Pass or Fail? Prediction of students' Exam Outcomes from Self-reported Measures and Study Activities

Bianca Clavio Christensen, Brian Bemman, Hendrik Knoche, and Rikke Gade

Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark,

{bcch|bb|hk|rgg}@create.aau.dk

Abstract. Technical educations often exhibit poor student performance and consequently high rates of attrition. Providing students with early feedback on their learning progress can assist them in self-study activities or in their decision-making process regarding a change in educational direction. In this paper, we present a set of instruments designed to identify at-risk undergraduate students in a Problem-based Learning (PBL) university, using an introductory programming course as a case study. Collectively, these instruments form the basis of a proposed learning ecosystem designed to identify struggling students by predicting their final exam grades in this course. We implemented this ecosystem and analyzed how well the obtained data predicted the final exam scores. Best-subset-regression and lasso regressions yielded several significant predictors. Apart from relevant predictors known from the literature on exam scores and drop-out factors such as midterm exam results and student retention factors, data from self-assessment quizzes, peer reviewing activities, and interactive online exercises helped predict exam performance and identified struggling students.

Keywords: Academic performance, Student retention, Learning Management System, Learning Tools Interoperability, Problem-Based Learning, Flipped learning

1 Introduction

Students enrolled in educations with technical content often struggle with passing technical courses and frequently drop out as a result [1, 2]. Much of the research on student dropouts or retention has focused on the personality traits of a student, typically without also considering their actual learning progress, e.g., in [3–6]. Both struggling students and course instructors, however, can benefit from an understanding of how the learning process of students is progressing. Such an understanding, for example, might encourage students to engage more deeply with the learning material or allow instructors to better direct resources to those in need. With many openly accessible learning resources, such as *Massive Open Online Courses (MOOCs)* now available, teachers can construct diverse learning ecosystems for their students which extend far beyond institutionally managed, digital *Learning Management Systems (LMSs)*. The research question addressed here is how we can identify struggling students from information gathered from their diverse interactions with these learning resources. While the midterm exam result is known as a relevant predictor from the literature, we investigate how data from student profile and interactive online activities in a learning ecosystem can help improve prediction of exam performance.

We present a set of instruments designed in a learning ecosystem to identify struggling first-semester undergraduate students enrolled in an introductory programming course at Aalborg University (AAU) - a *Problem-based Learning (PBL)* university in Denmark. These instruments consisted of both student self-reported personal attributes and self-assessed measures of the learning progress. We used these instruments in the construction of a multiple linear regression model for predicting student final exam scores. Our study continued the work presented in [7] and, in particular, expanded upon the original predictive model with a lasso regression analysis while providing an evaluation of the best performing model when no midterm exam score was available. Our proposed model consists of significant predictors from this set of aforementioned instruments that collectively suggest a possible relationship between the academic success of a student and select personal attributes as well as measures of their learning progress. These predictors could provide the university with the means to identify struggling students at risk of leaving the education. This will allow administrations to offer guidance to these individuals as early in their education as possible.

2 Background

Previous research on student retention has identified a number of factors for decreasing the risk of students leaving educational programs: *growth mindset* [3], *grit* (i.e., perseverance when faced with challenges) [4], *study habits* [6], *high school habits* [5, 8], and *social support for studying* [9]. Although this research has documented a wide range of potential predictors of student retention, agreement between studies is low [10–12]. For this reason, continued research would be better served by considering case studies [1]. This could be done, for example, by detecting students at risk of leaving the education and then directing adequate resources to those individuals based on relevant features of the study program from which these students left. One previous study on student dropouts in 2016 [2], looked at first semester students in an undergraduate Media Technology (Media Tech) program at AAU. While their findings from questionnaires, interviews, and study diary logs suggested that reasons for dropping out were quite diverse, they provided some evidence that the required skills and levels in mathematics and programming were higher than students initially expected resulting in dropouts. Natural science courses, for example, are notorious for attrition and low first-time success rates, particularly in the first year of study [13]. It is essential then to investigate interdisciplinary educations, such as Media Tech, that combine technical, scientific, and design skills.

