Analysing & Predicting Students Performance in an Introductory Computer Science Course

Alexander Tillmann, Detlef Krömker, Florian Horn und Thorsten Gattinger
Goethe-Universität Frankfurt, studiumdigitale
Varrentrappstr. 40–42
60486 Frankfurt am Main
tillmann@sd.uni-frankfurt.de
kroemker@sd.uni-frankfurt.de
horn@sd.uni-frankfurt.de
gattinger@sd.uni-frankfurt.de

Abstract: Students of computer science studies enter university education with very different competencies, experience and knowledge. 145 datasets collected of freshmen computer science students by learning management systems in relation to exam outcomes and learning dispositions data (e. g. student dispositions, previous experiences and attitudes measured through self-reported surveys) has been exploited to identify indicators as predictors of academic success and hence make effective interventions to deal with an extremely heterogeneous group of students.

Keywords: Learning analytics, Learning dispositions, Dispositional learning analytics, Formative assessment, Blended learning, heterogeneity.

1 Introduction

Due to the dramatically increasing student numbers in computer science studies, the heterogeneity of university entrants is growing continuously. Approaches like learning analytics (LA) try to address this diversity by identifying, collecting and analyzing data features from different sources like student information systems (SIS) or learning management systems (LMS). Through systematically analyzing learning-related data it should be possible to support the learning processes of heterogeneous students by recognizing and considering individual requirements. By regularly collecting and analyzing data it could be possible to directly react and offer support to prevent students from dropping out of university. The aim is to provide informative feedback on an individual and organizational level and offer opportunities for institutions to support student progress [BFM12; WI12]. Pattern recognition and predictive analytics have not yet been widely used in educational settings [WI12] and much of the data on which LA applications depend comes from learning management systems where grades are only included in some cases. While such learning analytics approaches demonstrated the potential of LA for feedback on the individual level, the findings based on demographics, grades and data tracked from the LMS were rather limited to the descriptive functions of LA. The limits of SIS and LMS data and the lack of important markers of students' heterogeneity thus make it difficult to design pedagogically informed interventions [CH15]. To overcome this shortcoming recent studies have proposed a Dispositional Learning Analytics (DLA) approach. A DLA infrastructure combines data of learning dispositions that impact learning processes (e. g., student dispositions, values, and attitudes measured using self-reported surveys) with data extracted from LMSs and SISs to optimize the connection with learning interventions [RCZ17]. In our empirical research focusing on a introductory module in computer science and programming we provide an application of the theoretical framework of DLA [BC12]. Furthermore the DLA perspective of individual difference characteristics that impact learning processes is expanded by an approach that focuses on the requirements of the currently heterogeneous student body. In order to ensure the implementation of an adequate learner-centered course design particular attention has to be paid to the increasing heterogeneity of the students. Different dimensions of heterogeneity are of particular relevance for higher education [MSM12]: social heterogeneity (age, family status, migrant background, academic background of the family), cognitive heterogeneity (skills, competences), expectancy (vocational and practice-orientation), motivational heterogeneity (procrastination, pragmatism, self-organization) heterogeneous situations in life (professional activity, part-time study, commuter). The main aim of this predictive modeling endeavor is to identify a set of key variables from a rich set of data as a basis for evidence-based decisions concerning a range of interventions intended to deal with the heterogeneity of undergraduate computer scientists. On the one hand, major academic success factors could be identified from the derived datasets, curricula adjusted accordingly and targeted academic support offered for disadvantaged students (e. g. with little prior knowledge of computer science). On the other hand, the LA infrastructure could serve as a learning coach to provide timely, formative feedback about learning activities and success and indicate a risk of underperformance.

2 Learning Analytics and learning dispositions data

2.1 Learning Analytics in Higher Education

Learning analytics tries to apply the outcomes of analyzing data collected by monitoring and measuring the learning process and its contexts [Te17]. Empirical studies furthermore show that academic success can be well predicted by a range of demographic, psycho-emotional, cognitive, methodological and social factors [CN12]. Models based on learning psychological and pedagogic theories are often only able to explain up to thirty percent of variance. Learning analytics research that predicts academic success from log data in virtual learning environments [Ag14] indicates that the combination of interaction data from the LMS and learning dispositions data can substantially improve the explained variance of learning success. Demographic characteristics, motivation, conscientiousness and commitment, prior knowledge, skills, competences, talent and personality have all proved to be features of learning dispositions with a significant impact on learning in higher education. By looking at the role of several alternative data sources this study extends the analysis of predictive modeling of academic performance and learning behavior.

