination. The criteria for exclusion in each step was $p < 0.1$, for the courses separately and $p < 0.05$ for all courses combined. As the assumption of homoscedasticity was often not met, robust regressions were used. Initially, regressions were run on all courses separately.

Table 2: Predictor variables used in predicting student performance using LMS data

ID	Predictor	N	*M*	*SD*	Used in
tot1	#LMS hits	4989	605	630	[1, 3, 9, 19, 21]
tot2	#Online sessions	4989	30.3	21.2	[6, 8, 20, 21]
tot3	Total time online	4989	4.89e4	4.07e4	[8, 19, 20, 21]
tot4	#Course page views	4989	208	144	[21]
pat1	Regularity of study time	4989	1926	993	[20]
pat2	Regularity of study interval	4989	3.09e5	2.52e5	
pat3	Largest period of inactivity	4989	1.23e6	7.86e5	
pat4	Time until first activity	4989	1.03e6	6.75e5	
pat5	Average time per session	4989	1.63e3	910	[6]
res1	#Resources viewed	2277	7.38	14.2	[21]
url1	#Links viewed	4037	7.81	17.0	[8, 21]
con1	#Content page views	4989	84.6	80.5	[11, 14]
for1	#Discussion posts views	4989	1.92	7.36	[8, 11, 21]
for2	#Discussion posts	2831	.058	.458	[6, 8, 21]
pol1	#Poll answers submitted	136	.705	.711	
pol2	#Poll views	136	1.79	1.34	
qui1	#Quizzes started	2256	22.4	11.9	[8, 21]
qui2	Average #attempts per quiz	4927	1.02	.278	
qui3	#Quizzes passed	4927	9.43	6.91	[8, 21]
qui4	#Quiz views	4927	110	80.3	[21]
qui5	Quiz grades	4927	34.2	35.2	[14]
qui6	Time spent on quizzes	4927	2.33e5	9.92e5	[8]
qui7	Average time per quiz	4927	2.43e4	1.04e5	
ass1	#Assignments submitted	774	1.95	1.36	[8, 21]
ass2	#Assignment submission views	2665	3.79	7.76	[11]
ass3	Time till assignment deadline	220	3.06e4	1.15e5	[11]
wik1	#Wiki views	136	68.0	61.7	[21]
wik2	#Wiki edits	4989	.382	1.04	[21]
gra1	Average assessment grade	4913	6.78	1.96	[19, 14]

As there was overlap between the students, multi-level analyses with crossed random effects were run to determine the amount of variance explained at course and student level. Thereafter, all courses were combined into one regression, using dummy coding with interactions for the courses, and cluster variables for the students. After the regressions with data from the whole course, analyses were run using the data available at the end of every week during the course, to analyze to what extent the prediction changed over time. Robustness of all models was checked with 10-fold cross-validation, using the function "crossfold", which runs ten regressions on subsamples and takes the average of these regressions.
Although most previous studies only report how well the regression or classification model performed in terms of (pseudo) R-squared values, this is not always a very useful metric. It might be more insightful to know how far away the predictions are from the true value, on average, or how much better the classification accuracy than a baseline, such as just predicting that everyone will pass. This could for example give more insight in whether the model could be used for automated assessment. For this reason, we decided to calculate such fit statistics as well.

4. RESULTS

First, we discuss the results of the correlation analyses and regressions on all courses separately. Thereafter, we report on the findings and implications when all courses are analyzed simultaneously. Finally, we show regression analyses on the LMS data as they get available on a week-by-week basis, to determine if early intervention seems possible.

4.1 Predicting student performance

Pearson correlation analyses showed that for all courses combined 26 of the 29 examined variables had a significant correlation with final exam grade (see Table A1). Only regularity of study time, number of discussion posts, and average time per quiz did not correlate significantly. Midterm grade had a large effect size ($r = 0.54$, $p < 0.001$), and the number of wiki views a moderate effect size ($r = 0.43$, $p < 0.001$). All other variables had an effect size below 0.3. The correlations between all variables and the final exam grade for all courses separately shows mixed results. Only midterm grade correlated significantly with final exam grade for all courses (in which it was available). Number of sessions, resources viewed, quizzes started, quizzes passed, and quiz grade correlated significantly in at least 75% of the courses. These variables are the most stable predictors in our sample. Most other predictor variables correlate significantly for 30% to 60% of the courses. Discussion forum and wiki usage had the least amount of significant correlations.

4.1.1 *Predicting final exam grade: separate regressions per course*

To analyze to what extent the variables combined can predict final exam grade, multiple linear regressions were run using the basic and pattern variables. As the correlations showed that there is substantial diversity in available predictor variables per course (not all courses use the same Moodle modules), regressions were run separately for each course. For 16 out of the 17 courses a significant prediction model could be created, with an average R^2 of 0.19 (see Table 3). For one course, all predictors were removed from regression, thus no model could be constructed. A 10-fold cross-validation safeguarded the robustness of the models. Table A2 shows the coefficients and p-values of the variables included in the final models per course. All variables were included in the final models of at least six courses, with the total time online having a significant (partial) correlation most often (11 out of 16 times). However, the direction of the coefficient of time online varied across courses: in some courses time online has a 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 till the first activity have a negative influence in all eight and six final models, respectively. This implies that more general conclusions based on these data should be restricted to these 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 models give some insight in which predictors have an influence on the final exam grade and on the fact that the findings are not consistent across courses. However, with an average 95%-confidence interval around the prediction (averaged across all courses) of (-1.56, +1.56), the models are quite far away from an accurate prediction.

4.1.2 *Predicting final exam grade: regressions for all courses simultaneously*

To determine the variance at the different levels, analyses were run in which all courses are analyzed simultaneously. We saw that in the prediction models constructed per course, the predictor variables often have different influences in different courses. Ordinary least squares regression

Table 3: Final model statistics

ID	Model	p	Mean residual	R^2	R^2 cf
1	F(4,432) = 21.83	<.001	1.63	.17	.17
2	F(9,1111) = 33.58	<.001	1.56	.19	.18
3	F(3,223) = 12.99	<.001	1.69	.17	.19
4	F(3,131) = 12.43	<.001	1.81	.18	.23
5	F(2,70) = 4.66	.002	1.33	.11	.12
6	F(3,116) = 5.04	.003	1.82	.10	.11
7	F(5,70) = 6.73	<.001	1.27	.33	.28
8	F(5,162) = 14.61	<.001	1.50	.18	.16
9	*All variables excluded, no model created*				
10	F(3,132) = 21.96	<.001	1.51	.32	.29
11	F(7,828) = 36.60	<.001	1.32	.23	.22
12	F(6,815) = 19.76	<.001	1.66	.12	.12
13	F(3,70) = 9.47	<.001	1.91	.27	.27
14	F(4,194) = 11.89	<.001	1.10	.16	.21
15	F(2,63) = 6.82	.002	1.36	.17	.25
16	F(6,319) = 8.69	<.001	2.01	.10	.09
17	F(2,59) = 7.44	.001	1.51	.26	.32

on all courses with courses coded as dummies and interaction effects for each course with the other predictors indeed shows that the effects of the predictors depend heavily on the course. For many predictors it can be shown that its coefficient varies substantially and significantly with the course (15 out of 18 predictors exhibit this pattern). However, running standard regressions is an obvious simplification of the true structure of the data. Some students followed multiple courses, thus the cases do not represent unique students. Moreover, the data are of course clustered by course. To take this into account, we ran multi-level regressions with final exam grade as our target variable and with crossed random effects for course and student. In such a model without predictors included, we estimated that 8% of the variance in final exam grade resides at the course level, and 48% at the student level. This implies that in case one wants to analyze all courses simultaneously, one should at least not ignore the student level clustering (and should proceed with caution when ignoring the clustering in terms of courses). On the other hand, it also shows that final grade is to a large extent an individual characteristic, which in principle offers hope to be able to capture this individual variance through LMS characteristics.

4.2 Predicting student performance over time

As it would be useful to be able to timely intervene, we reran our analyses with the Moodle LMS usage variables grouped per week. Only basic predictors (*tot1-tot4*) were used, as usage patterns are often not available or not yet meaningful in the first few weeks. In addition, the variables per module were also excluded from the analyses, as they did not provide enough variability in the first weeks. As only the basic predictors were used, standard regressions with courses coded as dummies and interaction effects for each course with the mean and deviances from the mean within courses as predictors were used. We therefore will report this standard regression, using student clustered standard errors, as this is easier to interpret.

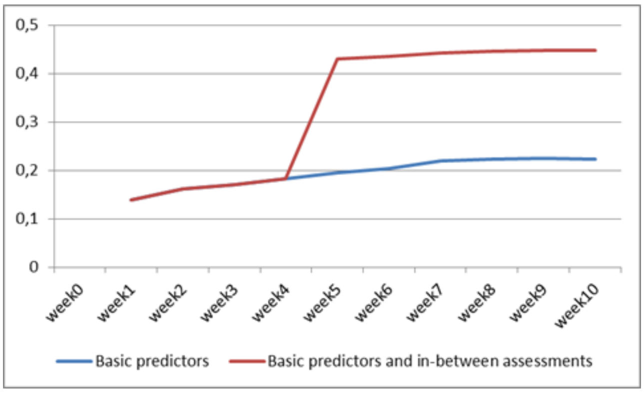

Figure 1: R^2 of linear regression over weeks

4.2.1 *Predicting final exam grade over time*

In the week before the course starts (week 0), all predictor variables were excluded from the step-wise backwards regression, so no valid prediction model could be produced based on LMS data. From week 1 untill the end of the course more data became available, which resulted in an improvement in prediction. However, the prediction only improved slightly from an adjusted R^2 of 0.14 to 0.22 (see Figure 1). However, when in-between assessment grades, available after week 5, were taken into account into the prediction, a larger improvement in prediction was found (week 5: $F(177, 4811) = 22.32$, adjusted $R^2 = 0.43$). With the LMS data of all weeks included, LMS data could explain an extra 2% of the variance in final grade, over and above the average midterm grades. Thus, LMS data, at least the way in which it was implemented here, is of decidedly smaller value than the in-between assessment grades. However, these midterm grades are often not available until the course is halfway, and sometimes not available at all. For early intervention (that is, before assessment grades become available), LMS data can be used, but to predict the final exam grade they are of limited value.

4.2.2 *Predicting pass-fail probabilities*

Fortunately, 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 for determining whether an intervention is needed. Therefore, binary logistic regressions were conducted as in [8, 21] with exam grade > 5.5 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, also binary logistic regressions did not lead to a high accuracy. After week 0, 1548 out of 2704 students who passed were correctly predicted as pass, and 1423 out of the 2266 students who failed where correctly classified as 'at risk'. This lead to a total classification accuracy of 60%. The classification accuracy increased to 69% when all LMS data was used, with 1950 out of 2720 students correctly classified as pass, and 1479 out of 2268 students correctly classified as 'at risk'. The best compromise between early feedback and classification accuracy seems to be after week 3, where 1922 out of 2720 students were correctly classified as pass, and 1397 of the 2268 failing students were correctly classified as 'at risk', resulting in an overall accuracy of 67%.
If we would intervene with students at risk based on this information, still 871 students who fail do not get an intervention. To improve learning and teaching, it might be more useful to intervene as much students 'at risk' as possible, at the cost of intervening with more students who do not need it. To consider this, the specificity (or true negative rate) was changed to 95%. This resulted in 656 out of 2720 students who were correctly classified as successful, and 2158 out of 2268 students who were correctly classified as 'at risk'. Thus, to be able to intervene with 95% of the students who fail, we need to intervene with 85% of the students, where 49% of the intervened students did not need the intervention.
Above binary logistic regressions were run on final exam grade, as in-between assessment grade was a proportion of the final grade. However, as in-between assessment grade was excluded from these models, we could as well run the binary logistic regressions on final course grade. As the final exam accounted for 68% of the final grade on average for all courses, the results were very similar, with only a somewhat higher accuracy for the predictions of final course grade.

5. CONCLUSION AND DISCUSSION

In the current study we investigated how LMS data can be used for predictive modeling of student performance after the course has finished and for predicting student performance during the course, considering 17 different courses at the same institution. Similar to previous research, we used basic predictors and predictors found in the different modules of the LMS. In addition we included more complex variables based on study patterns. First of all, we found that no consensus could be reached on which variables are consistent in predicting 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, showing that the in-between measurement of performance is in line with the measurement of performance at the end of the course. Discussion fora and wiki usage showed significant correlations in the least amount of courses, indicating that these variables might not be the best (or at least most stable) predictors for final exam grade. This unfortunately corroborates previous studies that have also shown different results with respect to which variables helped predict student performance. The differences in previous research could be explained by the fact that these studies varied in the sets of predictor variables, the analytical techniques, and the type of LMS used. As we used the same set of predictor variables and method in 17 blended courses in Moodle in one institution, our hope was that this would lead to more stable results, but apparently this is not the case. Thus, even though contextual effects are kept (more) constant than in the combined literature, nevertheless substantial differences in the size and signs of predictor variables were found.
These differences might be due to the fact that the models are very sensitive to how the courses are organized and how the LMS is used. For example, the low predictability of forum and wiki usage could be due to the fact that variability was low in these modules. As Nandi et al. [12] have argued previously, the discussion forum and wiki might be a better predictor when it is used more extensively. Multi-level analyses and regressions with interaction effects for each course confirmed that the effects of the predictors differ to a great

extent between among the courses. Next to course characteristics, students differ in how they use LMSs while studying. Multi-level analyses indeed show that a high proportion of variance could be explained at the student-level. This offers hope, in principle, that the student characteristic that predicts final exam grade is perhaps the studentâĂŹs usage characteristic, but none of the usage characteristics that have been used in the literature before seem to pick up this variance.

Even though we cannot really use predictive modeling with LMS data to determine which factors have the highest predictive value for predicting student performance, the models can provide us with information about which students are likely to fail, and which not. We found that we can explain up to 43% of the variance in final exam grade when in-between assessment grades are included. As could be expected, our model had a higher performance than a previous study that used only one predictor next to in-between assessment grade [14], and a lower performance compared to a study using four other sources of data next to the LMS data and in-between assessment data [19]. However, when in-between grades are not available, LMS data turned out to be a weak predictor, with an average R^2 of 0.19. Most other studies have been able to explain more of the variance in final grade [11, 20, 8, 21]. Thus, using only LMS data, we are quite far away from an accurate prediction even though we have used the same variables and methods to a very large extent.

The prediction of student performance over time has shown to only increase slightly over weeks, with a serious improvement after week 5, when in-between assessment grades became available. For early intervention, before the assessment data becomes available, this results in only a weak prediction. To be able to intervene at least 95% of the students at risk at the end of week 2, 86% of the students should receive an intervention, half of which would not actually need it. Thus for early intervention, when midterm grades are not available yet, our LMS data are of limited value.