Engaging students in the learning process and holding them accountable for their own progress, especially during the early semesters, is one of the primary goals of university educations and PBL, in particular. One important reason for this is the comparatively less interaction and feedback students receive at a university than in high school. The principle of *pre-training* [14] suggests that providing students with basic information ahead of their actual lectures can reduce cognitive overload. This principle is often implemented online as self-study activities with *flipped learning*. Such approaches leave more time for the instructor to facilitate classroom activities that are essential in a PBL framework, in which students analyze, evaluate, and create content in a hands-on fashion [15]. These learning goals and activities in PBL can be achieved through scaffolding more complex concepts and skills through interaction, group work, peer feedback, and immediate teacher support [16]. This stands in contrast to, for example, the lower levels in Bloom's taxonomy of learning (remember, understand, and apply) [17].

Breaking material down into smaller parts is one way to reduce the cognitive load of students [18] and doing so can make that content more accessible, focused, and easier to digest. Self-assessment questions are one way to increase this sort of in-depth learning in students [19], e.g., when students are trying to understand where they went wrong on a quiz. Through self-assessment quizzes, instructors can efficiently assess and manage student learning by creating, for example, online assignments with automated grading and feedback. This is a fundamental approach advocated in *Learning Tools Interoperability (LTI)* [20]. Moreover, self-assessment quizzes of this type can be designed to adapt and grow in response to student performance based on, for example, previous answers provided to the system. However, creating such content is time-consuming and provides little personal control for the teacher when implemented in an LMS, such as Moodle [20–22]. This ability to adjust feedback to a student's zone of *proximal development* [23] is often considered the gold standard of education that current digital tools and systems, unfortunately, do not meet. Even so, more and more teachers are relying on online activities for instruction. An important feature of this approach is the teacher's ability to not only monitor student progress but also target struggling students for early interventions.

Monitoring student progress online allows for immediate response and communication between teacher and students when adjusting instructions, moderating difficult learning content, and addressing student misunderstandings [24]. Additionally, instructors can get an idea of a student's level of engagement in courses by observing the relationship between that student's performance and their use of Moodle [25]. However, the relationship between grades and behavioral data such as Moodle activity logs is complex. While findings from a correlation analysis of these factors proved inconclusive [26], the authors suggested possible methods for continuing this work through neural networks, statistical classifiers, rule induction, or fuzzy rule learning.

Other retention studies used regression models to predict student performance [27–29]. The use of regression models vary in selection criterion e.g., employing linear regression models in combination with different variable-selection techniques such as adaptive lasso and cross-validated r^2 statistics, found that undergraduate academic performance (as measured by student grade point average (GPA)) can explain 54% of the variance in graduate-level performance [29].

Although previous studies showed changes in results when using different prediction methods [22, 29], a study found that course interventions (e.g. a midterm exam) could affect other assessment predictors in a model and showed how four commonly used measurements (raw change scores, normalized gain scores, normalized change scores, and effect sizes) can lead to misleading conclusions when excluding a control for the course interventions [28]. Instead, the authors proposed building a multiple linear regression using pre- and post-test scores from an introductory biology course that controlled for possible differences in student ability and preparation in order to estimate the effect of any non-randomized instructional intervention on student performance [28]. In summary, it is unclear how data from student profiles and instruments for automating grading and feedback (e.g. interactive online activities) in a learning ecosystem can help improve prediction of exam performance and identify struggling students.

3 Case study context and the learning ecosystem

AAU operates according to a PBL model which assumes that students learn best when applying theory and research-based knowledge to collaborative working strategies aimed at real-world problems. The students at AAU learn to take an active role as problem-solvers in finding and solving real-world situations [30], and AAU primarily teaches PBL project-oriented work. In any one educational program at AAU, each student must enroll in semester study activities corresponding to a total workload of 30 ECTS credits, where a single ECTS is anywhere between 25 to 30 work hours. These 30 ECTS credits typically include a semester project worth 15 ECTS and three courses worth 5 ECTS each. The mandatory study activities at AAU (i.e., semester projects and courses) require students to make connections between them that span from course-to-course in a single semester as well as across semesters.

In the Media Tech program, the introductory programming course required in the first semester constitutes an important building block for a student's academic success in further semesters. The instructor integrated several methods of flipped instruction [16] in the Fall semester 2017. These methods include online self-study activities consisting of self-assessment quizzes (SA), exercises on khanacademy.org (KA), and mandatory hand-in assignments on peergrade.io (PG). A Moodle course page served as the LMS for providing access to these self-study activities in addition to a collection of other learning resources, such as supplementary video content. The course utilized a combination of online instructions and face-to-face lessons, and when combined with these self-study activities, formed a learning ecosystem which encourages student learning beyond the boundaries of the classroom [31–33]. Figure 1 shows an overview of this learning ecosystem in its introductory programming course.

