In search for the most informative data for feedback generation: Learning Analytics in a data-rich context

Abstract. Learning analytics seek to enhance the learning processes through systematic measurements of learning related data and to provide informative feedback to learners and educators. Track data from Learning Management Systems (LMS) constitute a main data source for learning analytics. This empirical contribution provides an application of Buckingham Shum and Deakin Crick's theoretical framework of dispositional learning analytics: an infrastructure that combines learning dispositions data with data extracted from computer-assisted, formative assessments and LMSs. In a large introductory quantitative methods module, 922 students were enrolled in a module based on principles of blended learning, combining face-to-face Problem-Based Learning sessions with e-tutorials. We investigated the predictive power of learning dispositions, outcomes of continuous formative assessments and other system generated data in modelling student performance and their potential to generate informative feedback. Using a dynamic, longitudinal perspective, computer-assisted formative assessments seem to be the best predictor for detecting underperforming students and academic performance, while basic LMS data did not substantially predict learning. If timely feedback is crucial, both use-intensity related track data from e-tutorial systems, and learning dispositions, are valuable sources for feedback generation.

Keywords: blended learning; dispositional learning analytics; e-tutorials; formative assessment; learning dispositions

1. Introduction

Learning analytics provide institutions with opportunities to support student progression and to enable personalised, rich learning (Bienkowski, Feng, & Means, 2012; Oblinger, 2012; Siemens, Dawson, & Lynch, 2013; Tobarra, Robles-Gómez, Ros, Hernández, & Caminero, 2014). With the increased availability of large datasets, powerful analytics engines (Tobarra et al., 2014), and skilfully designed visualisations of analytics results (González-Torres, García-Peñalvo, & Therón, 2013), institutions may be able to use the experience of the past to create supportive, insightful models of primary (and perhaps real-time) learning processes (Author B, Submitted; Baker, 2010; Stiles, 2012). According to Bienkowski et al. (2012, p. 5), "education is getting very close to a time when personalisation will become commonplace in learning", although several researchers (García-Peñalvo, Conde, Alier, & Casany, 2011; Greller & Drachsler, 2012; Stiles, 2012) indicate that most institutions may not be ready to exploit the variety of available datasets for learning and teaching.

Many learning analytics applications use data generated from learner activities, such as the number of clicks (Siemens, 2013; Wolff, Zdrahal, Nikolov, & Pantucek, 2013), learner participation in discussion forums (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Macfadyen & Dawson, 2010), or (continuous) computer-assisted formative assessments (Author A, 2012a, 2012b; Wolff et al., 2013). User behaviour data are frequently supplemented with background data retrieved from learning management systems

(LMS) (Macfadyen & Dawson, 2010) and other student admission systems, such as accounts of prior education (Arbaugh, 2014; Author A, 2012a; Richardson, 2012). For example, in one of the first learning analytics studies focused on 118 biology students, Macfadyen and Dawson (2010) found that some (# of discussion messages posted, # assessments finished, # mail messages sent) LMS variables but not all (e.g., time spent in the LMS) were useful predictors of student retention and academic performance.

Buckingham Shum and Deakin Crick (2012) propose a dispositional learning analytics infrastructure that combines learning activity generated data with learning dispositions, values and attitudes measured through self-report surveys, which are fed back to students and teachers through visual analytics. For example, longitudinal studies in motivation research (Author B, 2012a; Järvelä, Hurme, & Järvenoja, 2011) and students' learning approaches (Nijhuis, Segers, & Gijselaers, 2008) indicate strong variability in how students learn over time in face-to-face settings (e.g., becoming more focussed on deep learning rather than surface learning), depending on the learning design, teacher support, tasks, and learning dispositions of students. Indeed, in a study amongst 730 students Author A (2012a) found that positive learning emotions contributed positively to becoming an intensive online learner, while negative learning emotions, like boredom, contributed negatively to learning behaviour. Similarly, in an online community of practice of 133 instructors supporting EdD students, Nistor et al. (2014) found that self-efficacy (and expertise) of instructors predicted online contributions.

However, a combination of LMS data with intentionally collected data, such as self-report data stemming from student responses to surveys, is an exception rather than the rule in learning analytics (Author A, 2013a; Buckingham Shum & Ferguson, 2012; Greller & Drachsler, 2012; Macfadyen & Dawson, 2010). In our empirical contribution focusing on a large scale module in introductory mathematics and statistics, we aim to provide a practical application of such an infrastructure based on combining longitudinal learning and learner data. In collecting *learner data*, we opted to use three validated self-report surveys firmly rooted in current educational research, including learning styles (Vermunt, 1996), learning motivation and engagement (Martin, 2007), and learning emotions (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). This operationalisation of learning dispositions closely resembles the specification of cognitive, metacognitive and motivational learning factors relevant for the internal loop of informative tutoring feedback (e.g., Narciss, 2008; Narciss & Huth, 2006). For *learning data*, data sources are used from more common learning analytics applications, and constitute both data extracted from an institutional LMS (González-Torres et al., 2013; Macfadyen & Dawson, 2010) and system track data extracted from the e-tutorials used for practicing and formative assessments (e.g., Author A, 2012b; Author A, 2013a; Wolff et al., 2013). The prime aim of the analysis is predictive modelling

(Baker, 2010; Sao Pedro, Baker, Gobert, Montalvo, & Nakama, 2013), with a focus on the roles of (each of) 100+ predictor variables from the several data sources can play in generating timely, informative feedback for students.

2 Literature review

2.1 Learning Analytics

A broad goal of learning analytics is to apply the outcomes of analysing data gathered by monitoring and measuring the learning process (Buckingham Shum & Ferguson, 2012; Siemens, 2013). A vast body of research on student retention (Credé & Niehorster, 2012; Marks, Sibley, & Arbaugh, 2005; Richardson, 2012) indicates that academic performance can be reasonably well predicted by a range of demographic, academic integration, social integration, psycho-emotional and social factors, although most predictive models can explain only up to 30% of variance. Recent studies in learning analytics (Agudo-Peregrina et al., 2014; Author A, 2013a; Macfadyen & Dawson, 2010; Wolff et al., 2013) seem to indicate that adding LMS user behaviour to these models can substantially improve the explained variance of academic performance. However, according to Agudo-Peregrina et al. (2014) there is no consensus in the learning analytics community on which user behaviour and interactions data are appropriate to measure, understand and model learning processes and academic performance.