To improve the prediction of student performance over time and at the end of the course, it might be useful to extract more complex variables from the LMS. This is certainly possible and a useful additional attempt, but we would like to highlight that in the current study we have used most variables as they have been used in previous research, so any extension along these lines would be a totally new endeavor. Such new variables could either include more qualitative LMS data, more quantitative LMS data, or data from other sources. Qualitative data, especially from the discussion forum, can give more insight in the type of participation of students and might therefore be more useful for predicting student performance [12, 5]. Additionally, this could give more insight in students who show high participation but receive low grades [11].

Adding more quantitative data also increases the risk of generating an even more diverse set of possible predictor variables. Therefore, we feel that before doing this, it is more fruitful to first establish more fine-grained theories or frameworks that connect LMS usage characteristics to performance. For example, [1] used different types of interactions in learning processes proposed by [10], for analyzing Moodle data. However, these more complex variables might be less useful early in the semester, as there is simply not enough data to be able to generate specific groups with high enough variability for prediction.

Finally, it might be useful to include other sources of data to improve the prediction. These other sources could also improve early prediction when they are available at the beginning of the course. Shum and Crick [17], among others, have already argue that traditional variables from the social sciences, such as learning dispositions, can give more detailed and timely information about the performance of students. While LMS data is a by-product of learner activity, self-disclosure data about dispositions might give a higher order of information about students' state, which is harder to infer from the raw LMS logs [17]. Accordingly, Tempelaar et al. [19] analyzed demographics, entry test, learning dispositions, motivation and engagement, LMS data, e-tutorials, and assessment data in two courses with 873 students. They found that entry test, 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 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. At this moment, we are supplementing our data with such other 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. Both additional theoretical argumentation and additional data sources need to be included to predict student performance and improve learning and teaching.

6. REFERENCES

[1] A. F. Agudo-Peregrina, S. Iglesias-Pradas, M. A. Conde-González, and A. 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. *Computers in Human Behavior*, 31:542–550, Feb. 2014.

[2] R. S. J. d. Baker and K. Yacef. The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM - Journal of Educational Data Mining*, 1(1):3–17, Oct. 2009.

[3] C. Beer, K. Clark, and D. Jones. Indicators of engagement. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference*, 2010.

[4] J. P. Campbell and D. G. Oblinger. Academic Analytics, 2007.

[5] J. Davies and M. Graff. Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, 36(4):657–663, 2005.

[6] S. Dawson, E. McWilliam, and J. P.-L. Tan. Teaching smarter: how mining ICT data can inform and improve learning and teaching practice. *Graduate School of Medicine - Papers (Archive)*, pages 221–230,

Jan. 2008.

[7] N. Hoic-Bozic, V. Mornar, and I. Boticki. A Blended Learning Approach to Course Design and Implementation. *IEEE Transactions on Education*, 52(1):19–30, Feb. 2009.

[8] L. Macfadyen and S. Dawson. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2):588–599, Feb. 2010.

[9] J. Milne, L. M. Jeffrey, G. Suddaby, and A. Higgins. Early identification of students at risk of failing. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference*, volume 2012, 2012.

[10] M. G. Moore. Editorial: Conferences and changes. *American Journal of Distance Education*, 3(3):1–6, Jan. 1989.

[11] L. V. Morris, C. Finnegan, and S.-S. Wu. Tracking Student Behavior, Persistence, and Achievement in Online Courses. *Internet and Higher Education*, 8(3):221–231, Jan. 2005.

[12] D. Nandi, M. Hamilton, J. Harland, and G. Warburton. How active are students in online discussion forums? In *Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114*, pages 125–134. Australian Computer Society, Inc., 2011.

[13] A. Piña. An overview of learning management systems. In *Virtual Learning Environments: Concepts, methodologies, tools and applications*, pages 33–51. IGI Global, Sullivan University System, USA, first edition edition, 2012.

[14] S. Rafaeli and G. Ravid. Online, web-based learning environment for an information systems course: Access logs, linearity and performance. In *Information Systems Education Conference (ISECON)*, volume 97, pages 92–99. Citeseer, 1997.

[15] S. Retalis, A. Papasalouros, Y. Psaromiligkos, S. Siscos, and T. Kargidis. Towards networked learning analytics - A concept and a tool. In *Proceedings of the 5th International Conference on Networked Learning*, 2006.

[16] C. Romero and S. Ventura. Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6):601–618, Nov. 2010.

[17] S. B. Shum and R. D. Crick. Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pages 92–101. ACM, 2012.

[18] S. B. Shum and R. Ferguson. Social Learning Analytics. *Educational Technology & Society*, 15(3):3–26, 2012.

[19] 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. *Computers in Human Behavior*, 47:157–167, June 2015.

[20] T. Yu and I.-H. Jo. Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education. In *Proceedings of the 4th International Conference on Learning Analytics and Knowledge*, pages 269–270. ACM, 2014.

[21] N. Z. Zacharis. A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, 27:44–53, Oct. 2015.

APPENDIX

Table A1: Correlations between dependent variable final exam grade and independent variables for all courses

	all	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
tot1	.04**	.16***	.01	.03	.11	.05	.10	-.16	.15	-.07	.41***	.41***	.32***	.29*	.17*	.08	.15**	.36**
tot2	.21***	.37***	.32***	.29***	.31***	.20	.22*	-.04	.20**	.04	.53***	.41***	.30***	.26*	.26***	.36**	.16**	.44***
tot3	.12***	.24***	.18***	.09	.33***	-.04	.22*	-.04	.12	-.06	.49***	.37***	.29***	.40***	.11	-.04	.04	.20
tot4	.19***	.32***	.23***	.20**	.18*	.22	.15	-.03	.09	-.09	.39***	.41***	.31***	.25*	.15	.27*	.14*	.37**
pat1	.03	-.03	-.04	-.06	.06	-.19	.18*	-.10	-.01	-.09	.05	.31***	.20***	.30*	-.21**	-.17	-.08	-.15
pat2	-.11***	-.33***	-.29***	-.28***	-.19*	-.17	.00	.09	.07	.01	-.33***	-.05	-.02	-.12	-.13	-.27*	-.07	-.35**
pat3	-.06***	-.16***	-.17***	-.32***	-.12	-.12	.06	-.01	.13	-.04	-.31***	.10**	.06	.02	.02	-.25*	.00	-.17
pat4	-.13***	-.15**	-.16***	-.08	-.32***	-.13	-.19*	-.36**	-.29***	-.20*	-.05	-.13***	-.13***	-.25*	-.04	-.06	-.18**	.04
pat5	-.05***	-.06	-.05	-.14*	.02	-.17	-.05	.07	-.04	-.07	.05	.16***	.15***	.06	-.20*	-.22	-.10	-.27*
gra1	.54***	.54***	.48***	.64***	.38***	.27*	.58***	.	.69***	.30***	.52***	.71***	.47***	.76***	.59***	.74***	.59***	.60***
res1	.13***	.	.	.	.	.28*	.00	.	.	.	.40***	.15***	.09**	.21	.23**	.	.	.26*
url1	.09***	.18***	.24***	.21**	.16	-.10	.	.	.05	.03	.	.06	.01	.	.	.	.	.21
con1	.17***	.34***	.28***	.24***	.21*	.25*	.16	.08	.19*	.01	.28***	.37***	.26***	.13	.16*	.23	.09	.40**
pol1	.18*	.	.	.	.	.	.	.	.	.	.18*	.	.	.	.	.	.	.
pol2	.19*	.	.	.	.	.	.	.	.	.	.19*	.	.	.	.	.	.	.
for1	.04**	-.02	.13***	.18**	.18*	.16	-.02	-.02	-.06	-.12	.24**	.07*	.05	.11	-.13	.12	.	.01
for2	.03	-.02	.04	.07	.	.	.	.	.	.	.24**	.	.	.	.	.	.	.
qui1	.11***	.19***	.13***	.03	.42***	.28*	.28**	.23*	.	.	.	.	.	.	.	.04	.	.
qui2	.08***	.	.06*	.	.08	.06	.05	.05	.21**	-.06	-.02	.22***	.15***	.18	-.08	.13	-.06	.
qui3	.20***	.26***	.25***	.12	.46***	.31**	.28**	.25*	.31***	.14	.15	.41***	.30***	.30**	.22**	.03	.19***	.
qui4	.14***	.19***	.12***	.05	.11	.08	.12	-.11	.04	-.12	.23**	.39***	.30***	.27*	.04	.02	.12*	.
qui5	.07***	.10*	.17***	.16*	.14	.11	.21*	.25*	.69***	.30***	.44***	.28***	.27***	.45***	.13	.08	.62***	.
qui6	.03*	.06	.06	-.04	.19*	-.05	.05	-.17	-.04	-.07	.12	.08*	.13***	.15	.09	-.07	.05	.
qui7	.01	-.05	.01	-.05	-.06	-.16	-.05	-.17	-.06	-.11	.03	.06	.06	.13	.03	.05	-.14*	.
ass1	.21***	.	.	.	.	.	.	.	.	.	-.03	.	.	.	.22**	.14	.17**	.18
ass2	.04*	.	.05	.03	.11	.	.	.	.	.	.11	.	.	.	.16	.29*	.16**	.23
ass3	.15*	.	.	.	.	.	.	.	.	.	.	.	.	.	.07	.17	.	.
wik1	.43***	.	.	.	.	.	.	.	.	.	.43***	.	.	.	.	.	.	.
wik2	.04**	-.02	.04	.07	.	.	.	.	.05	.	.27**	.	.	.	.23**	.24	.18***	.25

* p<0.05, ** p<0.01, *** p<0.001

Table A2: Final models multiple linear regression on all courses

	tot1		tot2		tot3		tot4		pat1		pat2		pat3		pat4		pat5		cons	
	b	sig	b	sig	b	sig	b	sig	b	sig	b	sig	b	sig	b	sig	b	sig	b	sig
1	-7.35e-4	.001					5.06e-3	.000			-1.18e-5	.000	1.19e-6	.036	-4.58e-7	.064			6.15	.000
2	-1.16e-3	.000	2.86e-2	.005	-9.57e-6	.046	5.30e-3	.000	-2.53e-4	.078	-9.39e-6	.000	1.06e-6	.006	-8.80e-7	.000	7.85e-4	.010	5.32	.000
3	-2.05e-3	.000					8.46e-3	.000					-1.70e-6	.000					7.42	.000
4					1.09e-5	.077					-1.61e-6	.006			-4.85e-7	.003			4.48	.000
5					-2.93e-5	.026	7.45e-3	.004											5.96	.000
6					1.69e-5	.080			4.18e-4	.023							-3.69e-4	.027	5.04	.000
7	-3.31e-3	.080							-1.29e-3	.028			-1.87e-6	.024	-2.34e-6	.000	1.07e-3	.009	11.38	.000
8	3.99e-2	.005					-4.79e-2	.002			-9.35e-6	.038	1.68e-6	.025	-1.74e-6	.000			7.93	.000
9																				
10			5.91e-2	.000	2.17e-5	.012	-3.56e-3	.003											2.79	.000
11	2.82e-3	.031	1.24e-1	.000	-4.67e-5	.001			6.42e-4	.001	-3.15e-6	.003	1.36e-6	.000			6.15e-4	.056	1.16	.020
12	2.82e-3	.016	4.29e-2	.011	-1.68e-5	.040					-7.49e-7	.070	4.16e-7	.004			3.60e-4	.006	2.95	.000
13	7.89e-2	.005			1.66e-4	.000	-1.06e-1	.002											2.79	.001
14			5.80e-2	.002	5.59e-5	.014	-1.04e-2	.007	-1.37e-3	.003									6.41	.000
15			1.81e-1	.003	-1.31e-4	.027													4.67	.000
16					1.26e-4	.056			-1.27e-3	.063	-6.51e-6	.000	2.91e-6	.002	-8.52e-7	.017	-1.23e-3	.010	4.85	.002
17											-1.48e-5	.001	2.68e-6	.005					6.50	.000

How to succeed?

The effect of different data sources on predicting student performance

Rianne Conijn – 0740635

December 2015

Supervisors:
Prof. dr. C.C.P. (Chris) Snijders
Dr. ir. P.A.M. (Ad) Kleingeld

Abstract

Much research in the field of learning analytics focusses on the predictive modelling of student performance with data from Learning Management Systems (LMSs). Unfortunately, these studies are often exploratory and use different methodologies. This results in different outcomes which are hard to compare. Even when the same method is used, differences are found. Moreover, Gašević and colleagues (2016) found that the prediction models using LMS data even differ per course within one institution. Hence, the portability of the prediction models across courses might be low. Additionally, the studies mostly focus on LMS data only, while ignoring learner data such as ability, personality, and motivation. These variables have been found significant and robust predictors. However, the prediction models using learner data and LMS data have rarely been combined or compared.

In the current study, first, the portability of seventeen blended courses at Eindhoven University of Technology are studied. Contrary to Gašević and colleagues (2016), predictors are used which are available in all courses. It is again found that the portability across courses is low. Thus, although prediction models might be useful for a specific course, data of multiple courses cannot be aggregated. When learner data were added to the LMS data, the prediction models still differed. Thus, adding student characteristics alone is not sufficient for increasing the portability. Exploratory analysis showed that course characteristics do have some effect on the portability, but still differences remain. Hence, one should still be careful when aggregating data from courses with the same characteristics.

Secondly, the predictability of student performance using learner data is compared with LMS data. It is found that learner data outperform LMS data. However, LMS data outperform learner data when in-between assessment data are added to LMS data. The addition of learner data to LMS data is especially useful when in-between assessments are not (yet) available, as is the case with early intervention. However, both sources combined still are not accurate for early prediction of final exam grade and pass/fail probabilities. To improve the early intervention, more data needs to be included.

Contents

1 Introduction 48

2 Previous work 49

2.1 Goals of learning analytics 49

2.2 Predicting student performance 49

2.2.1 Predicting student performance using LMS data 49

2.2.2 Portability of prediction models 51

2.2.3 Predicting student performance using learner data 52

2.2.4 Predicting student performance using learner data and LMS data 53

2.2.5 Predicting student performance over time 54

3 Method study 1: LMS data 55

3.1 Participants and study context 55

3.1 Data pre-processing 55

3.2 Data analyses 56

4 Results study 1 57

4.1 Portability of LMS data 57

4.2 Predictability of student performance using LMS data 57

4.3 Post-hoc analyses 59

4.3.1 Effects between and within students 59

4.3.2 Effects of course characteristics 60

5 Conclusion study 1 60

6 Method study 2: LMS data and learner data 61

6.1 Participants and study context 61

6.2 Learner data 61

6.3 Data analyses 62

7 Results study 2 63

7.1 Preliminary analysis 63

7.1.1 Differences between the subsample and the whole sample 63

7.1.2 Detecting outliers 64

7.2 Portability of learner data 65

7.3 Predictability of student performance 67

7.3.1 Predicting student performance with learner data 67

7.3.2 Predicting student performance with learner data and LMS data: compared 68
7.3.3 Predicting student performance with learner data and LMS data: combined 70
7.3.4 Predicting student performance over time 71
7.3.5 Predicting pass/fail probabilities 72
8 Discussion and Conclusion 73
8.1 Portability of prediction models 73
8.2 Predictability of student performance 75
9 References 78
Appendix A: Questionnaire learner data (Dutch) 83

1 Introduction

The use of technology in educational institutions has brought new opportunities for learning and teaching. Nowadays, a vast majority of educational institutions make use of Learning Management Systems (LMSs), also known as Virtual Learning Environments (VLEs) (Retalis, Papasalouros, Psaromiligkos, Siscos, & Kargidis, 2006). LMSs support student learning by providing content online and by allowing for additional benefits such as presentations, quizzes, assignments, and forums (Piña, 2012). LMSs support teachers by enabling them to provide such content in a relatively easy and integrated way. Moreover, as every action in an LMS is monitored and stored, insight can be gained in students' online behaviour, which in turn can be used to improve learning and teaching. The analysis of LMS data is often referred to as learning analytics (Siemens & Baker, 2012). We define learning analytics as the contextualization and interpretation of behavioural data, contextual data, and learner data, to improve learning and teaching.