2.2 Learning dispositions data

In contrast to previous DLA approaches which based their research on a single newly constructed instrument to collect dispositions data [BC12], our study employs well-established and validated instruments. Some learner characteristics and attitudes towards learning can be influenced by education (e. g. promotion of self-reliance, learning strategies). However, other personality qualities like psychometric properties also play a significant role in influencing academic achievement [ZZ16] and are quite stable over time [EKS17; HB17]. Consequently, we cannot try to influence the personality traits of the students so that they better fit to our organization of study programs and curriculum. Having accepting this, we then have to turn the tables: can the various requirements of the students be satisfied by the study program or are adjustments necessary? According to social-constructivist learning theories, learning is an active process of learning construction in which prior knowledge plays an important role [BP12]. We assume that when learning computer programming the effect of prior knowledge is of particularly crucial importance. Unfortunately the application of an entry diagnostic test of programming skills is time-consuming and could have a dissuasive effect on freshmen students with less prior knowledge. Therefore we would like to find out the extent of the predictive power of self-reported prior knowledge in programming for overall course performance. Other DLA research covering various aspects of affective, behavioral and cognitive antecedents of learning processes has found quite weak relationships between the instruments applied and academic performance [Te17]. Our contribution to the DLA research is based on a case study of freshman students in computer science and focuses on the influence of personality traits on academic achievement and the role of prior knowledge.

3 Method

While an increasing body of research is becoming available about the relationship between LMS data and academic performance [e.g. AG14; MD10], about the effects of formative assessment and feedback on learning [BF06], and about the relationship between learning dispositions data covering aspects of affective, behavioral and cognitive antecedents of learning processes [BC12], our empirical research examines the relationship between personality traits and the role of prior knowledge on academic achievement and how all these elements (LMS, formative assessment, learning dispositions data) can be integrated into a research context to analyze the relative contributions of each of the elements to student achievement. With a focus on combining longitudinal learning data extracted from an institutional LMS (including track data extracted from formative tests/quizzes), results from weekly exercises and self-reported learning dispositions data, this study aims to answer the following research questions:

- To what extent does the data of LMSs, formative assessments, personality traits and self-reported learning dispositions predict academic performance over time and what is the most effective data for predictive modeling?
- What is the relationship between learner data sources (LMSs and self-reported data) and assessment data (formative and summative) and to what extent do these predictions overlap?

- Which data source (LMSs, formative assessments or self-reported learner data) has the most potential to generate timely, informative feedback for students?
- To what extent does the LA approach support decisions about pedagogical interventions (e. g. redesign of courses)?

3.1 Description of the course and participants

145 datasets of freshmen computer science students enrolled in 2016/2017 on a module in introductory computer science and programming at Frankfurt Goethe-University could be derived. Most of the students, 70 %, are male. The educational setting in which students learn can be described as a blended learning system. One component consists of lectures which are also recorded and available as e-lectures. Attendance of the lectures is optional. Another component is a face-to-face tutorial course with small groups (15 students) and a problem-based learning approach with weekly exercise sheets (11 sheets about basic knowledge and concepts in computer science and seven sheets introducing programming and code writing). Students have to prepare the exercises and subsequently discuss the solution approach with a tutor. Participation in these tutorial groups is also optional, as are the online components of the blended setting. The LMS system Moodle is used to share basic course information and learning resources like PDFs and quizzes. Overall 24 quizzes are available and consist of items which expand on the content of the lecture and tutorial courses. The use of quizzes and participation in tutorial courses is stimulated by making bonus points available for good performance. Overall the bonus is maximized to 20 % of what can be scored in the exam. Due to the very diverse levels of prior knowledge in computer programming we chose this pedagogical setting as it stimulates students with less prior knowledge to make intensive use of learning resources. Learners with less prior knowledge may realize from the continuous feedback that they are falling behind other students, and therefore need to achieve a good bonus score for both, so as to compensate and improve their learning. A good way to do this is to attend the lecture and tutorial courses regularly and use the learning resources on the LMS with e-lectures, quizzes and literature.