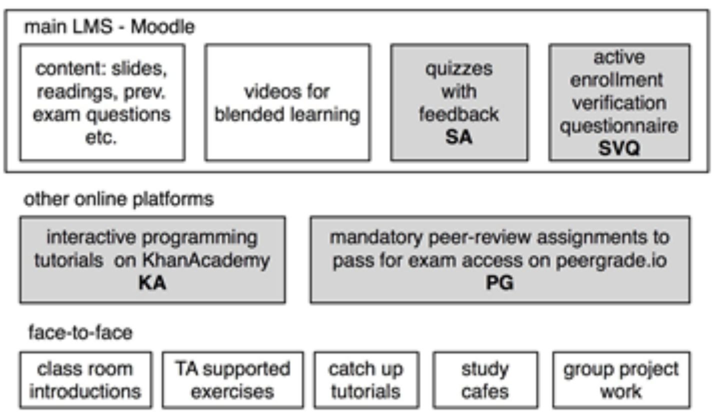

Fig. 1. Overview of the learning ecosystem in the introductory programming course for Media Tech students in the Fall semester 2017.

The learning ecosystem consisted of several online self-study activities, including course readings, videos, SA and KA, which encourage student learning prior to class. During lectures, the teacher presented the topic in a shortened format, followed by hands-on programming exercises in which the students may either work alone or in groups with the help of the teacher, teaching assistants, or their peers as part of study cafes. These in-class opportunities were designed to reinforce the concepts learned in the self-study activities through practical experience. At various points in the semester, students were asked to complete programming assignments and evaluate those created by their peers in PG. Such peer learning provided students with the opportunity to critically apply their knowledge.

Media Tech students have diverse backgrounds (e.g., in nationality, high school specialization, and proficiency in math) and study interests (e.g., in design or programming) [2]. Table 1 describes the background and study interest of the Media Tech undergraduate student from the study on student dropouts in 2016.

Table 1. Media Tech undergraduate student background and study interest from the study on student dropout in 2016 [2].

	Frequency	%	Average
Gender (N=195)			
Male	150	*76.9*	-
Female	45	*23.1*	-
Nationality (N=195)			
Danish	146	*74.9*	-
International	49	*25.1*	-
Education (N=195)			
Non-technical high school	118	*60.5*	-
Technical high school	37	*19.0*	-
Other	40	*20.5*	-
Design and creativity study interest (N=51)			
High importance	36	*70.6*	-
Some importance or less	15	*29.4*	-
Technology study interest (N=51)			
High importance	20	*39.2*	-
Some importance or less	31	*60.8*	-
Grade point average (N=195)	-	-	7.7
Math grade in high school (N=195)	-	-	6.4

In order to gather information about this diversity in first-year students as control variables, we designed a survey called the *Study Verification Questionnaire (SVQ)*. It consisted of a set of 111 self-reported questions based on established factors for student retention discussed in Section 2, such as grit and study habits, among others. Table 2 shows an overview of the SVQ questions and describes their respective SVQ category. We categorized the SVQ questions into 14 categories based on the retention factors discussed in Section 2 and the findings in [2] of Media Tech student dropouts. These 14 categories in SVQ were (a) social support for studying, (b) attitude towards education, (c) reasons for going to University, (d) education choice factors, (e) high school behaviour, (f) high school trust, (g) belonging uncertainty, (h) grit, (i) growth mindset, (j) self-control, (k) personal trait comparison, (l) perceived academic abilities, (m) studying and working hours, and (n) understanding of Media Tech.

Table 2. Summary of the Study Verification Questionnaire (SVQ) containing descriptions of its 14 categories and the number of questions ("Item") found in each.