Although the field and the term learning analytics are rather new, analysing student data to understand how students learn and to improve learning and teaching has been a topic of research for over decades. Formerly, analyses on student data were mostly conducted using student characteristics, also known as learner data, measured with validated questionnaires. For example, Jenson (1953) used standardized tests and grade point average (GPA) to predict student achievement. With the advancement of computers and internet, the field entered a whole new era. Because LMSs can provide data about all students without intervention needed, researchers started using LMS data instead of learner data to improve learning and teaching.

Currently, much research in the field of learning analytics is focussed on predictive modelling of student performance (Baker & Yacef, 2009; Romero & Ventura, 2010; Shum & Ferguson, 2012). Specifically, to predict students' grades and to predict students who are at risk of failing a course (Gašević et al., 2016). Predictive modelling of student performance is an important step in learning analytics, as it informs the implementation of intervention, such as personalized feedback. Contrary to student characteristics questionnaires, LMSs provide raw log data, not concrete measurements. Thus, the question is how LMS data can be used to predict student performance. To date, most studies use different methodologies with various sets of predictors, generated from the raw log data. Because of these differences, the studies are hard to compare, and the best way to predict student performance remains unknown.

Moreover, the question is if there is actually one best way to predict student performance. When similar methods and predictors are used, studies still found different results in the correlational analysis and prediction models. Thus, the effects of LMS behaviour on student performance might differ per institution or even per course. Gašević et al. (2016) indeed found differences between all models predicting final grade in nine courses within one institution. Thus, the portability of prediction models across courses might not be that high. However, Gašević et al. (2016) used predictors which were related to specific modules in the LMS, which were not available in all courses. Thus, the differences in the prediction models could be explained by the fact that not the same set of predictors was used in every course. Therefore, in our first study we determine the portability of seventeen courses within one institution while using only predictors which are available in all courses.

Contrary to LMS data, learner data such as ability, personality, and motivation have been found significant and robust predictors across courses (e.g., Britton & Tesser, 1991; O'Connor & Paunonen, 2007). Learner data might even be a better predictor for student performance, as it can provide more detailed and timely information (Shum & Crick, 2012). However, the prediction models using learner data and LMS data have rarely been compared. Therefore, our second study aims to find out which source can be best used for predicting student success (at the end of the course and during the course). Moreover, it examines whether these sources explain a unique part of the variance in final grade, and determines the portability of the prediction models using LMS data, when learner data are included.

Thus, in the current paper we explore the value of predicting student performance with LMS data and learner data within a course, and the portability of these models across courses. First of all, we provide a brief overview of the recent findings in predicting student performance, followed by two empirical studies. Study 1 uses LMS data in seventeen courses, and in study 2 learner data is added to the prediction models.

2 Previous work

2.1 Goals of learning analytics

Previous work in the field of learning analytics can be divided into three topics. First of all, much research can be found on the predictive modelling of student performance (Baker & Yacef, 2009; Romero & Ventura, 2010; Shum & Ferguson, 2012). Secondly, a significant amount of literature focusses on learning analytics tools, to facilitate the analyses and visualization of the data (e.g., Dawson, McWilliam, & Tan, 2008; Graf, Ives, Rahman, & Ferri, 2011; Mazza & Dimitrova, 2007; Petropoulou, Retalis, Siassiakos, Karamouzis, & Kargidis, 2008; Retalis et al., 2006; Zaïane & Luo, 2001). Lastly, an upcoming topic is based on successfully implementing learning analytics to improve learning and teaching (e.g., Campbell & Oblinger, 2007; Dyckhoff, Lukarov, Muslim, Chatti, & Schroeder, 2013; Greller & Drachsler, 2012). In the current study we only focus on predicting student performance.

2.2 Predicting student performance

Data used for predictive modelling can come from different sources, including data from LMSs, performance data, and data about the learner itself, such as learning strategies, motivation, beliefs, and demographics. Previously, often learner data were used for predicting student performance, but in recent years data from LMSs are increasingly used (Shum & Crick, 2012; Tempelaar, Rienties, & Giesbers, 2015).

2.2.1 Predicting student performance using LMS data

Even though predicting student success using LMS data is relatively new, already multiple studies can be found in the field. To get an overview of these studies, a literature review was conducted. While some similarities are found, the studies are highly different (see Table 1). Differences are found in the types of LMS used, including OnLine, eCore, Blackboard, LON-CAPA, and Moodle for blended as well as fully online courses. Even within one LMS, different modules are used per course and institution. For example, some courses use blogs, wikis, and quizzes, while others only provided content online. This results in a variety of available predictor variables per course. Often cited predictors include aggregated measures such as the number of clicks in an LMS, the number of online sessions, and the total time online. As many courses use a discussion forum, predictors such as the number of discussion posts, the number of discussion views, and the number of replies on discussion posts were often included as well. Some less regular predictors used are the regularity of study time, the time from handing in an assignment until the actual assignment deadline, the number of quizzes failed, the number of course grade views, and the number of wiki edits.

Table 1: Differences and similarities in studies predicting student performance using LMS data

	LMS	# students	# courses	Analytical techniques	Regression accuracy (R^2)	Classification Accuracy (pass/fail)
(Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014)	Moodle	356	8	Classification of predictors, Multiple Regression		
(Beer, Clark, & Jones, 2010)	Moodle + Blackboard	91,284 + 1,515	2,674 + 40	Correlation		
(Davies & Graff, 2005)	Blackboard	122	6	Kruskal-Wallis test		
(Dawson et al., 2008)	Blackboard	1,026	1	*t*-tests		
(Macfadyen & Dawson, 2010)	Blackboard	118	1	Correlation, Multiple Regression, Logistic Regression, Network Analysis	33%	74%
(Milne, Jeffrey, Suddaby, & Higgins, 2012)	N/A	658	9	Chi-squared test		
(Minaei-Bidgoli & Punch, 2003)	LON-CAPA	227	1	Classification (6 techniques)		87%
(Morris, Finnegan, & Wu, 2005)	eCore	284	3	Correlation, Multiple Regression	31%	
(Nandi, Hamilton, Harland, & Warburton, 2011)	Blackboard	645	2	Graph (visualization)		
(Rafaeli & Ravid, 1997)	OnLine	178	3	Multiple Regression	22%	
(Romero, Espejo, Zafra, Romero, & Ventura, 2013)	Moodle	438	7	Classification (21 techniques)		66%
(Tempelaar et al., 2015)	Blackboard	873	2	Correlation, Hierarchical Multiple Regression	4%	
(Yu & Jo, 2014)	Moodle	84	1	Multiple Regression	34%	
(Zacharis, 2015)	Moodle	134	1	Correlation, Multiple Regression, Logistic Regression	52%	81%

The outcome variable student performance is often quantified by final grade, or by whether the student passed a course or not. This results in different types of analytical techniques needed, such as binary logistic regression and multiple linear regression. Next to regression analyses, also *t*-tests, Kruskal-Wallis tests, Chi-squared tests, correlational analysis, network analysis, and a variety of classification analysis (for classifying the outcome variable as well as the predictors) are conducted.

These different analytical methods result in a variety of accuracy metrics reported. Most often cited measures are Pearsons' *r* for correlational analysis, R^2 for regression accuracy, and the classification accuracy (the proportion of correct classified cases among all cases).

The different methodologies result in a variety of outcomes, which makes it hard to compare the studies. Even when comparing studies which use the same analytical method, the different predictors used result in a wide variety of results. For example, using multiple linear regression, the amount of explained variance in final grade ranged from 4% (Tempelaar et al., 2015), 22% (Rafaeli & Ravid, 1997), to values around 33% (Macfadyen & Dawson, 2010; Morris et al., 2005; Yu & Jo, 2014), and even 52% (Zacharis, 2015). Even when similar predictor variables are used, they are not always robust. For example, Macfadyen and Dawson (2010) and Morris and colleagues (2005) found that the number of discussion post views correlated significantly with final grade, while Zacharis (2015) did not find a significant correlation.

Next to the different methodologies used, the explorative nature of the studies could also explain the different outcomes. Many studies are exploratory in nature, in the sense that they only analyse LMS data of one or a few institutions, using one or only a few courses, or they only describe special cases (such as courses using tailor-made e-tutorial packages). This makes it unclear whether the outcomes apply in general, or to a specific institution, course, or group of students. Beer, Clark and Jones (2010) did conduct a large scale study on 2,674 Blackboard courses with 91,284 students and 40 Moodle courses with 1,515 students. They showed that there is a positive correlation between the number of clicks and the final grade in both types of LMSs (Beer et al., 2010). Unfortunately, only correlational analysis of the amount of clicks with the final grade was conducted, so no general conclusions could be drawn on other prediction variables.

The exploratory nature and the different methods used make it hard to compare the different studies and draw general conclusions about what the best and most stable predictors are for predicting student performance. Moreover, the different outcomes in the studies raise questions on the portability of the prediction models (Gašević et al., 2016; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014): can the same models be used in multiple courses and multiple institutions? With more insight in the portability of the models we could gain better understanding how models developed in one course can be used in another course or institution.

2.2.2 Portability of prediction models

The issue of the portability of the prediction models was already recognized in 2011, when the Open Academic Analytics Initiative (OAAI) was initiated. OAAI has the goal to advance the field of learning analytics by exploring the challenges in scaling learning analytics across all higher education institutions (Jayaprakash et al., 2014). The first two subgoals of this initiative especially focus on the scaling of prediction models: investigating the portability of the models and creating an open source model for predicting student success (Lauría, Baron, Devireddy, Sundararaju, & Jayaprakash, 2012). Lauría et al. (2012) tested the portability of a prediction model for final grade between two institutions: Purdue University and Marist College. Although these institutions differ in institutional type, size, approaches, and type of LMS used (Blackboard vs. Sakai), similarities were found in correlational analysis and the prediction models for final grade (Lauría et al., 2012). All variables analysed (the number of sessions, the number of content views, the number of discussions read, the number of discussions posted, the number of assignments submitted, and the number of assessments submitted) correlated significantly with final grade in both institutions and had a similar effect size. The prediction model used at Purdue University for classifying students as pass or fail, had a 10% lower accuracy on average on data from Marist college (n = 3,877) (Jayaprakash et al., 2014). Hence, the authors stated that the portability of prediction models for student performance might be higher than expected.

However, Gašević et al. (2016) found that the portability across courses in an Australian university is not that high. They analysed and 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, where not all modules were used by every course. As the differences per course could be explained by individual differences, student characteristics, such as age, gender, international student, part time student, and first course, were also included. Multiple linear regression showed that student characteristics could account for 5% of the variance in final grade. The addition of LMS data led to an increase of 16% of explained variance. The models for all courses separately differed from each other and from the generalized model which included all courses. The authors argued that analysing the whole sample might underestimate or overestimate the effects of the predictors. This indicates that the same model cannot always be used for multiple courses and it questions the portability of the models between courses.

These contradicting results show that there is need for more studies investigating the portability of the prediction models. Therefore, in the first study we compare the portability of the prediction models of seventeen courses using Moodle LMS, within the same university, using a set of predictor variables which are available in all courses. Moreover, the predictive power of LMS data for predicting student performance is assessed. Contrary to LMS data, learner data have been found a significant and robust predictor across courses. Therefore, in the second study we invoke learner data and examine and compare the portability and predictability of learner data.

2.2.3 Predicting student performance using learner data

Two of the most important and robust predictors of student performance using learner data are ability, measured by tests such as SAT and ACT, and past performance, quantified with past GPA (Bipp, Kleingeld, & Schinkel, 2013; Conard, 2006; Dollinger, Matyja, & Huber, 2008; Superby, Vandamme, & Meskens, 2006). However, ability and GPA cannot account for all variability in student success. Especially in higher education they have less predictive power, as the range of intelligence scores gets restricted. Therefore, researchers also tested other predictors, which included trait as well as state variables. Researchers often emphasize on the state variables as these can actually be changed by students to improve their success. To show that state as well as trait variables have been found to predict student success, we provide a brief overview of some of the studied variables.

Personality is a trait variable known to be useful in predicting student performance. Personality is frequently tested with the Big Five traits of personality: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism. Especially conscientiousness is found to be a stable predictor. A meta-analysis of papers using the Big Five traits as predictors showed that the mean correlation between performance and conscientiousness and performance was $r = .24$ (O'Connor & Paunonen, 2007). The mean correlations of the other factors were considerably lower: openness to experience $r = .06$, extraversion $r = -.05$, agreeableness $r = .06$, and neuroticism $r = -.03$.

The state variables motivation and time management have shown to be positively correlated with student success multiple times (Bipp et al., 2013; Britton & Tesser, 1991; Kaufman, Agars, & Lopez-Wagner, 2008). Kaufman et al. (2008) found that intrinsic and extrinsic motivation could explain an additional 6% of variance in student success next to GPA and parental educational level. Britton and Tesser (1991) found that time attitudes, i.e. the feeling that you are in charge of how your time is spent, could account for 15% of the variance in GPA, and short-range planning for an additional 6%. Long-range planning was not found to have a significant influence. Superby et al. (2006) found that perceptions of the environment and the academic context did not have a significant influence on academic success. However, students who felt they had made a thorough decision for what university they wanted to go to, did receive a higher average grade ($r = 0.18$).

Overall, research showed state and trait variables combined could account for 16% (Kaufman et al., 2008), 20-30% (Bipp et al., 2013), 36% (Britton & Tesser, 1991), and 43% (Dollinger et al., 2008) of the variance in student success. Thus, learner data are useful in predicting student performance. However, since the introduction of LMSs, researchers started using LMS data, while omitting learner data. Unfortunately, the two data sources are rarely combined, even though the combination might result in more accurate predictions, especially for early intervention, as learner data can give more detailed and timely information (Shum & Crick, 2012). Theory on self-regulated learning also suggests that we should not omit student characteristics (Winne & Hadwin, 1998). By including cognitive conditions including beliefs, motivation, and prior knowledge, we can control for the effects of these conditions on the learning behaviour in the LMS. Additionally, comparing LMS data with learner data might give insight in the usefulness of both predictors. Lastly, combining LMS data with learner data might result in a higher portability, as the models also account for differences in the student.