Clow (2013, p. 692) argues that "as a field, learning analytics is data-driven and is often atheoretical, or more precisely, is not explicit about its theoretical basis". Although several researchers have worked to link learning analytics to pedagogical theory (Clow, 2013; Dawson, 2008; Macfadyen & Dawson, 2010; Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008), this is still the exception, rather than the rule. However, Macfadyen and Dawson (2010, p. 597) note that "knowledge of actual course design and instructor intentions is critical in determining which variables can meaningfully represent student effort or activity, and which should be excluded". For example, Author A (2013a) found empirical evidence for the role of a broad range of learning dispositions in learning analytics applications in a study amongst 1832 students. Demographic characteristics, cultural differences, learning styles, learning motivation and engagement, and learning emotions, all proved to be facets of learning dispositions having a substantial impact on learning mathematics and statistics. This study extends the analysis of predictive modelling for generating learning feedback by looking at the role of any data source in a multivariate context, so in the presence of several alternative data sources.

In Verbert, Manouselis, Drachsler, and Duval (2012), six objectives are distinguished in using learning analytics: predicting learner performance and modelling learners, suggesting relevant learning resources, increasing reflection and awareness, enhancing social learning environments, detecting undesirable learner behaviours, and detecting affects of learners. Although the combination of self-report learner data with learning data extracted from e-tutorial systems (see below) allows us to contribute to at least five of these objectives of applying learning analytics(as described in Narciss & Huth, 2006)(as described in Narciss & Huth, 2006), we will focus in this contribution on the first objective: predictive modelling of performance and learning behaviour (Baker, 2010; Sao Pedro et al., 2013). The ultimate goal of this predictive modelling endeavour is to find out which components from a rich set of data sources best serve the role of generating timely, informative feedback and signalling risk of underperformance.

2.2 Formative Testing and feedback

A classic function of testing is that of taking an aptitude test. After completion of the learning process, we expect students to demonstrate mastery of the subject. According to test tradition, feedback resulting from such "classical" tests are typically limited to a grade (Boud & Falchikov, 2006; Whitelock, Richardson, Field, Van Labeke, & Pulman, 2014). Another limitation of classical summative testing is that feedback becomes available only after finishing all learning activities (Segers, Dochy, & Cascallar, 2003). An alternative form of assessment, formative assessment, has an entirely different function: that of informing student and teacher (Segers et al., 2003). This information should help to better shape teaching and learning and is especially useful when it becomes available prior to or during the learning process. Feedback plays a crucial part to assist regulating learning processes (Boud & Falchikov, 2006; Hattie, 2009; Lehmann, Hähnlein, & Ifenthaler, 2014; Whitelock et al., 2014). Several alternative operationalisations to support feedback are possible. For example, using two experimental studies with different degrees of generic and directed prompts, Lehmann et al. (2014) found that directed prereflected prompts encourage positive activities in online environments. In a meta-study of 800+ meta-studies, Hattie (2009) found that the way students receive feedback was one of the most powerful factors in enhancing learning experiences. Diagnostic testing is an example of this, just as is a test-directed learning approach that constitutes the basic educational principle of many e-tutorial systems (Author A, 2009).

Because feedback from tests constitutes a main function for learning, it is crucial that this information is readily available, preferably even instantly. At this point digital testing comes on the scene: it is unthinkable to get feedback from formative assessments in time without using computers. Previous research by Wolff et al.

(2013) found that a combination of LMS data with data from continuous assessments were the best predictor for performance drops amongst 7701 students. In particular, the number of clicks in an LMS just before the next assessment significantly predicted continuation of studies (Wolff et al., 2013). Similarly, in a study of six online and two blended courses, Agudo-Peregrina et al. (2014) found that interactions with assessment tools, followed by interactions with peers and teachers, and active participation significantly predicted academic performance in the six online courses. However, no clear paths of learning analytics data were found for the two blended courses. In contrast, Author A (2013a) did find that both dispositional data and data extracted from formative testing had a substantial impact on student performance in a blended course of 1832 students.

2.3 Case Study: Mathematics and Statistics

Our empirical contribution focuses on freshmen students in quantitative methods (mathematics and statistics) of the business & economics school at Maastricht University. This education is directed at a large and diverse group of students, which benefits the research design. As a basic LMS system, Blackboard is used to share basic course information to students. Given the restricted functionality of this LMS in terms of personalised, adaptive learning content with rich varieties of feedback and support provision (for a detailed critique on the limitations of LMS, see Conde, García, Rodríguez-Conde, Alier, & García-Holgado, 2014; García-Peñalvo et al., 2011), two external e-tutorials were utilised: MyStatLab (MSL) and MyMathLab (MML). These e-tutorials are generic LMSs for learning statistics and mathematics developed by the publisher Pearson.

Although MyLabs can be used as a learning environment in the broad sense of the word (it contains, among others, a digital version of the textbook), it is primarily an environment for test-directed learning and practicing. Each step in the learning process is initiated by a question, and students are encouraged to (try to) answer each question. If a student does not master a question (completely), she/he can either ask for help to solve the problem step-by-step (Help Me Solve This), or ask for a fully worked example (View an Example), as demonstrated in Figure 1. These two functionalities are examples of Knowledge of Result/response (KR) and Knowledge of the Correct Response (KCR) types of feedback; see Narciss (2008; 2006). After receiving this type of feedback, a new version of the problem loads (parameter based) to allow the student to demonstrate his/her newly acquired mastery. When a student provides an answer and opts for 'Check Answer', Multiple-Try Feedback (MTF, Narciss, 2008) is provided, whereby the number of times feedback is provided for the same task depends on the format of the task (only two for a multiple choice type of task as in Figure 1, more for open type tasks requiring numerical answers).