3.2 Data sources and procedure

Learning Management System: Every user action is stored as a database record by Moodle data extraction tool. Following a system-independent classification of interactions in LMSs [Ag14] three different types of interactions associated with virtual learning are possible: student-student interactions (online communication between students, using chats and messages in forums), student-teacher interactions (e. g. interactions involving synchronous and asynchronous tutoring) and student-content interactions (interactions that happen when students make use of the content resources).

In our educational setting students could access e-lectures, various course materials (documents, videos, textbooks) and quizzes and automated feedback on their quiz attempts. These student-content interactions are usually associated with browsing and accessing the different resources, tasks, etc. Communication tools like the E-Mail-Function or course discussion boards were not used systematically, because within our educational setting the students see each other at least twice a week, so there is limited need for online-communication via the LMS. According to the concepts of previous research [Ag14; MD10] and our educational setting, we pre-selected items from the 140 possible interactions tracked in Moodle that describe student-content interactions: accessing the different resources (LMS_content) and the quizzes (LMS_quiz).

Learning dispositions data: Three different types of learning dispositions were included in this study: prior knowledge in computer programming, marks in math, English, German and informatics at school, and personality traits. Prior knowledge in computer programming was measured within the first week of the module. The data is based on an instrument adapted to the German language from the computer programming self-efficacy scale (PSES) [AD09; RW98]. The 7-point scale consists of 28 items and the reliability of the scores was $\alpha = .98$; example item: *I could write a computer program that computes the average of three numbers.* In addition, items for collecting information related to gender and marks in high school in math, English, German and informatics were prepared and delivered to the students via the LMS, together with the third component, the measurement of personality traits. This was undertaken using a short version (21 items) of the Big Five Inventory (BFI-K) for assessment of the five factors of personality with very well-tested psychometric properties [RJ05]. The Big Five Factors are: extraversion, agreeableness, conscientiousness, neuroticism and openness to experience [JSS99].

The inventory was administered with a self-report survey scored on a 5-point Likert scale.

Students' academic performance: For the predictive modeling three measures of academic performance were included: score in written exam (only exam results, without bonus points), aggregated scores for the exercise sheets about basic knowledge and concepts in computer science (ProgrammingBasics [PBasics]) and the sheets for introduction to programming and code writing (ProgrammingPractice [PPractice]).

Data analysis: For further analysis the LMS data was exported into an Excel file and merged with final course exam data, learning dispositions data and the grade data of the exercise sheets. 145 datasets could be derived after merging all data files. Then the dataset was imported into SPSS for statistical analysis. The first step of the analysis consisted of an observation of bivariate correlations between the various datasets and the results of the course exam. Since simply relying on correlations for predictive power may lead to Type II errors, we carried on to confirm the results with multiple regression. For the analysis a data transformation was used to standardize variable data as Z-scores, because the dataset contains variables measured in different scales. The multiple regression was calculated with the final course score achieved by each student as the dependent variable and the LMS datasets, learning dispositions data and results of the exercise sheets as the independent variables. Multiple regressions calculate the variance of the dependent variable as linear combinations of the independent variables. This method makes it possible to generate predictive models for the dependent variable based on data from the independent variables and assigns regression coefficients to each independent variable, allowing us to assess their relative importance in the predictive model. In the study a backwards multiple regression was calculated as this has the advantage that there is no suppression effect such as occurs when independent variables interact with opposite effects, possibly leading to Type II errors [Br13].

4 Results and Discussion

Table 1 shows the predictive power, as multiple correlation r, of alternative longitudinal datasets. All correlation coefficients above 0.2 are significant at $p < 0.01$ (two-tailed) and those of 0.2 or smaller at $p < 0.05$ (two-tailed).

Table 1: Multiple bivariate correlations R between various data sources and three academic performance measures. (n. s. = not significant)

Data source	Exam	PBasics	PPractice
Gender	n. s.	n. s.	.23
Marks in math	.52	.32	.42
in English	.24	n. s.	.26
in German	.44	.27	.35
in informatics	.51	n. s.	.42
PSES	.31	.19	.35
Big five factors:			
extraversion	n. s.	n. s.	n. s.
agreeableness	n. s.	n. s.	n. s.
conscientious	n. s.	n. s.	n. s.
neuroticism	−.18	n. s.	n. s.
openness	n. s.	n. s.	n. s.