2.2.4 Predicting student performance using learner data and LMS data

Some studies supplemented LMS data with basic background information, such as age, gender, and prior education or prior GPA (Tempelaar, Heck, Cuypers, van der Kooij, & van de Vrie, 2013), but most do not state any statistics about the influence of these variables (Arnold & Pistilli, 2012; Beer et al., 2010). Hence, it cannot be determined whether background variables are of any added value next to the LMS data. Although it is useful to control for these background variables, it could be reasoned that these variables cannot be controlled by the student and are therefore not useful as indicators for students and teachers on how to improve student success (Yu & Jo, 2014). Other learner data, such as motivation and time management might be of more value, as these could be influenced. However, the combination of LMS data with learner data is rare.

Tempelaar et al. (2015) did combine and compare Blackboard LMS data, with entry test data, data from two test-directed environments, quiz performance data, demographics, and self-reports on culture, learning styles, motivation, and emotions. Linear hierarchic regression on 873 Mathematics and Statistics students showed that behaviour in the two test-directed environments could best predict performance (*R* = .51 - .66). Furthermore, entry test could predict performance (*R* = .41 - .45), followed by motivation and engagement (*R* = .27 - .34), and learning styles (*R* = .21 - .25). Interestingly, LMS data played just a minor role: only the number of clicks was a significant predictor, and all LMS data combined could explain a marginal 4% of the variance in performance. This low percentage of variance explained could be due the fact that most behaviour occurred in the e-tutorials, while the LMS was marginally used. Thus, with such rich data available, LMS data might not be of an added value.

However, when less rich data is available, for example when no e-tutorials are provided, it is not known whether learner data or LMS data can be best used to predict student performance. Therefore, in our second study we compare LMS data with learner data for five courses, where no e-tutorials or other online learning systems are provided for the course. This could give more insight in whether LMS data or learner data is more useful for predicting success, and whether the added value of combining the data is significant enough to warrant the extra time needed to collect and combine the data. Moreover, as learner data can be available before the course starts, it might be useful for early prediction as well. Therefore, we also analyse and compare the prediction of LMS data and learner data over time.

2.2.5 Predicting student performance over time

Most studies that have tried to predict student performance analysed the behaviour of students in the LMS during the whole course. This indicates whether it is possible to infer study success from LMS data, but at a point in time where interventions are no longer meaningful (Campbell & Oblinger, 2007). Several but not many researchers have acknowledged this issue and decided to analyse potentially predictive data from early stages in a course.

For instance, Milne et al. (2012) analysed LMS data of the first week of a course for 658 students in 9 blended courses. They found that usage of the LMS in the first week of the course was significantly higher for successful students than for students who failed the course. Hu, Lo, and Shih (2014) analysed not only first week of the course, but predicted student performance of 300 students at three points in time during the course. In total fourteen different LMS variables were extracted, which were grouped for the first four, eight, and thirteen weeks of the course. Using three different classification techniques, it was found that prediction accuracy increased as the course progressed. The most significant predictors were the total time online, the number of course materials viewed, the average time per session, and the total time used for viewing materials.

Schell, Lukoff, and Alvarado (2014) also found that prediction accuracy increases over time, while analysing performance data (entry test, midterms, and quizzes) and self-efficacy. Multiple linear regression on 89 students in a blended course showed that 29% of the variance in final grade could be explained by the entry test, and the R^2 increased to 34% 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. (2015) also found that the prediction increases over time and that performance data are especially important. The number of clicks in the week before the course started (week 0) was found to have the highest predictive power. As the course progressed, the prediction of student performance gradually improved. Assessment data from the quizzes were shown to be the best predictor, but these data are typically only available after a couple of weeks. 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.

These studies show that the prediction improves over time, and that measures of performance are especially good predictors. In the second study we compare the predictive power of learner data, in-between assessment data, and LMS data for predicting student performance, at the end of the course, but also during the course. This gives us insight in the usefulness and the accurateness of both sources in predicting student performance early in the course, which in turn can be used for timely intervention.

3 Method study 1: LMS data

3.1 Participants and study context

The aim of the first study is to determine the portability of the prediction models using LMS data across courses. To achieve this, data were used from seventeen blended courses using Moodle LMS taught at Eindhoven University of Technology, The Netherlands (TU/e). All courses were taught in the first two quarters (fall and winter) of cohort 2014-2015. In total, data from 4,989 students in 17 courses were used, ranging from 62 to 1,121 students per course (*M* = 293, *SD* = 324). As students could enrol for multiple courses, not all cases resemble unique students. An overview of the courses and the number of students per course can be found in Table 2. The courses varied in level, type, and the way LMS was implemented in the course, i.e. which Moodle modules were used. Next to the content also the discussion forum, poll, wiki, quiz, scorm, and assignment modules were used. The scorm module is a special module which can integrate quizzes from an external source into Moodle.

The courses consisted of eight weeks of lectures and two weeks of final exams. LMS log data were collected over these ten weeks, as well as the week before the start of the lectures (week 0), and was grouped per week. Data from the two weeks long Christmas break were also included, as these fell into the lecture weeks of the winter quarter, making for a total of thirteen weeks of LMS data (week 0, eight lecture weeks, two break weeks, and two exam weeks). Next to LMS data, assessment data were collected. A full description of the dataset can be found in Conijn and Snijders (submitted).

3.1 Data pre-processing

To align the raw Moodle log data, data were pre-processed using R. Four basic aggregated predictors per course were used: the total number of clicks, the number of online sessions, the total time online, and the total number of views. A session was defined similarly as in Zacharis (2015), as the sequence of behaviour from the first click after the login to the LMS until the last click before logging out, or the last click before staying inactive for at least 40 minutes. Additionally, each session had to consist 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. Next to the basic predictors, more complex predictors based on study patterns were included: the irregularity of study time (*SD* of the time per session), the irregularity of study interval (*SD* of the time between sessions), the largest period of inactivity (time between two sessions), the time until first activity, and the average time per session.

Table 2: Course characteristics study 1 & study 2

	Course name	Quarter	Level (year)	Type	Most used module	N study 1	N study 2
1	Calculus A	1	1	Basic	Scorm	438	122
2	Calculus B	1	1	Basic	Scorm	1121	297
3	Calculus C	1	1	Basic	Scorm	227	
4	Calculus pre M Architecture	1	Pre M	Basic	Scorm	135	
5	Set theory and Algebra	1	1	Mathematics	Scorm	73	
6	Linear Algebra and Vector Calculus	2	2	Mathematics	Scorm	120	
7	Linear Algebra	1	Pre M	Mathematics	Scorm	76	
8	Experimental Physics 1	1	1	Physics	Quiz	168	
9	Experimental Physics 2	2	1	Physics	Quiz	155	
10	Behavioural Research Methods	2	2	Psychology	Quiz	136	
11	Applied Physical Sciences formal	2	1	Basic	Quiz	836	45
12	Applied Physical Sciences conceptual	2	1	Basic	Quiz	822	350
13	Condensed Matter	2	3	Physics	Quiz	74	
14	Introduction to Psychology & Technology	1	1	Psychology	Quiz	154	74
15	Linear Algebra 1	1	1	Mathematics	Multiple	66	
16	Statistics	2	2	Mathematics	Quiz	326	
17	The Effectiveness of Mathematics	2	1	Mathematics	Multiple	62	

The assessment data consisted of in-between assessment grades, the final exam grade, and the final course grade. All grades are on a scale from 0 to 10, where grades < 5.5 imply a student does not pass a course and all grades ≥ 5.5 represent a pass. 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. In-between assessment data were used as a predictor variable, which was available for 16 courses. The amount, weight, and type of in-between assessments differed among the 16 courses, 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 week 5, we have analysed the data assuming that the grades would be available at the end of week 5.

3.2 Data analyses

After data pre-processing in R, all analyses were run with Stata 14. Correlational analyses, multi-level analyses, ordinary least squares regressions, and multiple linear regressions were conducted. As the assumption of homoscedasticity was often not met, robust regressions were used. As there was overlap between the students, regressions were run on all courses separately. Thereafter, some exploratory analyses were conducted to determine the effects of course characteristics on the prediction models.

4 Results study 1

To analyse the portability of the prediction models across courses, we first analysed whether there is a difference in the predictor variables between the courses. Correlational analysis, multi-variate regression, and ordinary least squares regressions were run to test this. These analyses are described in more detail in our previous work (Conijn & Snijders, submitted). Here we only report a short summary and expand on the findings in Section 4.3.

4.1 Portability of LMS data

First of all, correlational analysis showed that almost all predictor variables correlated significantly in the whole sample of all courses combined. For all courses separately, the correlational analyses showed mixed results. Only in-between assessment grade correlated significantly in all courses, while most other predictors correlated significantly in 30% to 60% of the courses. Some of the variables even showed significant and substantial differences in the direction and the effect size of the correlation across courses. This indicates that the effects of the variables as predictors might differ across courses. A multi-variate analysis on final exam grade with crossed-random effects for course and student was run to check whether there indeed is an amount of variance residing at course level. The analysis showed that 8% of the variance could be explained at course level and 48% at student level. This means that we cannot simply ignore the clustering at course and student level, and that the highest gain in explaining the variance can be found on the student level. Ordinary least squares regression were run to investigate to what extent the effects of the predictors on final exam grade differ per course. All nine predictors varied significantly and substantially with the course (all p's $< .001$). These results show that we cannot simply combine the LMS data of all courses into one analysis. Hence, the portability of the prediction models for final exam grade using LMS variables across these courses is low.

4.2 Predictability of student performance using LMS data

First of all, we investigated whether LMS data could 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 run 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 student level dropped from 48% to 38%, and at course level raised from 8% to 18%. Thus, LMS data can indeed be used to explain part of the variance in final exam grade. Moreover, when LMS data is included, also a substantial amount of variance could be explained at the course level, indicating that it might be useful to take course characteristics into account as well.

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. All predictors with a significance level below .2 were removed from the models. The results of the final models with standardized coefficients for the predictor variables are shown in Table 3.

Table 3: 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	-0.20**	-0.39***	-0.36***		-0.31	-0.14	-0.30*	1.30**			0.14*	0.21*	1.91**				
Number of online sessions		0.24**	0.12						0.14	0.55***	0.44***	0.24*		0.60***	0.46**		0.27
Total time online		-0.19*		0.17	-0.23	0.25*	0.29			0.34**	-0.34***	-0.18*	0.76***	0.16	-0.23*	0.19	
Total number of views	0.31***	0.33***	0.35**		0.63***			-1.38**	-0.17	-0.42**			-2.28**	-0.52**			
Irregularity of study time		-0.09				0.26*	-0.57**				0.17***			-0.40*		-0.16	
Irregularity of study interval	-0.37***	-0.32***		-0.19**			0.38	-0.52*	0.27	-0.10	-0.22**	-0.12		-0.34		-0.53***	-0.77
Largest period of inactivity	0.16*	0.15**	-0.25**				-0.56*	0.56*	-0.24		0.32***	0.19**		0.33*		0.50**	0.72*
Time until first activity	-0.08	-0.10***		-0.27**		-0.17	-0.58***	-0.27***	-0.21							-0.12*	
Average time per session		0.22**				-0.22	0.35**				0.14	0.13**	-0.15	0.20		-0.18*	-0.20
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

[a)] Standardized betas for all variables
[b)] * $p < .05$, ** $p < .01$, *** $p < .001$
[c)] Constants omitted from the table

The results show that LMS data does explain some of the variance in final exam grade in every course, but the amount of explained variance differs to a great extent: from 8% for course 9 (where none of the predictors were significant), to 37% in course 7. Additionally, the predictor variables included in the final models differ to a great extent as well. None of the predictors is present in all of the models. The total time online and the irregularity of study interval are most often present in the models (12 out of 17), whereas the irregularity of study time per session is the least present (6 out of 17). Most predictors also differ in sign across courses, except for the amount of sessions which always shows a positive coefficient, and the time until the first hit which always show a negative coefficient. This implies that more general conclusions based on these data should be restricted to these two variables; more online sessions and less time until the first session (i.e. starting early) go with a higher grade.

To conclude, the multi-level analysis showed that LMS data can explain a substantial part of the variance in final exam grade on the student level. However, we do not know whether this is an effect within a course, between students, or within a student, across courses. Therefore, post-hoc analyses are run to determine these effects. The separate multiple regressions showed that all variables examined are useful to some degree, as they correlate in at least 5 of the 17 courses and are significant predictors in at least 6 of the 17 models. However, the predictors used and the sign and the size of the coefficients for these predictor vary greatly across the courses. This again indicates that the portability across courses of models using LMS data is low. These differences might be due to instructional conditions or course characteristics. When these course characteristics are studied we might better understand why certain variables are significant in some of the courses and not in others. Therefore, we report on some post-hoc analyses which explore the effects of course characteristics on the prediction models.

4.3 Post-hoc analyses

4.3.1 Effects between and within students

As data is available over multiple courses and some students followed multiple courses, we could further analyse the effects of LMS data on final exam grade between and within students. For example, students might get a higher grade in courses where they show more online activity (within student), or students might get a higher grade when they (on average in all courses) show more online activity than other students (between students). Therefore, multi-level analyses for each predictor were run with the mean and deviance from the mean for that predictor, dummies for all courses, and random intercepts for students. It was 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 within and between students (all *p*'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 ($p < .001$). The largest period of inactivity and the average time per session did not have significant effects between and within students. This shows that it might be useful to not only compare students with their peers, but also with their behaviour in other courses. Thus, although we cannot simply combine all courses into one regression, it is still useful to look at behaviour of the same student in other courses.

4.3.2 Effects of course characteristics

After including LMS data into the multi-level analysis, it was shown that a lot of variance could be explained at the course level. The differences between the prediction models across courses might be due to these course characteristics. To test the effect of the course characteristics, post-hoc analyses were run using ordinary least squares regressions on subsamples of the courses. The subsamples were based on the course characteristics. Courses were clustered based on type of course, level of course, and most used module in Moodle (see Table 2). The results showed that type of course might have some effect on the predictors, as variability in predictors between courses was lower when only courses of the same type were compared. For Mathematics courses, the effect of only 2 out of the 9 predictors differed between the courses; for Physics courses 6 out of the 9 predictors differed significantly; and for Psychology courses 4 out of the 9 predictors differed significantly. The variability in the predictors between the basic courses, offered for multiple disciplines, remained high. This indicates that we might aggregate LMS data when the courses are in the same discipline, as the predictors show less differences between courses. However, there are still some differences between courses within the same discipline, indicating that more assumptions need to be met to aggregate LMS data from multiple courses.