3. Research methods

3.1 Research questions

While an increasing body of research is becoming available how students' usage and behaviour in LMS influences academic performance (e.g., Arbaugh, 2014; Macfadyen & Dawson, 2010; Marks et al., 2005; Wolff et al., 2013), how the use of e-tutorials or other formats of blended learning effects performance (e.g., Lajoie & Azevedo, 2006), and how feedback based on learning dispositions stimulates learning (Buckingham Shum and Deakin Crick (2012), to the best of our knowledge no study has looked at how all these factors can be combined into one research context, and what the relative contributions of LMSs, formative testing, e-tutorials, and applying dispositional learning analytics to student performance are. In our empirical contribution focusing on a large scale module in introductory mathematics and statistics followed by 922 students, we aim to provide a practical application of such an infrastructure based on combining longitudinal learning data from our LMS, the two e-tutorials, and (self-reported) learner data. The prime aim of the analysis is predictive modelling (Baker, 2010; Sao Pedro et al., 2013; Wolff et al., 2013), with a focus on the role each of these data sources can play in generating timely, informative feedback for students.

Q1 To what extent do (self-reported) learning dispositions of students, LMSs and e-tutorial data (formative assessments) predict academic performance over time?

Q2 To what extent do predictions based on these alternative data sources refer to unique facets of performance, and to what extent do these predictions overlap?

Q3 Which source(s) of data (learning dispositions, LMS data, e-tutorials formative tests) provide the most potential to provide timely feedback for students?

3.2 Methodology

3.2.1 Context of study

The educational system in which students learn mathematics and statistics is best described as a 'blended' or 'hybrid' system. The main component is face-to-face: problem-based learning (PBL), in small groups (14 students), coached by a content expert tutor (Author A, 2009; Author B, 2009; Schmidt, Van Der Molen, Te Winkel, & Wijnen, 2009). Participation in these tutorial groups is required, as for all courses based

on the Maastricht PBL system. Optional is the online component of the blend: the use of the two e-tutorials (Author A, 2013a). This optional component fits the Maastricht educational model, which is student-centred and places the responsibility for making educational choices primarily on the student (Author A, 2013a; Schmidt et al., 2009). At the same time, due to strong diversity in prior knowledge in mathematics and statistics, not all students, in particular those at the high end, will benefit equally from using these environments. However, the use of e-tutorials and achieving good scores in the practicing modes of the MyLab environments is stimulated by making bonus points available for good performance in the quizzes. Quizzes are taken every two weeks and consist of items that are drawn from the same item pools applied in the practicing mode. We chose for this particular constellation as it stimulates students with limited prior knowledge to make intensive use of the MyLab platforms. Students with limited prior knowledge may realise that they fall behind other students, and therefore need to achieve a good bonus score both to compensate, and to support their learning. The most direct way to do so is to frequently practice in the MML and MSL environments. The bonus is maximised to 20% of what one can score in the exam.

The student-centred characteristic of the instructional model requires, first and foremost, adequate informative feedback to students so that they are able to monitor their study progress and their topic mastery in absolute and relative sense. The provision of relevant feedback starts on the first day of the course when students take two diagnostic entry tests for mathematics and statistics (Author A, 2013a). Feedback from these entry tests provides a first signal of the importance for using the MyLab platforms. Next, the MML and MSL-environments take over the monitoring function: at any time students can see their progress in preparing the next quiz, get feedback on the performance in completed quizzes, and on their performance in the practice sessions. The same (individual and aggregated) information is also available for the tutors in the form of visual dashboards (Clow, 2013; González-Torres et al., 2013; Verbert et al., 2012). Although the primary responsibility for directing the learning process is with the student, the tutor acts complementary to that self-steering, especially in situations where the tutor considers that a more intense use of e-tutorials is desirable, given the position of the student concerned. In this way, the application of learning analytics shapes the instructional support.

3.2.2 Participants

The most recent cohort of freshmen (2013/2014) containing 922 students were included, who in some way participated in learning activities (i.e., have been active in BlackBoard). A large diversity in the student

population is present: only 24% were educated in the Dutch high school system. The largest group, 46% of the freshmen, were educated according to the German Abitur system. High school systems in Europe differ strongly, most particularly in the teaching of mathematics and statistics. Therefore, it is crucial that the first module offered to these students is flexible and allows for individual learning paths (Author A, 2009, 2012a, 2013a). In the investigated course, students work an average 38.2 hours in MML and 24.4 hours in MSL, 30% to 50% of the available time of 80 hours for learning in both topics.

3.3 Instruments and procedure

As illustrated in Figure 2, we will investigate the relationships between a range of of datasources, leading to in total 102 different variables. In the subsections that follow, the several data sources are described that provide the predictor variables for our predictive modelling.

→ Insert Figure 2 about here

3.3.1 Registration systems capturing demographic data

In line with academic retention or academic analytics literature (Marks et al., 2005; Richardson, 2012), several demographic factors are known to influence performance. A main advantage of this type of data is that institutions can relatively easily extract this information from student admission, and are therefore logical factors to include in learning analytics models. Demographic data were extracted from concern systems: nationality, gender, age and prior education. Since, by law, introductory modules like ours need to be based on the coverage of Dutch high school programs, we converted nationality data into an indicator for having been educated in the Dutch high school system. 24% of students are educated in the Dutch higher education system, 76% of students in international systems, mostly of continental European countries. About 39% of students are female, with 61% males. Age demonstrates very little variation (nearly all students are below 20), and no relationship with any performance, and is excluded. The main demographic variable is the type of mathematics track in high school: advanced, preparing for sciences or technical studies in higher education, or basic, and preparing for social sciences (the third level, mathematics for arts and humanities, does not provide access to our program). Exactly two third of the students has a basic mathematics level, one third has an advanced level. (See Author A, 2009, 2012a; 2013a for detailed description)..