	Item	Category	Description
a.	1-9	*Social support for studying*	To which degree the students feel encouragement from friends and family [9].
b.	10-15	*Attitude towards education*	How the students feel and think of taking a university education [34].
c.	16-23	*Reasons for going to university*	Which factors motivate the students to commit to a university education [35].
d.	24-33	*Education choice factors*	How the students feel and think about the interdisciplinary tasks of technology and design [2].
e.	34-45	*High school behaviour*	How the students acted socially, and how they engaged in mandatory school activities [5, 8].
f.	46-50	*High school trust*	How the students experienced their high school and teachers [36, 37].
g.	51-56	*Belonging uncertainty*	To which degree the students feel socially integrated at university [38].
h.	57-61	*Grit*	How persistent and passionate the students are in achieving their goal [4].
i.	62-64	*Growth mindset*	To which degree the students believe that they can develop their abilities through effort [3].
j.	65-74	*Self-control*	How well the students regulate behavior, attention, and emotions in service of valued goals [6].
k.	75-87	*Personal trait comparison*	How the students compare their social and abilities to their peers [5, 8].
l.	88-92	*Perceived academic abilities*	How satisfied students are with their academic abilities [5, 8].
m.	93-97	*Studying and working hours*	How much effort and time the students allocate to studying.
n.	98-111	*Understanding of Media Tech*	How well the students understand the important topics related to *Media Tech* [2].

4 Data collection and method

In order to discover possible assessment predictors for predicting the final exam score at the introductory programming course in Fall 2017, we gathered data from all students for each of the self-assessment activities shown in grey in Fig. 1. Additionally, we collected the students' scores from the midterm exam (MT) and the final exam (FE). The predictive analysis explores the use of these assessments, SVQ, SA, KA, PG, and MT for predicting final exam scores. This data analysis first explores the assessment predictors with the midterm exam score and then without it, as the midterm exam is a costly resource and may be discontinued at the Media Tech education. We will analyze the assessment variables using two predictive methods: best-subset-regression using maximum five predictors and sequential replacement (Section 5.1) and lasso regression with best subset selection to choose the best-fitting features in a multiple linear regression model (Section 5.2).

4.1 Data set description

We collected data of students who enrolled in Media Tech in 2017 and who took the exam in the programming course on first semester. The data set consists of 22 assessment variables collected for 72 students. While 30 students failed the midterm exam, they were still eligible to participate in the final exam if they submitted the mandatory Peergrade assignments. At the final exam 58 students failed. Table 3 groups the assessment variables into: student enrolment, course activities, and course performance. As a part of the learning ecosystem, we collected data on the interactions the students had with these learning resources, e.g. the number of attempts that a student took for completing Khanacademy exercises.

Table 3. Grouping, description, and number of assessment variables.

Assessment	Description	Var
Student profile		**14**
SVQ	Study Verification Questionnaire	
Course activities		**6**
SA-avg	Average score in the self-assessment quizzes	
SA-N	Completed self-assessment quizzes	
KA-N	Attempts in Khanacademy exercises	
PG-submit	Peergrade submission score	
PG-feedback	Peergrade feedback score	
PG-combi	Peergrade combined score	
Course performance		**2**
MT-S	Midterm exam score	
FE-S	Final exam score	
Total		**22**

For the SVQ categories, we graded the student answers according to risk factors for dropping out from the literature as shown in Table 2. We graded the category of *understanding of Media Tech* based on our teacher experience of what constitute general attitudes or behaviours of a good student at Media Tech, e.g. when a student enjoys designing and building things towards solving real-world problems.

5 Predictive analysis results

In order to predict the final exam scores with and without the *midterm exam score*, we first applied a best-subsets-regression using maximum five predictors and sequential replacement, i.e. a combination of forward and backward selections. Then we applied a lasso regression. Both analyses are reported in the following sections using the 22 assessment predictors shown in Table 3 collected from the 72 students taking the introductory programming course in the fall of 2017.

5.1 Best-subset-regression

We began by using the students' *MT scores* to construct base linear regression model (*BSR1*). When excluding the midterm exam score from the best-subsets-regression analysis, the subset selected *Peergrade submission score* (*BSR6*). Table 4 shows the results of the different regression models, when continuing adding or changing other assessment predictors to *BSR1* and *BSR6*.

Table 4. Adjusted R-squared, predicted R-squared and p-values of the best-subset-regression models, predicting student performance in the final exam. The best performing model is highlighted in grey with MT and without MT.