The level of course does not explain the differences between the effects of the predictors across courses. The effect of all predictors still differed per course when only courses for first-year students were considered. For second-year and third-year courses, 4 out of the 9 predictors differed per course. The modules used in Moodle did have an influence on the effect of the predictors per course. Seven courses used the scorm module most often, with at least 70% of the total amount clicks in the scorm module, eight courses the quiz module (at least 45% of the total clicks), and in two courses no clear module was used most often. Courses which used the quiz module most often, showed different effects for 8 out of the 9 predictors, whereas courses using the scorm module, showed different effects for only 3 of the 9 predictors. This is interesting, as both modules provide quizzes either in Moodle itself, or in an external source integrated in Moodle using scorm. The differences might be caused by the fact that the quiz modules are implemented in different ways across courses. Scorm modules, on the other hand, are mostly used for Mathematics courses and may therefore be implemented similarly across courses. This suggests that the way an LMS is implemented in a course, i.e. which modules are used and how these modules are implemented, might have some effect on whether the LMS data could be aggregated for further analyses.

5 Conclusion study 1

The first study aimed to determine whether prediction models using LMS data are portable across courses. First of all, we found differences in the correlational analysis of final exam grade with the predictors over the different courses. Moreover, substantial differences were found between the sign and the size of the predictors across courses. This shows that we cannot simply run analyses on the data of multiple courses combined and the portability of the models for predicting student performance thus appears to be low. For individual courses the prediction models still provide useful information for the instructor to improve learning and teaching, but it cannot simply be assumed that the models can be used for other courses as well.

The differences in the prediction models might be explained by differences in course characteristics and student characteristics across courses. Exploratory analyses on the course characteristics showed that there indeed is some effect of discipline and how the LMS is used in the course, but these course characteristics cannot completely explain differences between the prediction models. Winne and Hadwin (1998) stated that learning is not only affected by task conditions (such as course characteristics), but also by internal factors, such as student dispositions and motivational factors. Hence, student characteristics could also influence the behaviour in the LMS and explain the differences in the prediction models. Therefore, in our second study we include student characteristics, also known as learner data, to find out whether these can explain the differences between the models. Moreover, we test if learner data can improve the prediction of student performance and which source, LMS data or learner data, has the highest power in predicting final grade.

6 Method study 2: LMS data and learner data

6.1 Participants and study context

For the second study, LMS data and performance data from study 1 were combined with learner data. Learner data came from a test for prospective students, which was only available for students of the departments of Industrial Engineering & Innovation Sciences and Built Environment. In total there were 426 students who conducted the test and who completed at least one course that made use of the LMS at the TU/e. Data from these students were combined with LMS data and performance data available per course. Only courses where at least 45 students had taken the test were included. As some students followed multiple courses (32 students followed 1 course, 326 followed 2, and 68 followed 3), this resulted in a total of 888 cases in 5 courses: Calculus A, Calculus B, Applied Physical Sciences formal, Applied Physical Sciences conceptual, and Introduction to Psychology & Technology (see Table 2).

6.2 Learner data

The learner data were extracted from an online questionnaire, which was part of the a study choice test for prospective students of Eindhoven University of Technology. Such a study choice test is available in all Dutch higher education institutions, to give students the opportunity to make a deliberate decision with respect to their further education. Data used in the current study came from a pilot of the online questionnaire, where only prospective students of the departments of Industrial Engineering & Innovation Sciences and Built Environment filled-in the questionnaire. The questionnaires were send to prospective students, a week before they followed an orientation activity at the TU/e. These activities took place between two to seven months before the start of the study program. When students did not complete the questionnaire before the orientation activity, they got the time to complete the questionnaire during the activity. This resulted in a really high-response rate. The questionnaire measured demographics and a total of nine factors related to capacities and motivation for the study program. Based on the online questionnaire, an advice concerning the study choice was given to the prospective students, categorized in capacities and motivation for their choice.

Most of these factors were adapted from validated questionnaires and were shown to be significant predictors in a previous longitudinal study on student performance and study continuation at the TU/e (Bipp et al., 2013). An overview of the capacity and motivation factors and their descriptive values can be found in Table 4, the complete questionnaire can be found in Appendix A: Questionnaire learner data (Dutch). The demographical measures consisted of gender, chosen Bachelor program (Industrial Engineering (IE), Psychology & Technology (P&T), Sustainable Innovation (SI), or Built Environment (BE)), and profile in prior education (science-oriented or society-oriented). Skills and capacities consisted of: GPA prior education, conscientiousness, time management, lack of study strategy, and self-efficacy. Motivation for study choice consisted of: bond with study program, confidence study choice, amotivation study choice, and external regulation.

Table 4: Predictor variables used in predicting student performance using learner data

Predictor	Scale	N	*M*	*SD*	Source
GPA prior education [a)]	[0-10]	819	6.87	0.52	
Conscientiousness	[1-5]	888	3.77	0.50	(Denissen, Geenen, van Aken, Gosling, & Potter, 2008)
Time management	[1-5]	888	3.76	0.64	(Kleijn, Topman, & Ploeg, 1994)
Study strategy (lack of)	[1-5]	888	2.14	0.92	(Harackiewicz, Barron, Tauer, Carter, & Elliot, 2000)
Self-efficacy	[1-5]	888	4.94	0.66	(Pintrich & De Groot, 1990)
Bond with study program	[1-7]	888	5.53	0.64	
Confidence study choice	[1-7]	888	5.57	0.89	
Amotivation study choice	[1-7]	888	1.49	0.64	(adapted from Guay, Vallerand, & Blanchard, 2000)
External regulation	[1-7]	888	2.03	0.94	(adapted from Guay et al., 2000)

[a)] For calculating prior GPA, a higher weight was given for courses which are required for entering the study program.

6.3 Data analyses

As in study 1, all analyses were conducted with Stata 14. As only students who filled-in the questionnaire were used as sample for this study, this study uses a subsample of the sample of the first study. Therefore, several regression analyses were run to compare the subsample used in this study to the whole sample used in study 1. Thereafter, the differences of the predictors per course were analysed using correlational analyses, multi-level analysis, and ordinary least squares regressions. Finally, multiple linear regressions were run to compare the prediction models using learner data, LMS data, and learner data combined with LMS data. All predictors with a significance level below .2 were removed from the models. As the assumption of homoscedasticity was often not met, robust regressions were used. Robustness of all models was checked with 10-fold cross-validation, using the function 'crossfold', which runs ten regressions on subsamples and takes the average of these regressions. Although most previous studies only report how well the regression or classification model performed in terms of (pseudo) R-squared values, this is not always a very useful metric. In most cases, it is more insightful to know how far away the predictions are from the true value, on average. This could for example give more insight in whether the model could be used for automated assessment. For this reason, we calculated such fit statistics as well.

7 Results study 2

7.1 Preliminary analysis

7.1.1 Differences between the subsample and the whole sample

As learner data were not available for all students, analyses in the second study are conducted on a subsample of students within five courses. The subsample was not randomly chosen, hence it is important to check whether the subsample significantly differs from the complete sample in these five courses. Paired-samples *t*-tests showed that there is a significant difference between the subsample and the whole sample, for almost all predictor variables (all *p*'s $< .05$). Only the total amount of views did not differ between the two samples ($t(888) = -1.56$, $p = .12$). The outcome variable final exam grade also differed significantly between the two groups ($t(888) = 2.32$, $p = .02$). Students in the subsample received a significant lower grade ($M = 5.32$, $SD = 2.10$), compared to students in the whole sample ($M = 5.44$, $SD = 2.34$). To investigate whether these differences affect the prediction of student performance, we ran four regressions on final exam grade, comparing students within the subsample with students in the whole sample. The four multiple linear regressions shown in Table 5 indicate that being in the subsample has an effect on the prediction models of final exam grade.

The first model, with the dummy in_subsample as only predictor, shows that being in the subsample does not have a significant effect on final exam grade. However, when we look at the separate courses (model 2) we do see a significant effect of being in the sample for one of the courses. For course 11, being in the subsample leads to a 1.3 lower grade, compared to the other students of course 11. Moreover, when the basic and pattern predictors are added to the model, we see that being in the subsample does have a significant (negative) effect on final grade (model 3). Thus, students in the subsample who show the same online behaviour as students in the whole sample, have a significant lower grade than students who are not in the subsample. Lastly, the interaction effects of being in the subsample with the predictors were included (model 4). This model shows that the predictors have a different effect inside and outside the subsample. The total number of clicks and the total time online have a significantly less negative effect on final exam grade in the subsample, compared to the whole sample. Contrary, the total amount of views and the average time online have a significantly less positive effect on final exam grade in the subsample.

Thus, the models show that there indeed is a difference between the effects of the predictors on final exam grade between students within the subsample and students outside the subsample. This difference might be explained by the study program, as only students from the departments of Industrial Engineering & Innovation Sciences and Built Environment took the test. The whole sample also consisted of students from more traditional engineering programs, as Physics and Mathematics. These students might perform better on the basic Calculus and Applied Physics courses, with similar amounts of learning in the LMS. The difference between the whole sample and the subsample points out that we cannot use the results from the subsample to draw conclusions about the whole sample, especially not about the predictors which show different effects. Moreover, the findings corroborate the results of study 1, showing that the effects of predictors are different per sample and that we cannot generalize the effects of a single predictor. Nevertheless, the comparisons of the models within the subsample remain valid. Thus, we can still compare the effects of using learner data and LMS data for predicting student performance.

Table 5: Effects of being in the subsample on final grade, compared to the whole sample

	Model 1	Model 2	Model 3	Model 4
in_subsample	-0.05	0.18	- 0.19*	- 0.17
Course 1		0.00		
Course 2		0.01		
Course 11		0.40*		
Course 12		- 0.71***		
Course 14		0.60**		
Course 1 * in_subsample		0.00		
Course 2 * in_subsample		- 0.27		
Course 11 * in_subsample		- 1.28**		
Course 12 * in_subsample		0.27		
Course 14 * in_subsample		0.40		
Total number of clicks			- 0.81***	- 0.98***
Number of online sessions			0.84***	0.80***
Total time online			- 0.40***	- 0.43***
Total number of views			0.75***	0.98***
Irregularity of study time			0.01	- 0.01
Irregularity of study interval			- 0.35***	- 0.35***
Largest period of inactivity			0.44***	0.45***
Time until first activity			- 0.11*	- 0.07
Average time per session			0.22*	0.28*
in_subsample * Total number of clicks				0.39**
in_subsample * Number of online sessions				- 0.17
in_subsample * Total time online				0.58*
in_subsample * Total number of views				- 0.81***
in_subsample * Irregularity of study time				0.07
in_subsample * Irregularity of study interval				- 0.40
in_subsample * Largest period of inactivity				0.15
in_subsample * Time until first activity				- 0.19
in_subsample * Average time per session				- 0.43*
R^2	.00	.03	.14	.15
N	3371	3371	3371	3371

a) Standardized values for all predictors
b) * p < .05, ** p < .01, *** p < .001
c) Constants omitted from table

7.1.2 Detecting outliers

Some of the students spend a lot of time online, while others rarely login to an LMS. These students who show extreme values on one or more of the predictors might influence the prediction models resulting in less generalizable models. These outliers cannot simply be removed from the analyses, as they still correspond to valid behaviours of students. To determine if there were influential cases, Cook’s distance was calculated for the models on the five courses separately using all data sources. For all 5 models Cook’s distance was below 1, indicating that there were no highly influential cases. Thereafter, we analysed students who had an extreme value on one or more of the predictors. Here, we defined an extreme value as a value which is at least three standard deviations lower or higher than the mean. In course 1, 19 students were found as an outlier, 35 in course 2, 12 in course 11, 42 in course 12, and 8 in course 14. To analyse whether these outliers also influenced the prediction models, we reran the multiple linear regressions with these outliers excluded. These prediction models did not differ much from the prediction models in which all cases were included. The models without the outliers had a .04 lower R^2 in course 1, an equal R^2 in course 2, and a .03 to .04 higher R^2 in courses 11, 12, and 14. Also, the effect of the predictors only differed marginally. Thus, the outliers did not influence the prediction models too much, indicating that we can leave the outliers in the models.

7.2 Portability of learner data

Before we analysed whether learner data can be used to predict student performance, we first examined whether the effect of the learner data predictors differed per course (as we did in study 1 for the effects of LMS data). If the predictors do not differ per course, one model could be used for all courses, with all data aggregated. To analyse whether the learner data are portable across courses, we ran correlational analysis, a multi-variate regression, and ordinary least squares regressions.

First of all, Pearson correlational analyses were conducted for all five courses, with the learner data variables and final exam grade. The results showed that most predictors correlate significantly in none, only one, or two of the courses. The results further showed that the predictors which correlate significantly with final grade differ per course (Table 6). Only past GPA correlated significantly in every course, with a moderate to large effect size (r = .38 - .54). The other capacity factors - conscientiousness, time management, and lack of study strategy - showed a significant correlation with low effect size (r < .3) in two courses, while self-efficacy did not correlate significantly in any of the courses. The motivational factors - bond with study program, certainty about study choice, and amotivation - only correlated in one of the courses with a low effect size (r < .3). External regulation of study choice did not correlate significantly in any of the courses. Interestingly, the significant correlations of the motivational variables were in the opposite direction of what was expected; a higher bond and certainty about the study program, and lower amotivation were correlated with a lower grade. This might be due to the fact that the analysed courses are basic courses, which every student has to take and which are often not directly related to the students' major.

The type of major correlated significantly only in the two Calculus courses, where students from Sustainable Innovation and Psychology & Technology had a negative correlation with grade in Calculus A, and students from Built Environment had a positive correlation. For Calculus B, students from Industrial Engineering had a negative correlation, while Psychology & Technology students had a positive correlation. These effects could be due to the fact that all Industrial Engineering students had to take variant B, and all Built Environment students had to take variant A, whereas students from Psychology & Technology and Sustainable Innovation were allowed to choose between the different Calculus courses. Students who had difficulties with mathematics in prior education were advised to choose the more conceptual variant (A), while the others were advised to choose the more formal variant (B). Thus, students from Psychology & Technology and Sustainable Innovation who choose variant A might have had less mathematical skills than average, and students who choose variant B might have had better mathematical skills on average compared to other students who were not allowed to choose between the two variants.

Table 6: Bi-variate correlations of the predictors with final grade per course (Pearson's *r*)

	Course 1	Course2	Course11	Course12	Course14
Male	.084	.220***	.022	- .044	.274*
Major IE		- .139*	- .001	- .009	
Major P&T	- .260**	.146*	- .005	.014	
Major SI	- .192*	.018	.010	- .065	
Major BE	.324***			.025	
Science-oriented profile	- .021	- .023	- .068	.093	.183
Prior GPA	.406***	.427***	.535***	.377***	.394**
Conscientiousness	.174	.166**	.173	.070	.294*
Time management	.232*	.217***	.180	.047	.226
Lack of study strategy	- .225*	- .126*	- .015	- .017	- .161
Self-efficacy	.089	- .040	.040	.059	- .172
Bond with study program	.049	- .025	.114	- .112*	- .005
Confidence study choice	- .007	- .026	- .296*	- .001	.079
Amotivation for study choice	- .006	.073	- .044	.159**	- .196
External regulation study choice	.015	.004	.237	.079	- .063
N	122	297	45	350	74

[a] * $p < .05$, ** $p < .01$, *** $p < .001$

The differences in correlations between final exam grade and the predictor variables for each course give an indication that different predictor variables might be important in different courses. A multi-variate analysis on final exam grade with crossed-random effects for course and student was run to check whether there indeed is an amount of variance residing at course level. The analysis showed that 9% of the variance could be explained at course level and 37% at student level. This means that we cannot simply ignore the clustering at course and student level. Moreover, it shows that a lot of variance can be explained using student variables. Later on, we show whether learner data and LMS data can explain this variance. Compared to study 1, significantly less variance resides at the student level, which is probably due to the significantly lower sample size, where a greater amount of variance is attributed to error.