4.1 Predictive power of demographic data

Apart from gender (code: female = 1, male = 2) the other tested demographic variables (age and number of semesters) did not show significant correlations with exam results or results from the exercise sheets (PBasic and PPractice). A statistically significant relation was measured between gender and academic performance in solving the exercise sheets with computer programming tasks (PPractice). These exercises were performed better by males. This could be explained by the fact that the group of women had less prior knowledge in computer programming (variable PSES) ($M = 2.7$; $SD = 1.4$) than men ($M = 3.6$; $SD = 1.8$). This difference is significant with $t(137) = -4.2$, $p < 0.01$. Furthermore, there are no gender differences in high school marks, which are also significant predictor variables. This shows the impact of prior knowledge in programming. Even without differences in math, informatics, English or German, women have difficulties catching up with men in programming, because of a lack of prior knowledge.

4.2 Predictive power of learning dispositions

Marks in high school all show significant correlations with performance measures. The impact of marks in math is substantial: its beta weight in predicting the course exam is 0.52, explaining in itself 27 % of variation. Also the language marks (German, R = .44 and English, R = .24) and marks in informatics (R = .51, n = 97, not all participants took informatics in school) are significant predictor variables. The data of the computer programming self-efficacy scale (PSES) is also a powerful predictor for exam performance (R = .31) and performance in solving exercise sheets (PBasic, R = .19; PPractice, R = .35). 12 % of variation of the practical computer programming performance can be explained by the self-reported programming skills. In line with social-constructivist learning theories where prior knowledge plays an important role, prior education (marks in school and domain-specific competences) seem to be useful factors to include in learning analytics modeling.

The Big Five Factors of personality do not show substantial predictive power for academic success. In contrast to previous research [GW01] a systematic relation between the factor conscientiousness and learning performance could not be found. There is only one exception: a significant negative correlation between the factor neuroticism and the course exam (R = −.18). This means that students who are uncertain, frightened and nervous perform worse in the course exam, but performed as well as the others in the continuous exercises (PBasic and PPractice). The result shows that the educational setting discriminates against students with a high level of neuroticism. Providing students with past exam papers for practicing purposes or writing a trial exam or completely different forms of examination could be measures to support such students.

4.3 Predictive power of LMS data

Given the wealth of LMS student-content interaction data, preliminary analyses were applied to find out which indicators of learning intensity performed well in most of the consecutive weeks. The total number of clicks per week, merging the overall activities, showed the most consistent role in all of the weekly models to predict the impact of student-content interactions (LMS_content) on academic performance. Other variables like *total time online*, *#files viewed, #uploaded,* etc. did not show a consistent role for predictive modeling. The results show little progress in predictive power over time. The earliest predictions have a beta weight in predicting course exam results of 0.23, indicating that only about 5 % of performance variation can be explained by LMS student-content interactions data. After some weeks without significant correlations between overall LMS user activity and performance at the end of the semester coefficients increase to a medium-large effect size (R = .30 – .50) with values up to 0.37 in week 15. This is one week before the exam. This late time close to the exam is not very useful as a prediction model for providing early feedback to students.

A remarkable and consistent feature of prediction is the LMS quizzes dataset. The variable "first score" showed the highest consistency in all of the weekly models and was selected for prediction. Figure 1 demonstrates the predictive power in terms of the multiple correlation coefficients of longitudinal models developed on the Moodle quizzes data for three performance measures. Multiple correlations values demonstrate a medium–large effect size (R = .30 – .50) and develop from around R = 0.30 to R = 0.48 at the end of the semester.

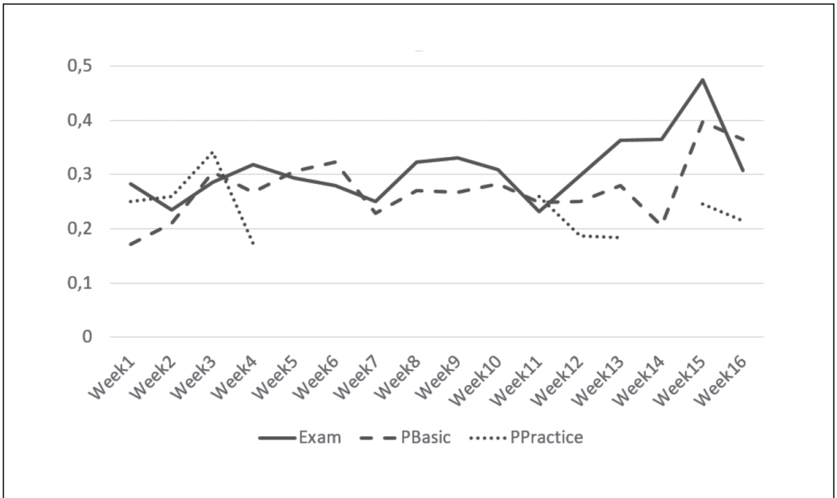

Figure 1: Predictive power of the Moodle quizzes dataset for three performance measures (only significant p < .05 correlations are shown)

In line with previous research [Wo13], the results of quizzes seem to be a reasonable indicator for learning in prediction modeling.