Model	Var	Predictors	*Adj, r^2*	*Pred. r^2*	P	
BSR1	1	*MT-S*	0.50	0.48	0.001	*
BSR2	2	*MT-S, KA-N*	0.53	0.51	0.001	*
BSR3	3	*MT-S, KA-N, high school trust*	0.56	0.53	0.001	*
BSR4	4	*MT-S, KA-N, high school trust, self-control*	0.58	0.54	0.001	*
BSR5	5	*MT-S, KA-N, high school trust, self-control, personal trait*	**0.62**	0.58	0.001	*
BSR6	1	*PG-submit*	0.16	0.13	0.001	*
BSR7	2	*PG-submit, KA-N*	0.20	0.16	0.001	*
BSR8	3	*PG-submit, KA-N, high school trust,*	0.24	0.19	0.001	*
BSR9	4	*PG-submit, KA-N, high school trust, self-control*	0.28	0.20	0.001	*
BSR10	5	*PG-submit, KA-N, high school trust, self-control, personal trait*	**0.30**	0.24	0.001	*

**Less than 0.001*

The best model from this selection method, *BSR5,* consisted of the following assessment predictors: midterm score, attempts in Khanacademy exercises, self-reported high school trust, self-reported self-control, and self-reported personal traits (*r^2 = 0.62, pred. r^2 = 0.58, $p < 0.001$*). When no midterm exam score was available, the best model from selection method, *BSR10,* consisted of the following assessment predictors: *Peergrade submission score, attempts in Khanacademy exercises, self-reported high school trust, self-reported self-control,* and *self-reported personal traits* (r^2 = 0.30, pred.r^2 = 0.24, $p < 0.001$). For comparison, an ANOVA test revealed significant differences from *BSR1* and *BSR5* (*F* = 6.49, *p* < 0.001) and from *BSR6* and *BSR10* (*F = 4.69, p < 0.003*).

Adding new predictors seems to improve the base models (*BSR1* and *BSR6*) in terms of the increasingly adjusted R-squared (i.e. how a new term improves the model more than would be expected by chance) and predicted R-squared values (i.e. how well the model predicts the removed observation). However, as we risk modelling random noise in the data when including many predictors, we began testing for overfitting in the best performing model with MT (*BSR5*) and the best performing model without MT (*BSR10*). To examine the models further, we used a 5-fold cross validation to estimate the accuracy and test for overfitting of the models. Table 5 presents the final cross validation results based on the five folds.

Table 5. Cross validation results of the best performing models from best-subset-regression (*BSR5* and *BSR10*).

Model	RMSE	R-squared	MAE	RMSESD	R-squaredSD	MAESD
BSR5	12.11	0.60	10.00	2.93	0.20	2.84
BSR10	15.85	0.33	12.86	2.98	0.34	2.60

The cross-validation accuracy results for *BSR5* (*r^2 = 0.60*) differ from the whole sample in Table 4 (*r^2 = 0.62*), accounting for 62% of the variance in exam scores for these participants. The resulting validation-set error in the form of the Relative Mean Absolute Error was ten percentage points (*RMSE = 10.00*). The accuracy results for *BSR10* differ from the cross validation (*r^2 =*

0.33) and the whole sample in Table 4 ($r^2 = 0.30$), and the validation-set error was almost 13 percentage points in this model (*RMSE = 12.86*).

5.2 Lasso regression

In our second analysis, we applied lasso regression for predicting the final exam score due to its robustness in a dataset with many features collected from a small sample size [39]. This analysis selected a different subset of assessment predictors compared to the previous predictive analysis that used best-subset-regression.

Lasso regression and best subset selection were used to choose the best assessment predictors in a multiple linear model. In lasso regression, the choice of the tuning parameter was based on a cross validation argument choosing the highest tuning parameter, which yielded a cross validation error within one standard error of the lowest cross validation error [39]. When repeating using all assessment predictors, the subset selected only the *midterm exam score* (*LASSO1*). When excluding the midterm exam score, the subset selected *Peergrade submission score* (*LASSO6*). Using best subset selection in an exhaustive search with a maximum subset size of five yielded AIC and BIC results shown in Table 6.

Table 6. Chosen subsets using best subset selection with exhaustive search with a maximum of size five for modelling final exam scores. The best performing model is highlighted in grey, called *LASSO5*.