3.3.2 Diagnostic entry tests

At the very start of the course, so shaping part of Week0 data, are entry tests for mathematics and statistics all students were required to do. Both entry tests are based on national projects directed at signalling deficiencies in the area of mathematics and statistics encountered in the transition from high school to university (See Author A, 2012b for an elaboration) for an elaboration). Topics included in the entry tests refer to foundational topics, often covered in junior high school programs, such as basic algebraic skills or statistical literacy.

3.3.3 Learning dispositions data

Learning dispositions of three different types were included: learning styles, learning motivation and engagement, and learning emotions. The first two facets were measured at the start of the module, and from the longitudinal perspective are assigned to Week0 data. Learning style data are based on the learning style model of Vermunt (1996, 1998). Vermunt's model distinguishes learning strategies (deep, step-wise, and concrete ways of processing learning topics), and regulation strategies (self, external, and lack of regulation of learning). Recent Anglo-Saxon literature on academic achievement and dropout assigns an increasingly dominant role to the theoretical model of Andrew Martin (2007): the 'Motivation and Engagement Wheel'. This model includes both behaviours and thoughts, or cognitions, that play a role in learning. Both are subdivided into adaptive and mal-adaptive (or obstructive) forms. Adaptive thoughts consist of Self-belief, Value of school and Learning focus, whereas adaptive behaviours consist of Planning, Study management and Perseverance. Maladaptive thoughts include Anxiety, Failure Avoidance, and Uncertain Control, and lastly, maladaptive behaviours include Self-Handicapping and Disengagement. As a result, the four quadrants are: adaptive behaviour and adaptive thoughts (the 'boosters'), mal-adaptive behaviour (the 'guzzlers') and obstructive thoughts (the 'mufflers').

The third component, learning emotions, is more than a disposition: it is also an outcome of the learning process. Therefore, the timing of the measurement of learning emotions is Week4, halfway into the module, so that students have sufficient involvement and experience in the module to form specific learning emotions, but still timely enough to make it a potential source of feedback. Learning emotions were measured through four scales of the Achievement Emotions Questionnaire (AEQ) developed by Pekrun et al. (2011): Enjoyment, Anxiety, Boredom and Hopelessness. All learning dispositions are administered through self-report surveys scored on a 7-point Likert scale.

3.3.4 Learning management system

User track data of LMS are often at the heart of learning analytics applications. Also in our context intensive use of our LMS, BlackBoard (BB), has been made. In line with Agudo-Peregrina et al. (2014), we

captured tracking data from six learning activities. First, the diagnostic entry tests were administered in BB, and through the MyGrades function, students could access feedback on their test attempts. Second, surveys for learning dispositions were administered in BB. Third, two lectures per week were provided, overview lectures at the start of the week, and recap lectures at the end of the week, which were all videotaped and made available as webcasts through BB. Fourth, several exercises for doing applied statistical analyses, including a student project, were distributed through BB, with a requirement to upload solutions files again in BB. Finally, communication from the module staff, various course materials and a series of old exams (to practice the final exam) were made available in BB. For all individual BB items, Statistics Tracking was set on to create use intensity data on BB function and item level.

3.3.5 E-tutorials MyMathLab and MyStatLab

Students worked in the MyMathLab and MyStatLab e-tutorials for all seven weeks, practicing homework exercises selected by the module coordinator. The MyLab systems track three scores achieved in each task, mastery score (MMLMastery), time on task (MMLHours), and number of attempts required to get to the mastery level achieved (MMLAttempts). Those data were aggregated over the on average 25 weekly tasks for mathematics, and about 20 tasks for statistics, to produce six predictors, three for each topic, for each of the seven weeks. Less aggregated data sets have been investigated, but due to high collinearity in data of individual tasks, these produced less stable prediction models.

The three (bonus) quizzes took place in the weeks 3, 5 and 7. Quizzes were administrated in the MyLab tools, and consisted of selections of practice tasks from the two previous weeks.

3.3.6 Academic performance

Six measures of academic performance in the Quantitative Methods module were included for predictive modelling: score in both topic components of the final, written exam (MathExam and StatsExam), aggregated scores for the three quizzes in both topics, MathQuiz and StatsQuiz, overall score in the module, QMTotal (weighting the final exam with weight 5, and the bonus score from quizzes and homework with weight 1), and module passing rate: QMPass.

3.4 Data analysis

Complete information was obtained for 873 out of 922 students (95%) on the various instruments. Prediction models applied in this study are all of linear, hierarchic regression type. More complex models have been investigated, in particular interaction models. However, none of these more advanced model types passed the model selection criterion that prediction models should be stable over all seven weekly intervals. Collinearity existing in track data in a similar way forced us to aggregate that type of data into weekly units; models based on less aggregated data such as individual task data gave rise to collinearity issues.

4. Results

The aim of this study being predictive modelling in a rich data context, we will focus the reporting on the coefficient of multiple correlation, R, of the several prediction models. Although the ultimate aim of prediction modelling is often the comparison of explained variation, which is based on the square of the multiple correlation, we opted for using R itself, to allow for more detailed comparisons between alternative models. Values for R are documented in Table 1 for prediction models based on alternative data sets. For data sets that are longitudinal in nature and allow for incremental weekly data sets, the growth in predictive power is illustrated in time graphs for BB track data, MyLabs track data and test performance data. To ease comparison, all graphs share the same vertical scale.

→ Insert Table 1 about here

4.1 Predicting Performance by demographic data

For the mathematics related performance measures, and for measures relating to completion of the module, there is only one significant predictor variable: mathematics track in high school. Its impact is substantial: its beta weight in predicting, for example, MathExam is 0.43, explaining in itself 20% of variation. However, performance in statistics is different: there exists a substantial impact of the internationalisation dummy, favouring students educated in the Dutch high school system. That impact finds its explanation in the extraordinary role of statistics in the Dutch high school system, in comparison to other continental European countries. Lastly, gender is significant in predicting StatsExam, favouring male students. However, more predictors do not imply better prediction: mathematics performance is much better predicted than statistics performance, with overall performance in an intermediate position, due to lack of coverage in so many high school programs. In other words, in line with previous research (Author A, 2013a; Marks et al., 2005; Richardson, 2012) prior education seems to be a useful factor to include in learning analytics modelling.