4.4 Predicting performance by exercise sheet results

A good predictor for exam performance are the weekly results of the exercise sheets. Exercise sheets about basic knowledge and concepts in computer science (PBasics) had to be prepared from the first week until week 11, and the sheets for the introduction to programming and code writing (PPractice) had to be prepared in weeks one and two and then every second week until week 12. Students who learn continuously throughout the semester and perform the exercises well, also perform better in the final exam. Therefore, performance measured by exercise sheets constitutes a reliable predictor for the exam. Multiple correlation values range from R = 0.20 to R = 0.59, which explains 35 % of variation.

4.5 Predictive modeling by multiple regression

Although different variable sets appear to show a significant correlation with final exam results, it would be erroneous to rely too heavily on the predictive power of simple correlations. A multiple backwards stepwise regression was conducted in order to develop a predictive model in which the variable "final course exam" was the continuous dependent variable. The final results of the backwards stepwise multiple regression are shown in Table 2.

Table 2: Coefficient data after multiple backwards regression

R2	**Corrected R2**	**Durbin-Watson**
.727	.712	2.04

Model overview after multiple backwards regression

Variable	**Standardized coefficient beta**	**t**	**Sig.**	**VIF**
LMS_Quiz	.313	5.30	.000	1.398
PBasic	.301	4.53	.000	1.765
PPractice	.258	3.74	.000	1.899
Marks German*	.131	2.31	.023	1.277
Marks math*	.119	2.00	.048	1.409
PSES*	.114	2.09	.039	1.186

* Learning Dispositions Dataset

The analysis generated a predictive model of students' exam results as a combination of LMS quiz tracking data, learning dispositions data and formative assessment data. The Durbin-Watson coefficient value indicates that there are no auto-correlation problems in the model. All six variables are statistically significant contributors ($p < 0.05$) and the values of the variance inflation factor (VIF) suggest that multicollinearity effects may be ruled out in this analysis. The multiple squared correlation coefficient for the model is .727, indicating that some 73 % of the variability in students' final performance in this course can be explained by this combination of LMS data and learner data. The single most predictive variable reported here, with a regression coefficient (Beta) of .313 is the LMS tracking variable LMS_Quiz, measuring the total achievement of the first scores of online self-test quizzes. The variable represents student engagement with the learning content, rather than their effort in completing graded course assessments.

The student-content interaction data, measured by the number of total clicks per week, is dominated by the predictive power of the other data components. The LMS student-content interactions data has no added value in predicting performance and was excluded from the predictive model by the backwards stepwise regression analysis. An important finding is that knowledge of actual course design (e. g. what online resources are available, online communication between students and students, and students and teachers) is critical in determining which students are not engaging with course materials in a manner that is indicative of an effective learning strategy. In line with previous findings [MD10; Wo13; Xi15] quizzes, integrated in the course design, seem to be a good indicator of students' engagement in learning and therefore also for predictive modeling. The second and third significantly predictive variables PBasic and PPractice represent student engagement in solving exercise sheets, also used as a formative assessment to stimulate continuous learning throughout the semester with the possibility of gaining nine bonus points for the final exam. The final three significantly predictive variables observed in this study come from the student learning dispositions dataset. They are related to prior knowledge. Students with better marks in German and math and prior knowledge of computer programming, measured by the computer programming self-efficacy scale (PSES), achieve higher overall final exam results. The Big Five Factors of personality do not show substantial predictive power for academic success. The variables were excluded from the model. Learning dispositions data is not as easily collected as tracking data from the LMS [BC12]. Whether the effort involved in collecting learning dispositions data, like prior knowledge in computer programming or high school marks, is worthwhile or not also depends on whether this improves the ability to predict the passing rate.