As the correlational and multi-variate analyses show that there is an effect of course, and correlational analysis showed that predictors differ per course, we ran ordinary least squares (OLS) regressions to investigate to what extent the effects of the predictors on final exam grade differ per course. OLS regressions were run with all learner variables on final grade, including the courses (coded as dummies), and the interaction effects of the predictor with the courses. The major was omitted, as some of the courses were only required for one major. All of the other eleven learner variables varied significantly and substantially with the course (all *p*'s < .001). These results show that we (again) cannot simply combine learner data of all courses into one regression. Moreover, the portability of the prediction models using learner data might be low. However, learner data might still explain the different effects of LMS data across courses.

7.3 Predictability of student performance

The multi-level analysis showed that 37% of variance in final exam grade can be explained on the student level, and 9% at course level. LMS data as well as learner data might be used to explain part of this variance. Study 1 indeed showed that LMS data can be used to predict student performance. The correlational analysis on learner data show that learner data might be useful as well. In the following we first examine whether learner data can indeed explain (a part of) this 37%. Thereafter, we compare the predictive power of models using LMS data with models using learner data. Finally, we examine whether the combination of LMS data with LMS data results in a better prediction, and how this prediction evolves over time.

7.3.1 Predicting student performance with learner data

First, we checked whether learner data can be used to predict student performance and thus can explain (a part of) this 37%. To determine this, a multi-level analysis on final exam grade was run with all learner data and crossed-random effects for course and student on the five courses. It was found that after adding these learner data to the analysis, the amount of variance that could be explained at student level dropped to 21%, and at course level raised to 15%. This indicates that learner data indeed can explain some of the variance and hence is a useful predictor, but does not account for all variance.

Table 7: Multiple linear regressions on final exam grade using learner data, separated per course

	Course 1		Course 2		Course 11		Course 12		Course 14	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Male	0.39		0.53*	0.38	0.22		-0.45	-0.37	0.04	
Major IE			0.00	0.00	0.00		0.00			
Major P&T	0.00	0.00	0.54	0.48	1.59		0.20		0.00	
Major SI	- 1.08	- 1.21	0.23	0.23	-0.41		-0.87			
Major BE	1.58**	1.30**					0.21			
Science-oriented profile	0.18		- 0.12		0.75		0.53		1.30	1.14
Prior GPA	1.17***	1.30***	1.59***	1.65***	2.48**	2.24***	1.53***	1.66***	0.90*	0.84**
Conscientiousness [a)]	- 0.08		- 0.00		-0.01		0.07		0.29	0.20
Time management [a)]	0.31**	0.26**	0.10	0.11	0.06		0.01		-0.06	
Lack of study strategy [a)]	- 0.11		- 0.07		0.06		0.02		-0.06	
Self-efficacy [a)]	0.05		- 0.18*	- 0.20***	-0.20		0.06		-0.28*	-0.31**
Bond study program [a)]	- 0.14		0.05		0.37	0.18	-0.14	-0.07	-0.12	
Confidence choice [a)]	- 0.07		- 0.08		-0.40*	-0.30*	0.03		0.02	
Amotivation choice [a)]	- 0.01		0.06		0.08		0.10	0.09	-0.21	-0.16
External regulation choice [a)]	0.14	0.11	0.06		0.39	0.30*	0.06		0.14	
R^2	.34	.30	.26	.24	.57	.47	.20	.17	.32	.28
N	116	116	273	273	38	38	328	328	64	64

[a)] Standardized betas reported
[b)] (1) Full model, (2) Predictors with *p*-value < .2 in full model
[c)] * p < .05, ** p < .01, *** p < .001
[d)] Constants omitted from table

Secondly, five multiple linear regressions were run on the five separate courses. The final models including all predictors, and the models including only the predictors with a *p*-value below 0.2 in the full model can be found in Table 7. The models show that learner data can on average explain 29% of the variance in final exam grade. Thus, learner data are a useful predictor for predicting student performance. As expected, past GPA was found to be a significant and substantial predictor in all courses. The effects of the other predictors differed highly across courses. Interestingly, self-efficacy was found a significant negative predictor in two of the five courses. This indicates that one standard deviation increase on the self-efficacy scale, with all other variables constant, results in a 0.2 to 0.3 standard deviation decrease in final exam grade in these two courses. Some other predictors were found significant in only one of the five courses: the major Built Environment, time management, and external regulation were positively related with final exam grade, whereas certainty about the study choice was negatively related. Gender, all other major programs, conscientiousness, the lack of learning strategy, bond with the study program, and amotivation were not found significant predictors in any of the courses.

To conclude, learner data can be used to predict final exam grade. The models give insight in which predictors influence final exam grade in these five courses. Especially past GPA is shown to be a robust predictor. The predictive value of the other predictors differs per course, and is substantially lower. Half of the predictors were not significant in any of the models and three of the twelve predictors were significant in only one of the five courses. For each separate course, the amount of variance explained is quite high, indicating that the models might be used for these specific courses to improve learning and teaching. However, with a mean residual of 1.78 for the prediction, the models are distant from accurate prediction. Moreover, the multi-level analysis showed that not all variance on the student level could be explained by learner data. Therefore, we analysed whether LMS data might be more useful for predicting student performance, and whether these two sources combined could significantly increase the prediction.

7.3.2 Predicting student performance with learner data and LMS data: compared

First of all, we compared the prediction models using learner data with the prediction models using LMS data, to investigate which data source works best for predicting student performance. As we showed in the preliminary analysis that the whole sample differs significantly from the sample using learner data, we cannot compare the models from study 1 using LMS data on the whole sample with the subsample using learner data. Therefore, we ran a multi-level analysis on final exam grade with all LMS data, in-between assessment grade, and crossed-random effects for course and student on the subsample. The results showed that when LMS data is added, only 19% of the variance can be explained at student level, and 12% at course level. Thus, LMS data is still a useful predictor on the subsample, but it cannot account for all variance on the student level. Moreover, the variance in final exam grade remaining at the student level is lower for LMS data with in-between assessment data than for learner data. This makes LMS data with in-between assessment data a potentially better predictor for student performance than learner data.

To determine this, multiple regressions on final exam grade were run on the subsample, for each course separately, using LMS data (with and without in-between assessment grades), learner data, and LMS combined with learner data. To facilitate the comparison, all models are shown in Table 8.

Table 8: Multiple linear regressions on final exam grade using learner data and LMS data, separated per course

	Course 1				Course 2				Course 11				Course 12				Course 14			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Total number of clicks [a)]	-0.18*	-0.17		-0.32*	-0.34***	-0.38***		-0.06					0.07							
Number of online sessions [a)]				0.00																
Total time online [a)]					0.13*				-0.05	0.61***						0.11*				
Total number of views [a)]	0.32***	0.25*		0.37*	0.29***	0.35***														
Irregularity of study time [a)]								-0.06						0.13*						-0.07
Irregularity of study interval [a)]				-0.21*		-0.26***			-0.61*			-0.03						-0.71***		-0.17
Largest period of inactivity [a)]									0.62*					0.10				0.52**		
Time until first activity [a)]		-0.23**						-0.09												
Average time per session [a)]	-0.15*	-0.09								-0.34*			-0.02			-0.05				
In-between assessment grade	1.13***			0.96***	0.73***			0.63***	0.85***			0.93***	0.62***			0.47***	0.85***			0.64***
Male							0.38	0.36							-0.37	-0.49*				
Major IE							0.00													
Major P&T			0.00				0.48													
Major SI			-1.21				0.23													
Major BE			1.30**																	
Science-oriented profile				0.72															1.14	
GPA prior education			1.30***	0.86**			1.65***	1.15***			2.24***				1.66***	1.27***			0.84**	0.79*
Conscientiousness [a)]																			0.20	
Time management [a)]			0.26**	0.21**			0.11	0.09												-0.19
Study strategy (lack of) [a)]																				
Self-efficacy [a)]							-0.20***	-0.17**								0.09			-0.31**	
Bond with study program [a)]											0.18				-0.07	-0.15**				
Confidence study choice [a)]											-0.30*									
Amotivation study choice [a)]															0.09				-0.16	
External regulation [a)]			0.11								0.30*									0.10
R^2	.40	.11	.30	.50	.34	.19	.24	.37	.55	.29	.47	.49	.19	.03	.17	.29	.32	.14	.28	.38
R^2 cross-validated	.27	.09	.11	.20	.24	.06	.14	.26	.20	.05	.11	.22	.12	.01	.14	.25	.22	.06	.08	.24
M residual	1.49	1.60	1.73	1.96	1.47	1.78	1.56	1.42	1.54	2.65	2.25	1.48	1.65	1.77	1.57	1.45	2.01	1.78	1.81	1.66
N	122	122	116	116	297	297	273	273	45	45	38	45	350	350	328	328	74	74	64	64

[a)] Standardized betas reported

[b)] (1) LMS data with final grade, (2) LMS data without final grade, (3) Learner data, (4) LMS data and learner data

[c)] * $p < .05$, ** $p < .01$, *** $p < .001$

[d)] Constants omitted from table

The models show that LMS data with in-between assessment grade can explain the highest amount of variance, with an average R^2 of .36. However, when in-between assessments are not (yet) available this drops to .15. Learner data can predict on average .29 of the variance in final exam grade. This indicates that learner data might be more useful than LMS data when in-between assessments are not (yet) available. However, the models show that the prediction of final exam grade with learner data is still on average 1.78 point away on the scale from 0 to 10, and LMS data with in-between assessment is on average 1.63 away. Moreover, multi-level analyses showed that still 19 to 21% of the variance could be explained at student level. Hence, to improve the prediction and explain more of the variance in final exam grade, LMS data is combined with learner data.

7.3.3 Predicting student performance with learner data and LMS data: combined

Multi-level analyses were run to determine whether LMS data combined with learner data could explain an additional part of the variance in final grade. Multi-level analyses were run with LMS data, in-between assessment grade, learner data, and random effects for course and student. The results showed that the variance residing at student level dropped to 13%, and the variance residing at course level raised to 15% when these variables were added. Thus, not all variance can be explained when LMS data, assessment data, and learner data are combined. However, the variance residing at student level is significantly less than when only learner data (21%) or LMS data with in-between assessment data (19%) were included. Thus, learner data can explain an additional part of the variance in final exam grade, next to LMS data. Five multiple linear regressions using both learner data and LMS data indeed showed that the explained variance increased when the two sources were combined. On average, 41% of variance in final grade could be explained when learner data as well as LMS data with in-between assessments were taken into account.

The models using both learner data and LMS data again show that especially the measurements of performance, such as past GPA and in-between assessment grade, have a high predictive power. In two of the five models there are even no other significant predictors left, next to the performance measures. The other three courses show some effect of the other predictors: the total amount of clicks, the total amount of views, the total time online, gender, the lack of time management, self-efficacy, and bond with study program are significant in one of these three courses. Thus, a lot of the predictive power comes from performance measures. This indicates that time-consuming questionnaires about capacities and motivation, and analyses of LMS data might not be necessary when some measures of performance are available. When a in-between grade is available, the addition of learner data indeed only slightly improves the prediction.

As the sample sizes per course are quite small, the models might explain too much of the error in the data. Therefore, 10-fold cross-validation was conducted on all models (see Table 8) to determine whether the models overfit the data. The cross-validation indeed results in a significant lower pseudo R^2 on average. For the models with LMS data the R^2 decreased from .15 to .05, for LMS data with in-between assessment grade from .36 to .21, for learner data from .29 to .12, and for all sources combined from .41 to .23. As expected, the difference between the original R^2 and the cross-validated R^2 is highest in the courses with the smallest sample sizes. Thus, the models do not only show variance between the courses, but also overfit the data. Both indicate that the models will perform less on new data. This again shows that the models are useful for a specific course, but the portability of the models across courses or other samples is low.

Although the cross-validated R^2 is significantly lower than the original R^2, the comparisons between the four different models per course stay about the same. Using learner data for predicting final exam grade results in a higher explained variance than using LMS data (average cross-validated R^2 = .12 and .05, respectively), whereas LMS data combined with in-between assessments results in a higher predictive power than learner data, with an average cross-validated R^2 of .21. When learner data is added to LMS data and in-between assessment data, the average cross-validated R^2 increases marginally to .23. Interestingly, for course 1, the model with all data performs even worse than the model with LMS data and in-between assessment data (cross-validated R^2 = .20 and .27, respectively). This indicates that adding learner data to LMS data and in-between assessment data does not have much added value for the prediction of final exam grade. However, when using these prediction models to improve learning and teaching, early intervention is needed. In the first few weeks, before in-between assessment data are available, learner data could still be useful.

7.3.4 Predicting student performance over time

To analyse whether early intervention is possible using LMS data and learner data, and how the prediction evolves over time, predictions were compared over the weeks. Learner data were available before the course started, LMS data were aggregated per week, and in-between assessment grade was available after week 5. For the LMS data, only the basic predictors were used, as usage patterns are often not available or not yet meaningful in the first few weeks. Multiple linear regressions were run on the eleven weeks of the course, with interactions for the courses. Six different combinations of the data sources were used: (1) learner data, LMS data, and in-between assessments; (2) learner data and in-between assessments; (3) LMS data and in-between assessments; (4) learner data and LMS data; (5) LMS data; (6) learner data. The R^2 and the mean residual of these six models over time are shown in Figure 1.