11

4.2 Predicting Performance by EntryTest data

Entry test data have substantial predictive power for both performance in mathematics (for MathExam, R=.43, for MathQuiz, R=.45) and overall performance (QMscore, R=.41). These correlations are very similar in value to those of the prior mathematics education variables, indicating that the entry tests provide a good summary of what students have learned in high school. Predictive power for statistics related performance is at a lower level (for StatsExam, R=.30, for StatsQuiz, R=.22), due to the circumstance that many of the students have not been educated before in statistics, so that the entry test cannot be very informative of later performance in the course.

4.3 Predicting Performance by Learning Dispositions data

In terms of predictive power, learning dispositions sit in between BB track data, and the three data sources containing data of more cognitive nature, as is clear from Table 1. Different from MyLab, EntryTest, and demographics predictors, the impact of learning dispositions is of a rather constant level, irrespective of the type of performance measure. For learning styles, R ranges from .21 for passing rate, to .25 for overall score, whereas for the motivation and engagement data the range is from .27 to .34. Learning emotions achieve even higher levels of prediction power, but as noticed before these variables are measured in the midst of the module, so are themselves best viewed as a mixture of disposition and the outcome of the learning process.

The prediction relationships take different shapes, depending on the performance measure. For instance, amongst the learning styles variables, critical processing of learning material, the processing strategy most indicative of deep learning, acts as the most powerful predictor for exam performance, both for mathematics and statistics. In contrast, the regulation strategy self-regulation of learning content is the strongest predictor of quiz performance (with a negative beta, indicating that students who follow their own learning agenda underperform relative to students who adopt the agenda built into the Mylabs).

4.4 Predicting Performance by learning management system data

Given the wealth of BB data, preliminary analysis was applied to find out which indicators of learning intensity performed well in each of the consecutive weeks. BB data is highly collinear, implying different choices of predictor variables in models for each of the seven weeks. The single variable playing a consistent role in all of the weekly models is overall activity in BB: the total number of clicks, per week. Figure 3 demonstrates the

predictive power in terms of the multiple correlation coefficients of longitudinal models developed on overall user activity.

➔ Insert Figure 3 about here

The figure signals two important features. First, there is little progress in predictive power over time: the earliest predictions are about as good as later predictions. It is indeed the case that Week0 BB usage, that is the use of BB in the week before the module starts, has the highest predictive power for the several performance variables of all individual weeks. In line with previous findings (Agudo-Peregrina et al., 2014; Macfadyen & Dawson, 2010), the second observation is that predictive power of our LMS remains low: the multiple correlations of all six performance indicators converges to a value of about 0.2, indicating that no more than about 4% in performance variation can be explained by BB track data. Although there is strong variation in LMS data, this variation is not consistently related with variation in performance. There is one exception to this general result: the number of downloads of old exams for practicing purposes is a reasonable predictor (beta equal to 0.25). However, nearly all of these downloads took place in Week8, the same week as the exam taking place, because of which it is not very useful for a prediction model for providing early feedback to students.

4.5 Predicting Performance by MyMathLab and MyStatLab e-tutorial data

After aggregation to weekly data, three use intensity data remain: mastery level, time on task, and average number of attempts per task, both for MML as for MSL. All of these variables are highly positively correlated: for mathematics e.g., the correlation between mastery level and time for the whole module is .49, the correlation between mastery level and number of attempts is .63, and between time and attempts .47. However, if we include all three predictor variables into one equation, the outcome becomes (for mathematics, estimated over all weeks):

$$MathExam = 0.671 \times MMLMastery - 0.161 \times MMLTime - 0.317 \times MMLAttempts$$

$$MathQuiz = 0.801 \times MMLMastery - 0.149 \times MMLTime - 0.153 \times MMLAttempts$$

with values of R being .51 and .66, respectively. A remarkable and very consistent feature of all prediction equations using mastery, time, and attempts data is that the beta of mastery is always positive, and the beta of time on task and number of attempts are always negative, although all bivariate correlations between time on

task and performance measures are positive. There is, however, a simple explanation for this sign reversal: mastery time on task and attempt variables are strongly collinear. Practising longer in the two MyLab systems increases expected performance, since students who practise more, achieve higher mastery levels. Similarly: redoing a task for a second or third time will generally increase mastery level[i].

Now that the potential of building prediction models for performance based on data from the two MyLab systems has been established, the next step was to design these prediction models using incremental data sets of track data. Starting with the Week1 data set, we extend the data set in weekly steps, arriving after seven weeks at the final set of predictor variables, containing mastery, time on task and number of attempts system data of seven consecutive weeks for MML and MSL systems. Figure 4 describes the development of the multiple correlation coefficients R in time, that is, over subsequent weekly data sets.

➔ Insert Figure 4 about here

Since the predictor data sets are incremental, the values of multiple correlations increase over weeks. Those for performance in the mathematics exam, and the overall score, start at values around 0.4 in Week0, and increase to values between 0.5 and 0.6 in the last week. In other words, being pro-active in the e-tutorials seems to be a good candidate to be included in learning analytics modelling.

4.6 Predicting Performance by Quiz data

That the best predictor for performance, is performance itself, will not surprise many teachers and researchers. Although quizzes in our context are more of formative, than summative type (bringing only a bonus score, to a maximum amount of 20% of what one can score), they constitute the most reliable predictor of all six performance measures. Focusing on performance in the exam (since predicting quiz scores, or total scores, from the quizzes themselves brings about endogeneity issues), multiple correlation values develop from R=.64 to R=.73 for mathematics, and from R=.54 to R=.62 for statistics, over the three quizzes. Figure 5 demonstrates this development in predictive power, where as the starting point of the time trajectories, the EntryTests are used. In line with Wolff et al. (2013), quizzes seem to be a good indicator for learning in prediction modelling.

➔ Insert Figure 5 about here

4.7 Predicting Performance by all Weekly Data

The very last step in assessing the quality of prediction models entails the combination of different data sources in each of the longitudinal models. Figure 6 provides an insight in the development of predictive power in time, when combining all available data.