To test the predictive power of the model in terms of whether or not an individual student is considered “at risk of failure”, binary logistic regressions were calculated. Students with course exam scores < 50 % were coded as “at risk” (0), while students with exam scores ≥ 50 % were coded as “performing adequately or better” (1). In our course scoring scheme < 42 % was considered a fail. The division point was selected to include students whose final exam scores indicate that they barely passed the course and may have benefited from earlier feedback, support and intervention. The logistic regression model (Tab. 3) accurately placed individual students in either the “at risk” or “performing adequately or better” category 91.9 % of the time. The model resulted in errors, classifying an “at risk” student as “performing adequately or better” at a rate of only 5.2 %: only seven students out of 135 were predicted to be performing adequately or better when their actual final exam score placed them in the “at risk” category. Out of these seven students, only three actually failed the course. They did not achieve scores over 42 %. Three out of 135 represents a predictive failure rate of only 2.2 %.

Table 3: Logistic regression “Risk of failure” classification results (N = 135)

Observed	**Predicted**		
	At risk	**Not at risk**	**Percentage correct**
At risk	37	7	84.1
Not at risk	4	87	95.6
Overall percentage			91.9

Note: “At risk” = exam score < 50 %; “Not at risk” = exam score ≥ 50 %

By placing four students in the “at risk” category even though these students eventually passed the course (achieving scores of > 50 %) the model predicted wrongly 3 % of the time. This error seems to be of less concern because it seems better to mistakenly identify a student as being at risk of failure than to neglect a student who requires additional learning support. If these results had been accessible earlier in the semester, the majority of students who failed or almost failed and those who would have been recognized “at risk of failure” could have been identified by teachers.

The value of the learning dispositions data for prior knowledge (expressed as self-reported programming skills and high school marks in German and math) can be demonstrated by a binary logistic regression analysis where only these three variables are included in the model (Tab. 4).

Table 4: “Risk of failure” classification results from three dispositions data variables (N = 141)

Observed	**Predicted**		
	At risk	**Not at risk**	**Percentage correct**
At risk	19	27	41.3
Not at risk	12	83	87.4
Overall percentage			72.3

Overall, the model accurately placed the students 72.3 % of the time (Tab. 5), which shows that even with only these three variables collected in the first week of the semester the model is still comparatively precise. In the model 27 students out of 141 were predicted to be performing adequately, while their final course score placed them in the “At risk” category. This is a failure rate of (only) 20 %. However, as soon as quiz and formative assessment data become available their predictive power is dominant.

5 Conclusions

In this exploratory study we integrated data from several different sources and found evidence for the strong predictive power of data from LMS tracked quiz data and formative assessment data. We also found evidence for the predictive power of some learning dispositions data. Predictions of academic performance using data of domain-specific skills like prior knowledge in computer programming or high school marks in – for the field of knowledge – relevant subjects were much more accurate than when using personality trait data. The Big Five Factors of personality could not be used for predictive purposes. The role of student-content interactions data from the LMS (total number of weekly clicks) also appeared to be minimal and was thus also excluded from predictive modeling. The combination of LMS quiz tracking data, the results of exercise sheets, data of prior education (high school marks), and self-reported prior knowledge led to the predictive model with the greatest significance. Therefore learning analytics should combine LMS data with learner data. As soon as this data is available the generation of timely feedback based on performance predictions and early signaling of underperformance is possible. Differences in prior knowledge and education of the students seem to have a high impact on learning success, which could be quite accurately predicted with our dataset. The high cognitive heterogeneity of the students becomes a key issue for course planning and learning process planning. One idea for dealing with this challenge is to select students with less prior knowledge in computer programming and invite them to a special tutorial so that they can catch up with the others. Future studies should be directed towards the investigation and analysis of potential indicators in relation to other dimensions of heterogeneity such as social heterogeneity, heterogeneous situations in life, or motivational heterogeneity. Even if previous findings show weaker relationships between motivational heterogeneity data and academic performance [Te17], these kinds of learning dispositions could be used in combination with LMS and assessment data to provide better predictions, e. g. in cases where the quality of quizzes and tests are weak (e. g. when they are focusing on factual knowledge only) and do not fit well to the requirements of the final course exam.

References

[Ag14] Agudo-Peregrina, Á. F. et al.: Can we predit 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, S. 542–550, 2014.

[AD09] Askar, P.; Davenport, D.: An investigation of Factors related to self-efficacy for java programming among engineering students. The Turkish online Journal of Educational Technology, 8(1), S. 26–32, 2009.