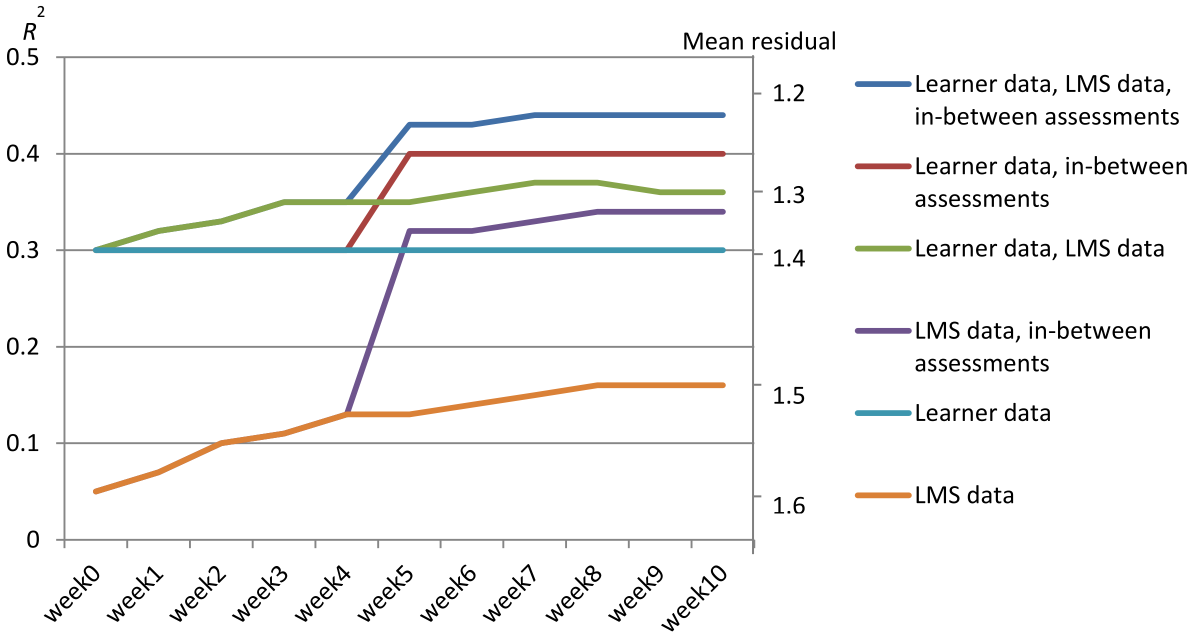

Figure 1: R^2 and mean residual (approximated) for predicting final exam grade over time for six combinations of the different data sources

Expectedly, it was found that the predictions using LMS data improve slightly over time. Also, when in-between assessment data are added at the end of week 5, a high increase in explained variance can be found. The combination of learner data, LMS data, and in-between assessment data results in the highest predictive power during the whole course. When there is no access to the raw LMS log data, using learner data with in-between assessment data is a good second best for predicting final exam grade. For early prediction, before in-between assessment data are available, learner data is the most useful source. Because these data are already available before the course starts, it has a high value for early intervention. The addition of LMS data in the first weeks leads to a slight increase in the prediction. The best compromise between early feedback and accuracy seems to be after week 3, as the prediction does not improve much after that. However, at that point in time, the mean residual is 1.35, hence the prediction is on average 1.35 off away from an accurate prediction of final exam grade (on a scale from 0 to 10). This may however not be a major issue as there is no need to predict the exact final exam grade. It would be enough for intervention to be able to predict whether a student will pass or fail a course.

7.3.5 Predicting pass/fail probabilities

To predict whether a student would pass or fail the course, binary logistic regressions were run on learner data, in-between assessment data, and LMS data grouped per week, with interactions for the courses. As we are particularly interested in whether a student would fail (to provide feedback), students with a final exam grade < 5.5 were coded as at risk (1), while student with a final exam grade ≥ 5.5 were coded not at risk (0). In total 450 of the 888 students were coded as at risk (51%). The same six combinations of the data sources were considered as in the multiple linear regressions. The pseudo R^2 for these six models over time are shown in Figure 2. Similarly as with predicting final exam grade, it was found that the prediction using LMS data improves slightly over time, and a high increase in the prediction can be found after the in-between assessments are added.

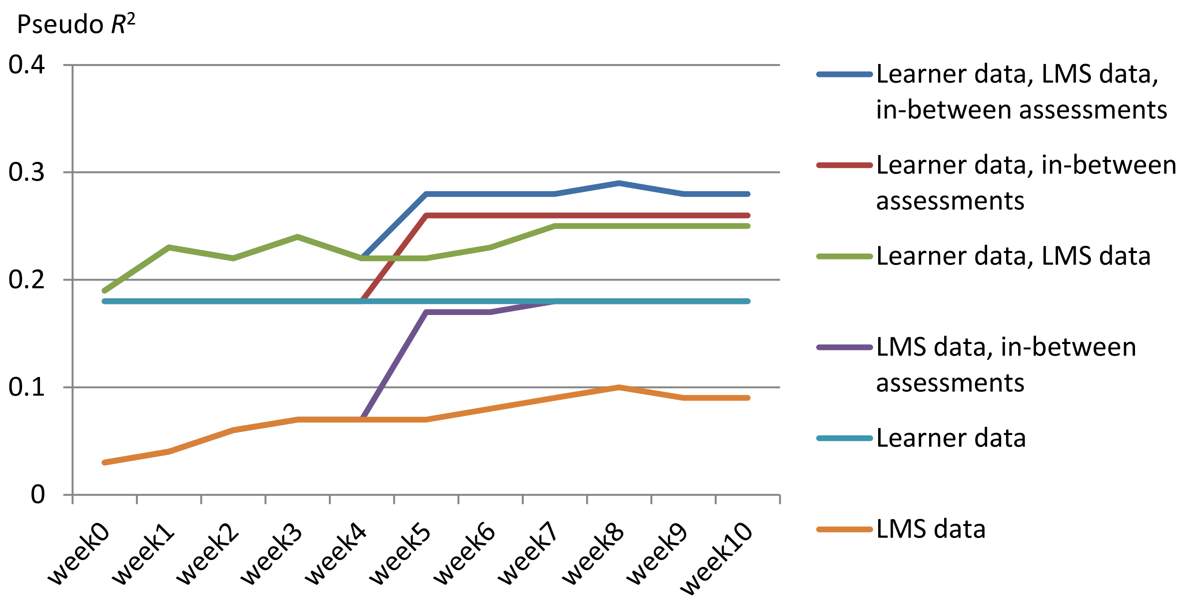

Figure 2: Pseudo R^2 for predicting pass/fail probabilities over time for six combinations of the different data sources

Contrary to predicting final exam grade, learner data is equal or even somewhat better in predicting pass/fail probabilities than LMS data, even after in-between assessment data have become available. Using only LMS data, the total classification accuracy was rather low and ranged from 54% after week 0 to 62% after week 10. Interestingly, when we divide the total prediction accuracy into the accurate predictions of students who passed and failed, we see that LMS data is especially bad in predicting whether a student will pass (specificity). In week 0 LMS data can only accurately predict 24% of the passing students as not at risk, increasing to 57% in week 5, while learner data can predict 69% of the passing students as not at risk. Thus, when the exact grade is not needed, learner data is of more value than LMS data.

Unfortunately, the prediction whether a student would pass or fail is also distant from accurate prediction. The binary logistic regression showed that after week 10, when all data sources are combined, the total classification accuracy equals 74%. Week 1 was the best comprise between early feedback and accuracy, with a total classification accuracy of 72%, a false positive rate of 29% and a false negative rate of 26%. Thus even when all data sources are combined, predicting whether a student would pass or fail is not accurate. Hence, one should proceed with caution when intervening students based on these statistics. With all data included still 26% of the students would not get an intervention, while they actually needed the help. Moreover, 29% of the students would get an intervention while they did not need it, which might influence a students' self-efficacy and motivation.

8 Discussion and Conclusion

8.1 Portability of prediction models

In this study we investigated the value of using LMS and learner data for predicting student performance and the portability of these models across courses. In our first study, we analysed the portability of the prediction models of seventeen blended courses, using LMS data. Similar to previous research (e.g. Tempelaar et al., 2015; Zacharis, 2015), basic predictors were used, including the number of clicks, the number of sessions, the total time online, and the number of views. Additionally, more complex variables based on the study patterns and (ir)regularities were included: the irregularity of study time, the irregularity of study interval, the largest period of inactivity, the time until the first activity, and the average time per session. It was found that the effect of all the predictors differed to a great extent across courses. This corroborates previous findings on predicting student success, which showed different results in correlations and prediction models. We tried to explain these differences between previous studies with the different analytical techniques, different sets of predictor variables, and different LMSs used. However, while keeping the contextual effects more constant, we still found substantial differences in the sign and size of the predictors. This shows that even within one institution, using one LMS, and one set of predictor variables, the portability of the prediction models across courses is low.

The findings are in line with Gašević et al. (2016) who found substantive differences in the prediction models of nine blended courses. Gasevic et al (2016) however used predictors based on the modules used per course. The modules used differed per course, leading to different predictors between courses. In the current study, only predictors were used that were available for all courses. While using a more generic set of variables, still differences were found between the effects of the predictors across courses. The data of several courses can thus neither be simply combined for analysis nor to construct general models. Moreover, this indicates that it might be even harder to compare whole institutions, as institutions probably show more differences than two courses within one institution. When institutions are compared based on prediction models with aggregated LMS data of all courses in the institution, this might overestimate or underestimate the effects per course (cf. Lauría et al., 2012). Especially when only a few institutions are compared, this might give similar results, while the models per course differ extensively.

Thus, LMS data could be used to predict student performance in a specific course, but one should proceed with caution when aggregating the data of multiple courses. Aggregation of data is less problematic when combining data of courses from the same discipline, as is shown with an exploratory analysis. The prediction models then showed more similarities, especially for the Mathematics courses, this might however not be the case for all disciplines (Gašević et al., 2016). Next to combining courses from the same disciplines, courses using the LMS in a similar way might also be combined. For example, it was found that models based on courses using a specific Moodle module (scorm) for their quizzes showed similarities. However, models of courses using the quiz module in Moodle were different. Thus while discipline and most used module in Moodle are factors that lead to less difference in prediction models, one must still be careful in aggregating LMS data from different courses based on these characteristics.

Next to course characteristics, the differences between the prediction models might also be explained by the differences in student characteristics across courses. Theory on self-regulated learning states that learning is not only affected by task conditions (such as course characteristics), but also by internal factors, such as student dispositions and motivational factors (Winne & Hadwin, 1998). In the second study, it was therefore tested if student characteristics can – partly – explain differences in behaviour in the LMS and in student performance.

The learner data used in the second study consisted of the demographical variables gender, science-oriented profile, and current major; the capacities prior GPA, conscientiousness, time management, lack of study strategy, and self-efficacy; and the motivational factors bond with study program, confidence study choice, amotivation study choice, and external regulation. It was shown that the prediction models using these data differed per course. This indicates that learner data of multiple courses can also not be aggregated. The differences could be due to the low amount of significant correlations between the predictors and final exam grade. When learner data were added to the LMS data, the prediction models still differed. Adding student characteristics alone therefore is not sufficient for increasing the portability of the prediction models; course characteristics still need to be considered.

To conclude, the prediction models are useful for specific courses. The portability of the prediction models across courses is however low, even when controlling for student characteristics. Course characteristics did have some influence on the prediction models. Unfortunately, as only data of seventeen courses were available in our case, we could not elaborate much on the effect of the course characteristics. In future work, courses should be analysed over multiple years, to determine whether the prediction models still differ when course characteristics are kept as similar as possible. Additionally, to explore which specific course characteristics influence the effects of the predictors, a larger amount of courses with more course characteristics should be analysed. In this way, it could be determined which characteristics need to be similar to be able to use a prediction model in multiple courses.

Fortunately, for a few predictors the portability between courses is high, and the prediction models could be used in multiple courses without controlling for student and course characteristics. We found that in-between assessments and prior GPA showed a high predictive power in all courses. Hence, these predictors show a much higher portability than the other LMS and learner variables. However, in-between assessments and GPA do not account for all variance in final exam grade, and sometimes in-between grades are not even available. Thus, when only performance measures are used, the predictability might be rather low.

8.2 Predictability of student performance

The second aim of our study was to compare the value of using LMS data and learner data for predicting student performance. Study 1 showed that LMS data could account on average for 20% of the variance in final grade within the seventeen blended courses. This is somewhat low compared to other studies who predicted student success (Macfadyen & Dawson, 2010; Morris et al., 2005; Rafaeli & Ravid, 1997; Yu & Jo, 2014; Zacharis, 2015). This could be due to the differences in types of LMSs used, the sets of predictor variables examined, and the course characteristics.

Moreover, the low predictability, and also the low portability of the LMS variables, can be explained by the fact that we do not really know what we are measuring. LMSs provide us with raw log data, but these are not concrete measurements. LMS data is for example not a (direct) measurement of motivation. To improve the predictions with LMS data, we need to get a better insight in what the LMS data represents, what the effects are, and how it can be converted into concrete measurements. Steps in this direction were taken in the post-hoc analyses of the first study. In the post-hoc analyses the effects of LMS data on student performance were separated, to determine whether these are effects between students or within a student. Differences in the LMS predictors within students across courses as well as the differences between students in one course, were shown to have an influence on student performance. This indicates that LMS data measures something that has a different effect on student performance across students and across courses. Thus, it is not only useful to compare LMS data with other students within a course, but also with LMS data of the same student in other courses.

Further determination of the meaning of the LMS data should be based on educational theory. Currently, raw log data are pre-processed in different sets of predictor variables, but these predictor variables are not grounded in theory of student learning. Future work should include frameworks based on theories for analysing LMS data. Agudo-Peregrina and colleagues (2014) for example generated LMS predictors from the raw data based on the types of interaction. This gives more insight in which predictors should be used for the prediction models, and could improve the predictability and portability of the prediction models.

Contrary to LMS data, learner data do provide more concrete and robust measurements, and might thus be more useful in predicting student performance. Therefore, in our second study we combined LMS data with learner data, to determine which source is most useful in predicting student performance, and whether learner data and LMS data explain a unique part of the variance in final exam grade. Unfortunately, as learner data was not available for all courses, the analyses of study 2 were restricted to five courses. As the subsample in study 2 was significantly different from the whole sample in study 1, no general conclusions could be draw about the whole sample. Therefore, all conclusions are restricted to these five courses. Multi-level analysis showed that for these five courses 38% of the variance could be explained at student level, and 9% at course level.

First, it was examined whether learner data could explain part of this variance at student level. It was found that learner data could account on average for 29% of the variance in final exam grade in the five courses. This amount is within the range of what other studies found when analysing the effects of trait and state variables on student performance (Britton & Tesser, 1991; Dollinger et al., 2008; Kaufman et al., 2008). The amount of variance explained was mostly due to prior GPA, which corroborates previous findings that past performance is an important and robust predictor for student performance. All other predictors showed no effect, or only a small effect in one or two of the courses. This is in contrast with previous literature in social sciences which reported robust effects of the predictors on student performance. For example, conscientiousness was found a stable predictor in a meta-analysis on personality traits (O'Connor & Paunonen, 2007), and time management and motivation have been pointed out as significant predictors as well (Britton & Tesser, 1991; Kaufman et al., 2008). Moreover, a previous longitudinal study on the same university, with similar measures for the capacities, external regulation, and amotivation, did find a significant result for all these measures on study progress and study drop-out (Bipp et al., 2013).

These differences in results can be (partly) explained by the fact that in the questionnaires in the current study were completed two to seven months before the students started their study program at the university. Thus, some of the state variables (all motivational variables, time management, (lack of) learning strategy, and self-efficacy), might have been changed in the meanwhile. Moreover, the motivational variables measured motivation for the study program as a whole, not for a specific course. Future work should include motivations for courses itself, measured right before the start of the course, as these might have more influence on the final exam grade of the specific course. Future work should also reassess the motivation when the course has started for a few weeks, when the students know somewhat better what to expect of the course. This might have an even better predictive power.