➔   Insert Figure 6 about here

As indicated before, the predictive power toward the Quiz performance components are an artefact of using predictor variables that more and more coincide with the predicted performance component. The main criterion is the prediction of both exam components of performance. In Week3, multiple correlations R for predicting MathExam and StatsExam are a substantial .72 and .62. Given the importance of Week3 data with regard to potential interventions for students at risk (ie., failing the course and/or dropping out),, Figure 7 provides scatterplots of the prediction equations for the two exam performance components in the first row, and the two quiz performance components in the second row, with mathematics in the first panel, and statistics in the second. Scatterplots produced for later weeks demonstrate higher predictive power, but less time to intervene: with still five full weeks to catch up, Week3 feedback appears to be the best compromise between timely feedback and sufficient high predictive power (the ribbon pattern in the first two panel are a consequence of exam scores expressed as integer numbers).
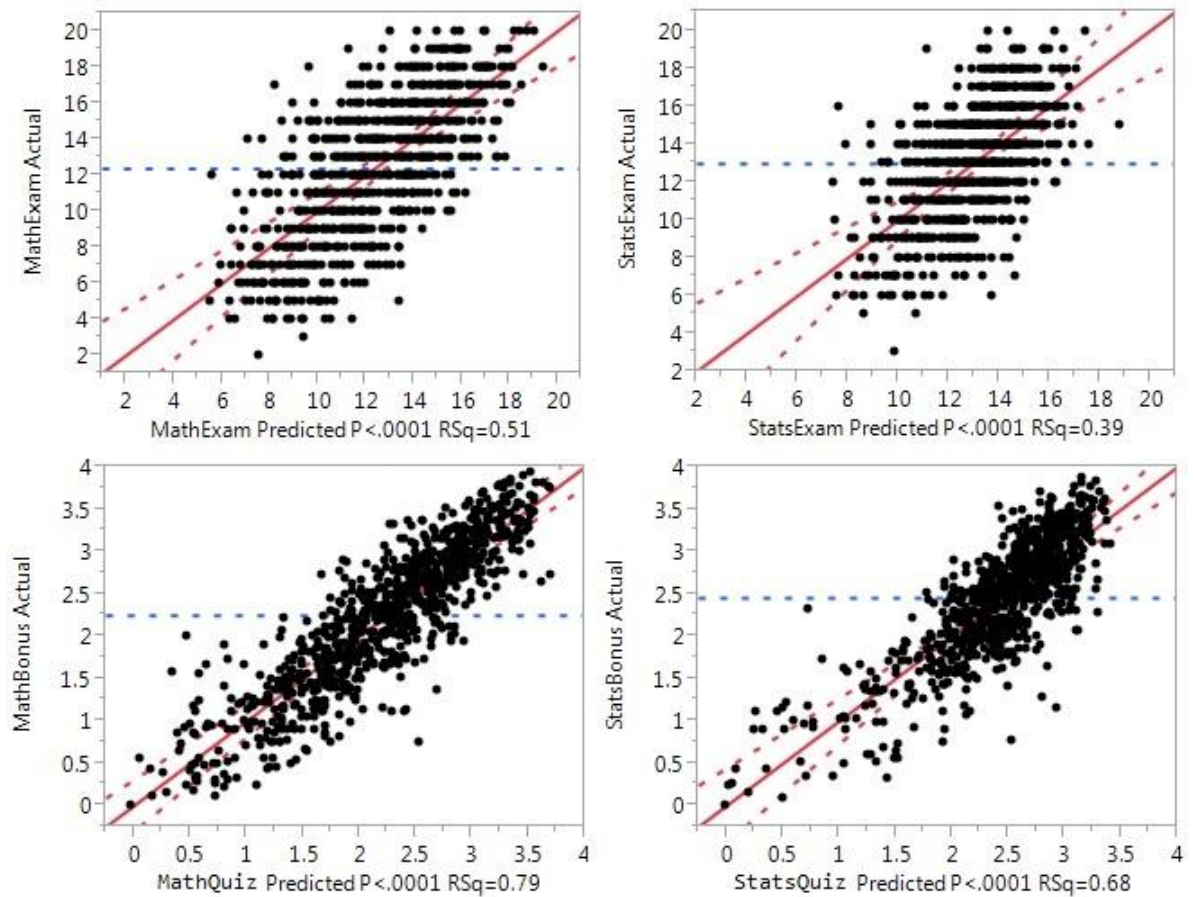
Fig 7: Scatterplots of prediction equations for exam (first row) and quiz (second row) performance, for mathematics (left) and statistics (right)

When we compare predictive power of all data combined, with that of prediction models based on a single data source, there is evidence of considerable overlap in the information content of various data sources. Especially MyLab track data, EntryTest data, Quiz data and prior education data share variation. From that perspective of providing unique information, the learning dispositions data set is most complementary. For example, in the Week0 data set, demographic variables predict MathExam with R=.43, StatsExam with R=.29. Adding learning dispositions to demographic variables increases R to .53 and .42, respectively, with entry testing and BB data having the limited effect of further increasing R to .59 and .45.

Part of the complementary nature of disposition data is in the specific position it takes in predicting the passing rate. Of all performance variables, the passing rate is by far the most difficult to predict, since the required score to pass the test is about at the top of the score distribution. So relatively small differences in test scores make the difference between failing and passing, making it a more difficult phenomenon to predict than the final score itself. From that perspective, disposition data do a relatively good job in pass/fail predictions,

providing support to the notions by Buckingham Shum and Deakin Crick (2012) that learning analytics should combine LMS data with learner data.

5. Discussion

In this empirical study into predictive modelling of student performance, we investigated several different data sources to explore the potential of generating informative feedback for students and teachers using learning analytics: data from registration systems, entry test data, students' learning dispositions, BlackBoard tracking data, tracking data from two e-tutorial systems, and data from systems for formative, computer assisted assessments. In line with recommendations by Agudo-Peregrina et al. (2014), we collected both dynamic, longitudinal user data and semi-static data, such as prior education. It appears that the role of BlackBoard track data in predicting student performance is dominated by the predictive power of any of the other data components, implying that in applications with such rich data available, BlackBoard data have no added value in predicting performance and signalling underperforming students. This seems to confirm initial findings by Macfadyen and Dawson (2010), who found that simple clicking behaviour in a LMS is at best a poor proxy for actual user-behaviour of students.