[BFM12] Bienkowski, M.; Feng, M.; Means, B.: Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology, S. 1–57, 2012.

[BF06] Boud, D.; Falchikov, N.: Aligning assessment with long-term learning. Assessment & Evaluation in Higher Education, 31, 4, S. 399–413, 2006.

[Br13] Brosius, F.: SPSS 21. mitp, Heidelberg, 2013.

[BC12] Buckingham Shum, S.; Crick, R. D.: Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In (Buckingham Shum, S.; Gasevic, D.; Ferguson, R. Eds.): Proceedings of the 2nd international conference on learning analytics and knowledge. New York, S. 92–101, 2012.

[BP12] Bulu, S. T.; Pedersen, S.: Supporting problem-solving performance in a hypermedia learning environment: The role of students' prior knowledge and metacognitive skills. Computers in Human Behavior. 28, 4, S. 1162–1169, 2012.

[CH15] Conde, M. A.; Hernandez-García, Á.: Learning analytics for educational decision making. In: Computers in Human Behavior, 47, S. 1–3, 2015. DOI: http://dx.doi.org/10.1016/j.chb.2014.12.034

[CN12] Credé, M.; Niehorster, S.: 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), S. 133–165, 2012.

[EKS17] Elkins, R. K.; Kassenboehmer, S. C.; Schurer, S.: The stability of personality traits in adolescence and young adulthood. Journal of Economic Psychology. 60, 37–52, 2017.

[GW01] Gibbons, D. E.; Weingart, L. R.: Can I do it? Will I try? Personal efficacy, goals, and performance norms as motivators of individual performance. Journal of Applied Social Psychology. 31, S. 624–648, 2001.

[HB17] Hopwood, C. J.; Bleidorn, W.: Stability and change in personality and personality disorders. Current Opinion in Psychology, In press, accepted manuscript, 2017.

[JSS99] John, O. P.; Svrivastava, J.; Svrivastava, S.: The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In (Pervin, L. A.; John, O. P. Eds.): Handbook of personality: Theory and research, 2, New York, S. 102–138, 1999.

[MD10] MacFadyen, L. P.; Dawson, S.: Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education. 54(2), S. 588–599, 2010.

[MSM12] Metzger, C.; Schulmeister, R.; Martens, T.: Motivation und Lehrorganisation als Elemente von Lernkultur. Zeitschrift für Hochschulentwicklung, 2012. DOI: http://zfhe.at/index.php/zfhe/article/view/433

[RW98] Ramalingam, V.; Wiedenbeck, S.: Development and Validation of Scores on a Computer Programming Self-efficacy Scale and Group Analysis of Novice Programmer Self-efficacy. Journal of Educational Computing Easearch, 9(4), S. 367–381, 1998.

[RJ05] Rammstedt, B.; John, O. P.: Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung des fünf Faktoren der Persönlichkeit. Diagnostica, 51,(4), S. 195–206, 2005.

[RCZ17] Rienties, B.; Cross, S.; Zdrahal, Z.: Implementing a learning analytics intervention and evaluation Framework: What works? In (Daniel, B. K. Ed.): Big data and learning analytics: Current theory and practice in higher education, Cham: Springer International Publishing, S. 147–166, 2017. DOI: http://dx.doi.org/10.1007/978-3-319-06520-5_10

[SDL13] Siemens, G.; Dawson, S.; Lynch, G.: Improving the quality of productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics. Society for Learning Analytics Research, 2013.

[Te17] Tempelaar, D. et al.: Student profiling in a dispositional learning analytics application using formative assessment. Computers in Human Behavior. S. 1–13, 2017.

[WI12] Wagner, E.; Ice, P.: Data Changes Everything: Delivering on the Promise of Learning Analytics in Higher Education. In: EDUCAUSE Review Online. 2012. http://educause.edu/ero/article/data-changes-everything-delivering-promise-learning-analytics-higher-education.

[Wo13] Wolff, A. et al.: Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Paper presented at the proceedings of the third international conference on learning analytics and knowledge, 2013.

[Xi15] Xing, W. et al.: Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. Computers in Human Behavior, 47, S. 168–181, 2015.

[ZZ16] Zhang, J.; Ziegler, M.: How do the big five influence scholastic performance? A big five-narrow traits model or a double mediation model. Learning and Individual Differences. 50, S. 93–102, 2016.