Furthermore, in study 2 the predictive value of LMS data was compared to learner data. It was found that learner data could explain 14% more of the variance in final exam grade compared to LMS data. However, when in-between assessment grades were added to the LMS data, LMS data could predict more than learner data.

Last, the combined predictive effect of both LMS data with in-between assessment data and learner data was explored. Multi-level analysis showed that after adding both data sources to the model, the amount of variance in final exam grade remaining at student level dropped to 15%, and 13% at the course level. Less variance remained at student level than when only LMS data with in-between assessment data, or learner data was added, thus both sources can explain a unique part of the variance in final exam grade. Hence, as expected, adding learner data to LMS data with in-between assessment data led to an increase in the prediction of the separate models per course. Though, the increase was only slightly. Regressions over time showed that learner data were especially useful early in the course, when in-between assessment grades are not available yet. These findings are in line with Tempelaar et al. (2015), who also found that up to in-between performance measures were available, learner dispositions were highly useful predictors. Thus, adding learner data is very useful for early prediction, but does not have much added value when in-between assessment grades are available. As learner data are not easily collected, one might argue to omit learner data in these cases.

Though adding learner data to LMS data is useful for early prediction, the prediction is still not accurate. Early prediction of final exam grade is on average 1.35 away from accurate prediction (on a scale from 0 to 10). Additionally, binary logistic regressions showed that predicting pass or fail probabilities is also less accurate than would be desirable. When these predictions would be used for intervention, 26% of the students will not get feedback, while they needed it, and therefore still might fail the course. Moreover, 29% of the students who did not need the intervention do get feedback. This might even influence their self-efficacy and motivation. For example, Jayaprakash et al. (2014) found that students who did get an intervention showed higher withdrawal rates than students who did not get an intervention. Hence, the prediction must be as accurate as possible, to avoid the chance of an unnecessary withdrawal.

To improve the (early) prediction, it might be useful to include more predictors. After adding LMS data and learner data to the model, the multi-level analysis showed that still 15% of the variance in final exam grade remained at student level, and 13% at course level. This shows that there are relevant predictors that were not included in the models in the present study. To improve the predictability at the student level, first of all, next to qualitative LMS data, quantitative LMS data could be added. Especially data from the discussion forum or wikis might give more information on the type of participation of the student in the LMS (Davies & Graff, 2005; Nandi et al., 2011) and could thereby improve the prediction models. This could also give more insight in students who show high participation but receive low grades (Morris et al., 2005). Secondly, as not all learning behaviour occurs within the LMS, behaviour outside the LMS should be considered too. For example, lecture attendance (Agudo-Peregrina et al., 2014), behaviour in informal networks, and behaviour in other (informal) learning tools (Tempelaar et al., 2015), could be included as well to improve the prediction models. Moreover, still 13% of the variance can be explained at the course level. This indicates that including course characteristics might not only improve the portability of the prediction models across courses, but also the predictability of student performance.

To conclude, this study gained more insight in what LMS data represents, its value in predicting student performance compared to learner data, and the portability across courses, when controlling for student and course characteristics. This has brought interesting new venues for further exploring the field of learning analytics.

9 References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). 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. *Computers in Human Behavior*, *31*, 542–550. http://doi.org/10.1016/j.chb.2013.05.031

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). New York, NY, USA: ACM. http://doi.org/10.1145/2330601.2330666

Baker, R. S. J. d, & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM - Journal of Educational Data Mining*, *1*(1), 3–17.

Beer, C., Clark, K., & Jones, D. (2010). Indicators of engagement. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference*. Retrieved from http://cms.ascilite.org.au/conferences/sydney10/procs/Beer-full.pdf

Bipp, T., Kleingeld, A., & Schinkel, S. (2013). Bachelor Entrance Study (BEST), Onderzoek naar studiesucces en drop-out binnen de Bachelor opleidingen aan de faculteit Industrial Engineering & Innovation Sciences aan de TU/e.

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, *83*(3), 405–410. http://doi.org/10.1037/0022-0663.83.3.405

Campbell, J. P., & Oblinger, D. G. (2007). Academic Analytics. Educause.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality*, *40*(3), 339–346. http://doi.org/10.1016/j.jrp.2004.10.003

Conijn, M. A., & Snijders, C. C. P. (submitted). The value of LMS data for predictive modeling of student performance - A study on 17 blended courses using Moodle LMS. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*. Edinburgh, Scotland.

Davies, J., & Graff, M. (2005). Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, *36*(4), 657–663.

Dawson, S., McWilliam, E., & Tan, J. P.-L. (2008). Teaching smarter: how mining ICT data can inform and improve learning and teaching practice. *Graduate School of Medicine - Papers (Archive)*, 221–230.

Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and Validation of a Dutch Translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, *90*(2), 152–157. http://doi.org/10.1080/00223890701845229

Dollinger, S. J., Matyja, A. M., & Huber, J. L. (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality*, *42*(4), 872–885. http://doi.org/10.1016/j.jrp.2007.11.007

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting action research with learning analytics. In *Proceedings of the 3rd International Conference on Learning*

Analytics and Knowledge (pp. 220–229). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2460340

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. http://doi.org/10.1016/j.iheduc.2015.10.002

Graf, S., Ives, C., Rahman, N., & Ferri, A. (2011). AAT: a tool for accessing and analysing students' behaviour data in learning systems. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 174–179). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2090145

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*(3), 42–57.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, *24*(3), 175–213.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology*, *92*(2), 316.

Hu, Y.-H., Lo, C.-L., & Shih, S.-P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, *36*, 469–478. http://doi.org/10.1016/j.chb.2014.04.002

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Jenson, R., E. (1953). Predicting scholastic achievement of first-year graduate students. *Educational and Psychological Measurement*, *13*, 322–329.

Kaufman, J. C., Agars, M. D., & Lopez-Wagner, M. C. (2008). The role of personality and motivation in predicting early college academic success in non-traditional students at a Hispanic-serving institution. *Learning and Individual Differences*, *18*(4), 492–496. http://doi.org/10.1016/j.lindif.2007.11.004

Kleijn, W. C., Topman, R. ., & Ploeg, H. M. van der. (1994). Cognities, studiegewoonten en academische prestatie: de ontwikkeling van de Studie-management en academische resultaten test (SMART). *Nederlandsch tijdschrift voor de psychologie Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, *49*(5), 231–233.

Lauría, E. J. M., Baron, J. D., Devireddy, M., Sundararaju, V., & Jayaprakash, S. M. (2012). Mining academic data to improve college student retention: an open source perspective (p. 139). ACM Press. http://doi.org/10.1145/2330601.2330637

Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. http://doi.org/10.1016/j.compedu.2009.09.008

Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, *65*(2), 125–139. http://doi.org/10.1016/j.ijhcs.2006.08.008

Milne, J., Jeffrey, L. M., Suddaby, G., & Higgins, A. (2012). Early identification of students at risk of failing. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference* (Vol. 2012). Retrieved from http://www.ascilite.org.au/conferences/wellington12/2012/images/custom/milne,_john_-_early_identification.pdf

Minaei-Bidgoli, B., & Punch, W. F. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In *Genetic and Evolutionary Computation—GECCO 2003* (pp. 2252–2263). Springer. Retrieved from http://link.springer.com/chapter/10.1007/3-540-45110-2_119

Morris, L. V., Finnegan, C., & Wu, S.-S. (2005). Tracking Student Behavior, Persistence, and Achievement in Online Courses. *Internet and Higher Education*, *8*(3), 221–231. http://doi.org/10.1016/j.iheduc.2005.06.009

Nandi, D., Hamilton, M., Harland, J., & Warburton, G. (2011). How active are students in online discussion forums? In *Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114* (pp. 125–134). Australian Computer Society, Inc. Retrieved from http://dl.acm.org/citation.cfm?id=2459952

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990. http://doi.org/10.1016/j.paid.2007.03.017

Petropoulou, O., Retalis, S., Siassiakos, K., Karamouzis, S., & Kargidis, T. (2008). Helping educators analyse interactions within networked learning communities: a framework and the AnalyticsTool system. In *Proceedings of 6th International Conference on Networked Learning* (pp. 5–7). Retrieved from http://nlc.ell.aau.dk/past/nlc2008/abstracts/PDFs/Petropoulou_317-324.pdf

Piña, A. A. (2012). An overview of learning management systems. In *Virtual Learning Environments: Concepts, methodologies, tools and applications* (First edition, pp. 33–51). Sullivan University System, USA: IGI Global.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Rafaeli, S., & Ravid, G. (1997). Online, web-based learning environment for an information systems course: Access logs, linearity and performance. In *Information Systems Education Conference (ISECON)* (Vol. 97, pp. 92–99). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.9119&rep=rep1&type=pdf

Retalis, S., Papasalouros, A., Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). Towards Networked Learning Analytics–A concept and a tool. In *Proceedings of the 5th International Conference on Networked Learning*. Retrieved from http://nlc.ell.aau.dk/past/nlc2006/abstracts/pdfs/P41%20Retalis.pdf

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. http://doi.org/10.1002/cae.20456

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *40*(6), 601–618. http://doi.org/10.1109/TSMCC.2010.2053532

Schell, J., Lukoff, B., & Alvarado, C. (2014). Using Early Warning Signs to Predict Academic Risk in Interactive, Blended Teaching Environments. Retrieved from http://digitalcommons.apus.edu/cgi/viewcontent.cgi?article=1042&context=internetlearning

Shum, S. B., & Crick, R. D. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92–101). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2330629

Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society*, *15*(3), 3–26.

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2330661

Superby, J. F., Vandamme, J.-P., & Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 37–44). Taiwan.

Tempelaar, D. T., Heck, A., Cuypers, H., van der Kooij, H., & van de Vrie, E. (2013). Formative Assessment and Learning Analytics. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 205–209). New York, NY, USA: ACM. http://doi.org/10.1145/2460296.2460337

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 Behavior*, *47*, 157–167. http://doi.org/10.1016/j.chb.2014.05.038

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice*, *93*, 27–30.

Yu, T., & Jo, I.-H. (2014). Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education. In *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (pp. 269–270). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2567594

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, *27*, 44–53. http://doi.org/10.1016/j.iheduc.2015.05.002

Zaïane, O. R., & Luo, J. (2001). Towards evaluating learners’ behaviour in a Web-based distance learning environment. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies* (pp. 357–360). http://doi.org/10.1109/ICALT.2001.943944

Appendix A: Questionnaire learner data (Dutch)

Cap2 – Consciëntieusheid [1 - 5]:
1. Ik ben een persoon die grondig te werk gaat
2. Ik ben een persoon die volhoudt tot de taak af is
3. Ik ben een persoon die doorgaans geneigd is tot slordigheid (1↔5)
4. Ik ben een persoon die geneigd is lui te zijn (1↔5)
5. Ik ben een persoon die een werker waar men van op aan kan
6. Ik ben een persoon die dingen efficiënt doet
7. Ik ben een persoon die plannen maakt en deze doorzet
8. Ik ben een persoon die gemakkelijk afgeleid is (1↔5)
9. Ik ben een persoon die een beetje nonchalant kan zijn (1↔5)

Cap3 – Timemanagement [1 - 5]:
1. Ik heb grote moeite om studie en vrije tijd te combineren (1↔5)
2. Ik kan studie en vrije tijd goed indelen
3. Ik heb grote moeite om geregeld te studeren (1↔5)
4. Ik begin op tijd een proefwerk/tentamen voor te bereiden

Cap4 – Leerstrategie [1 - 5]:
1. Ik weet niet zeker hoe ik moet studeren voor de vakken in de opleiding die ik op dit moment volg
2. Ik merk vaak dat ik niet weet wat ik moet bestuderen of waar ik moet beginnen
3. Het ontbreekt me aan een studiestrategie voor de opleiding die ik op dit moment volg

Cap5 – Academisch zelfvertrouwen [1 - 5]:
1. Ik verwacht goed te presteren vergeleken met andere studenten die deze opleiding gaan volgen
2. Ik denk dat ik in deze opleiding goede cijfers zal halen
3. Ik denk dat ik vergeleken met anderen een goede student ben
4. Ik weet dat ik in staat ben de lesstof van deze opleiding te leren
5. Mijn studievaardigheden zijn uitmuntend vergeleken met andere studenten die deze opleiding gaan volgen
6. Ik denk dat ik vergeleken met andere studenten in deze opleiding veel weet van het vakgebied
7. Ik verwacht het heel goed te doen op deze opleiding
8. Ik weet zeker dat ik uitstekend kan presteren bij de cases en taken die ik in deze opleiding moet doen
9. Ik ben er zeker van dat ik de stof kan begrijpen die in deze opleiding onderwezen wordt

Mot1 – Binding met opleiding [1 - 7]:
1. Deze opleiding past heel goed bij mijn interesses
2. De beroepen die ik na deze opleiding kan uitoefenen passen heel goed bij mijn interesses
3. Ik heb een goed beeld van wat deze opleiding inhoudt
4. Als ik deze opleiding zou kiezen, dan zou ik mijn toekomst met vertrouwen en optimisme tegemoet kunnen zien
5. Het is mij duidelijk wat de opleiding van mij verwacht
6. Ik heb een goed beeld van wat voor werk en carrière ik na mijn opleiding wil

Mot2 – Zekerheid studiekeuze [1 - 7]:
1. Ik weet zeker dat het een goede keuze is om deze opleiding te gaan volgen
2. Een HBO-opleiding is een reëel alternatief voor mij (1↔7)
3. Ik twijfel tussen meerdere TU/e opleidingen (1↔7)
4. Ik twijfel tussen TU/e en andere universiteiten (1↔7)

Mot3 – Motivatie studiekeuze [1 - 7]:
1. Er zijn wellicht goede redenen om deze opleiding te doen, maar persoonlijk zie ik er geen
2. Als ik deze opleiding zou volgen, zou ik er bij de eerste de beste tegenslag zomaar mee op kunnen houden
3. Ik zie niet in wat deze opleiding me oplevert

Mot4 – Zelfregulatie [1 - 7]:
Stel dat je deze opleiding kiest. In welke mate zijn onderstaande redenen dan van toepassing.
1. Omdat ik geen enkele keus heb
2. Omdat het iets is dat ik moet doen
3. Omdat ik verondersteld word om dit te doen
4. Omdat ik het gevoel heb dat ik het moet doen

Appendix: Code of scientific conduct

TU/e Technische Universiteit Eindhoven University of Technology

Declaration concerning the TU/e Code of Scientific Conduct for the Master's/PDEng/PhD thesis

I have read the TU/e Code of Scientific Conduct[1]

I hereby declare that my Master's/PDEng/PhD-thesis has been carried out in accordance with the rules of the TU/e Code of Scientific Conduct

Date

........30 november 2015........

Name

........Rianne Conijn........

ID-nr

........0740635........

Signature

....

[1] See: http://www.tue.nl/en/university/about-the-university/integrity/scientific-integrity/
The Netherlands Code of Conduct for Academic Practice of the VSNU can be found here also. More information about scientific integrity is published on the websites of TU/e and VSNU.