Data extracted from the testing mode of the MyLab systems, the quiz data, dominate in a similar respect all other data, including data generated by the practicing mode of MyLabs, indicating the predictive power of "true" assessment data (even if it comes from assessments that are more of formative, than summative type). However, assessment data is typically delayed data (Boud & Falchikov, 2006; Whitelock et al., 2014; Wolff et al., 2013), not available before midterm, or as in our case, the third week of the course. Up to the moment this richest data component becomes available, entry test data and the combination of mastery data and use intensity data generated by the e-tutorial systems are a second best alternative for true assessment data. This links well with Wolff et al. (2013), who found that performance on initial assessments during the first parts of online modules were substantial predictors for final exam performance.

A similar conclusion can be made with regards to the learning disposition data: up to the moment that assessment data become available, they serve a unique role in predicting student performance and signalling underperformance beyond system track data of the e-tutorials. From the moment that computer assisted, formative assessment data become available, their predictive power is dominated by that of performance in those formative assessments. Dispositions data are not as easily collected as system tracking data from LMSs or e-tutorial systems (Buckingham Shum & Deakin Crick, 2012). The answer to the question if the effort to collect

dispositional data is worthwhile (or not), is therefore strongly dependent on when richer (assessment) data becomes available, and the need for timely signalling of underperformance. If timely feedback is required, the combination of data extracted from e-tutorials, both in practicing and test modes, and learning disposition data suggests being the best mix to serve learning analytics applications. In contrast to Agudo-Peregrina et al. (2014), who found no consistent patterns in two blended courses using learning analytics, we did find that our mix of various LMS data allowed us to accurately predict academic performance, both from a static and dynamic perspective. The inclusion of extensive usage of computer-assisted tests might explain part of this difference, as well as more fine-grained learning disposition data allowed us to model the learning patterns from the start of the module.

Even in the case dispositions would more strongly overlap other predictor variables, like e.g. prior education, dispositions have a unique position with regard to the final aim of feedback. Feedback is informative if two conditions are satisfied: it is predictive, and allows for intervention. Feedback based on prior education may be strongly predictive, but is certainly incapable of designing interventions as to eliminate the foreseen cause of underperformance (Boud & Falchikov, 2006; Whitelock et al., 2014). Feedback related to learning dispositions, such as signalling suboptimal learning strategies, or inappropriate learning regulation, is generally open to interventions to improve the learning process (Lehmann et al., 2014; Pekrun et al., 2011). So, to the extent learning dispositions share predictive power with alternative aspects of learning, feedback in terms of these dispositions will generally be preferred over feedback framed in any of the other aspects of learning.

These findings strongly support the integrative approach to learning analytics as advocated by Buckingham Shum and Deakin Crick (2012). As '[t]*here is substantial and growing evidence within educational research that learners' orientation towards learning -their learning dispositions- significantly influence the nature of their engagement with new learning opportunities…*' (Buckingham Shum & Deakin Crick, 2012; p. 2), the combination of 'first generation' technology-driven learning analytics with insights from educational research provides the step toward 'second generation' learning analytics. In other words, a development of learning analytics that empowers students to become independent professionals, who can shape their own learning. Future developments should further investigate how to best present feedback based on learning disposition data in combination with technology-generated data to students.

A crucial limitation of our study is that we focussed our analyses based only on formal learning interactions, as measured by the three LMS data. Although we added self-reported data from a range of learning disposition instruments to get a more fine-grained, nuanced understanding of the data, several studies indicate

that students increasingly use informal networks (Agudo-Peregrina et al., 2014; Hommes et al., 2012) and learning tools (e.g., Facebook, twitter, texts) to share knowledge and learn together. For example, Hommes et al. (2012) found that informal social learning links primarily predicted academic performance amongst 300 medical students. Using dynamic social network analyses, Author B (2014) found that 30-80% of learning occurred outside formally assigned groups. Agudo-Peregrina et al. (2014) argue that learning analytics should take into consideration data from Personal Learning Environments (PLE), although several ethical issues (Author B, Submitted) need to be addressed in terms of informed consent if institutions are using PLE data, such as Facebook.

6. Conclusion

The generation of timely feedback based on early performance predictions and early signalling of underperformance are crucial objectives in many learning analytics applications. The added value of data sources for such applications will therefore depend on the predictive power of the data, the timely availability of the data, and the uniqueness of information in the data. In this study, we integrated data from many different sources and found evidence for strong predictive power of data from formative testing. However, not all modules will contain aspects of formative testing, and even if so, data from formative testing might not be timely enough. In that case, data from e-tutorial systems as the MyLabs, both in terms of mastery level and time on task and attempt data constitute a good second best information source, as will do entry test data or prior education data. Learning data from these various sources share a cognitive nature, and thus share important overlap in predictive power. Learner data in the form of learning dispositions have a unique role in such learning analytics applications since its contribution in performance prediction is indeed orthogonal to that of other data sources. In a rich data context as investigated here, the role of BB track data appeared to be minimal.

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*(0), 542-550. doi: 10.1016/j.chb.2013.05.031
Arbaugh, J. B. (2014). System, scholar, or students? Which most influences online MBA course effectiveness? *Journal of Computer Assisted Learning*. doi: 10.1111/jcal.12048
Author A. (2009). [details removed for peer review].
Author A. (2012a). [details removed for peer review].
Author A. (2012b). [details removed for peer review].
Author A. (2013a). [details removed for peer review].
Author B. (2009). [details removed for peer review].
Author B. (2012a). [details removed for peer review].
Author B. (2014). [details removed for peer review].

Author B. (Submitted). [details removed for peer review].

Baker, R. (2010). Data mining for education. *International encyclopedia of education, 7*, 112-118.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1-57.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399-413. doi: 10.1080/02602930600679050

Buckingham Shum, S., & Deakin Crick, R. (2012). *Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics.* Paper presented at the 2nd International Conference on Learning Analytics & Knowledge, Vancouver, British Columbia.

Buckingham Shum, S., & Ferguson, R. (2012). Social Learning Analytics. *Journal of Educational Technology & Society, 15*(3). doi: 10.1145/2330601.2330616

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683-695. doi: 10.1080/13562517.2013.827653

Conde, M. A., García, F., Rodríguez-Conde, M. J., Alier, M., & García-Holgado, A. (2014). Perceived openness of Learning Management Systems by students and teachers in education and technology courses. *Computers in Human Behavior, 31*(0), 517-526. doi: 10.1016/j.chb.2013.05.023

Credé, M., & Niehorster, S. (2012). Adjustment to College as Measured by the Student Adaptation to College Questionnaire: A Quantitative Review of its Structure and Relationships with Correlates and Consequences. *Educational Psychology Review, 24*(1), 133-165. doi: 10.1007/s10648-011-9184-5

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Journal of Educational Technology & Society, 11*(3).

García-Peñalvo, F. J., Conde, M. Á., Alier, M., & Casany, M. J. (2011). Opening learning management systems to personal learning environments. *Journal of Universal Computer Science, 17*(9), 1222-1240. doi: 10.3217/jucs-017-09-1222

González-Torres, A., García-Peñalvo, F. J., & Therón, R. (2013). Human–computer interaction in evolutionary visual software analytics. *Computers in Human Behavior, 29*(2), 486-495. doi: 10.1016/j.chb.2012.01.013

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society, 15*(3).

Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Hommes, J., Rienties, B., de Grave, W., Bos, G., Schuwirth, L., & Scherpbier, A. (2012). Visualising the invisible: a network approach to reveal the informal social side of student learning. *Advances in Health Sciences Education, 17*(5), 743-757. doi: 10.1007/s10459-012-9349-0

Järvelä, S., Hurme, T., & Järvenoja, H. (2011). Self-regulation and motivation in computer-supported collaborative learning environments. In S. Ludvigson, A. Lund, I. Rasmussen & R. Säljö (Eds.), *Learning across sites: new tools, infrastructure and practices* (pp. 330-345). New York, NY: Routledge.

Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2 ed., pp. 803-821). Mahwah, NJ: Erlbaum.

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior, 32*(0), 313-323. doi: 10.1016/j.chb.2013.07.051

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

Marks, R. B., Sibley, S. D., & Arbaugh, J. B. (2005). A Structural Equation Model of Predictors for Effective Online Learning. *Journal of Management Education, 29*(4), 531-563. doi: 10.1177/1052562904271199

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*(2), 413-440. doi: 10.1348/000709906X118036

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (3 ed., pp. 125-144). Mahaw, NJ: Lawrence Erlbaum Associates.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction, 16*(4), 310-322. doi: 10.1016/j.learninstruc.2006.07.003

Nijhuis, J., Segers, M., & Gijselaers, W. (2008). The extent of variability in learning strategies and students' perceptions of the learning environment. *Learning and Instruction, 18*(2), 121-134. doi: 10.1016/j.learninstruc.2007.01.009

Nistor, N., Baltes, B., Dascălu, M., Mihăilă, D., Smeaton, G., & Trăuşan-Matu, Ş. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application. *Computers in Human Behavior, 34*, 339-344. doi: 10.1016/j.chb.2013.10.051

Oblinger, D. G. (2012). Let's Talk... Analytics. *EDUCAUSE Review, 47*(4), 10-13.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology, 36*(1), 36-48. doi: 10.1016/j.cedpsych.2010.10.002

Richardson, J. T. E. (2012). The attainment of White and ethnic minority students in distance education. *Assessment & Evaluation in Higher Education, 37*(4), 393-408. doi: 10.1080/02602938.2010.534767

Sao Pedro, M., Baker, R. S. J., Gobert, J., Montalvo, O., & Nakama, A. (2013). Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill. *User Modeling and User-Adapted Interaction, 23*(1), 1-39. doi: 10.1007/s11257-011-9101-0

Schmidt, H. G., Van Der Molen, H. T., Te Winkel, W. W. R., & Wijnen, W. H. F. W. (2009). Constructivist, Problem-Based Learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist, 44*(4), 227-249. doi: 10.1080/00461520903213592

Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising New Modes of Assessment: In Search of Qualities and Standards.* Dordrecht: Kluwer Academic Publishers.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist, 57*(10), 1380-1400. doi: 10.1177/0002764213498851

Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the quality of productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics: Solarresearch.

Stiles, R. J. (2012). Understanding and Managing the Risks of Analytics in Higher Education: A Guide: Educause.

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education, 50*(4), 1103-1127. doi: 10.1016/j.compedu.2006.10.007

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. *Computers in Human Behavior, 31*(0), 659-669. doi: 10.1016/j.chb.2013.10.001

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-Driven Research to Support Learning and Knowledge Analytics. *Journal of Educational Technology & Society, 15*(3), 133-148.

Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education, 31*(25–50). doi: 10.1007/BF00129106

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

Whitelock, D., Richardson, J., Field, D., Van Labeke, N., & Pulman, S. (2014). *Designing and Testing Visual Representations of Draft Essays for Higher Education Students*. Paper presented at the LAK 2014, Indianapolis.

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). *Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment.* Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge.

i In follow-up multiple regression modeling, time and number of attempts have a negative impact: for a given mastery level, students who need more time to reach that level, or students who need more trials to reach that level, have lower expected performance, which is quite intuitive. Amongst students with complete mastery (mastery level > 95%), the negative correction for time on task is stronger than for students with incomplete mastery. The opposite is true for the correction for number of attempts: amongst students with high mastery, the negative correction for redoing a task is less than for students with low mastery. Although these interaction effects appeared to be stable patterns for both topic areas, we choose not to include them in the development of longitudinal models: the impact of predictive power is small, and it introduces more collinearity, and less parsimony.