Model	Var	Predictors	*df*	AIC	BIC
null	0		2.00	21.86	25.83
LASSO1	1	*MT-S*	4.00	-18.46	-10.50
LASSO2	2	*MT-S, PG-combi*	4.00	-17.32	-9.36
LASSO3	3	*MT-S, PG-combi, high school trust*	5.00	-20.52	-10.57
LASSO4	4	*MT-S, PG-combi, high school trust, self-control*	6.00	-21.67	-9.74
LASSO5	5	*MT-S, PG-combi, high school trust, personal trait, self-control*	7.00	-23.17	-9.25
LASSO6	1	*PG-submit*	3.00	13.17	19.14
LASSO7	2	*SA-avg, PG-submit*	4.00	12.04	20.00
LASSO8	3	*SA-avg, PG-submit, self-control*	5.00	10.30	20.25
LASSO9	4	*SA-avg, PG-submit, high school trust, self-control*	6.00	8.54	20.47
LASSO10	5	*SA-avg, PG-submit, high school trust, perceived academic abilities, self-control*	7.00	8.42	22.34

The model with the lowest AIC and BIC (*LASSO5*) is highlighted in grey in Table 6 and consisted of the following assessment predictors: *midterm exam score, Peergrade combined score, self-reported high school trust, self-reported personal traits,* and *self-reported self-control*. For comparison, an ANOVA test revealed significant differences from our base model (*LASSO1*) and *LASSO5* ($F = 4.74, p <$

0.01) and from *LASSO6* and *LASSO10* (*F = 3.20, p = 0.02*). The models in Table 6 were also compared with 10-fold cross validation with RMSE as a measurement of the prediction error. Figure 2 shows the cross-validation results of these models.

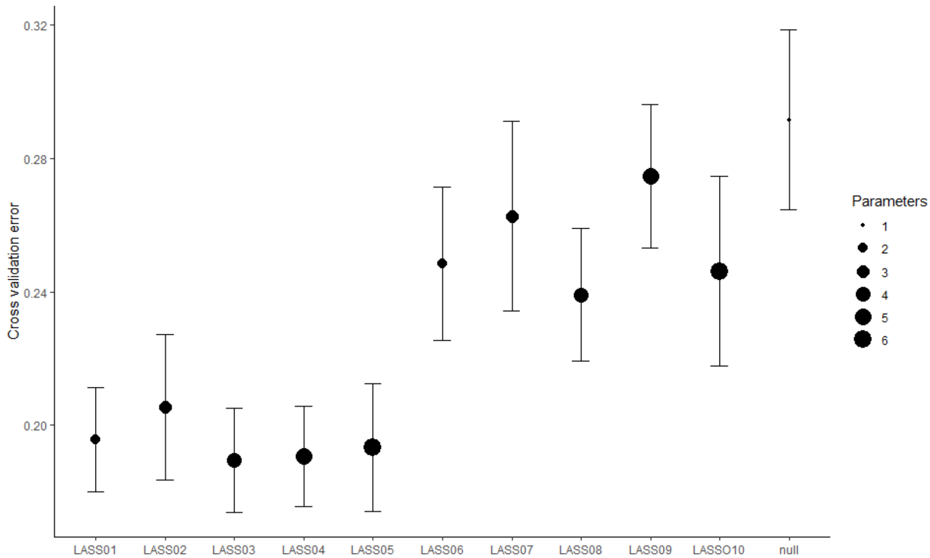

Fig. 2. Comparison of RMSE for linear models predicting the final exam score. The bars indicate plus/minus one standard error and the size of the dot indicates the number of parameters.

The five models on the left in Fig. 2 have the lowest cross-validation error. These models include the midterm exam score from *LASSO1* to *LASSO5*. The lowest cross validation errors were from models (*LASSO3, LASSO4,* and *LASSO5*) with the best subset of three, four, and five parameters when the midterm exam score was available.

5.3 Correlation between the predictors

We analyzed the correlations between the assessment predictors to measure the strength and direction of the linear relationship. For the correlation analysis, we selected the assessment predictors from the best performing models (i.e. *BSR5, BSR10, LASSO5, LASSO10*) for predicting the final exam score. Figure 3 shows a graphical display of the correlation matrix results using Spearman's correlation method for non-parametric rank-based correlation test. The variables in the correlation analysis include (listed from left to right in Fig. 3): self-reports of *high school trust, self-control (hsTrust), academic abilities (acadAbility), personal traits (personalTrait),* and *average score in self-assessment quizzes (SAAVG), final exam score (examScores), midterm exam score (MT), attempted Khanacademy exercises (KAN), Peergrade combined score (PGC),* and *Peergrade submission score (PGS)*.

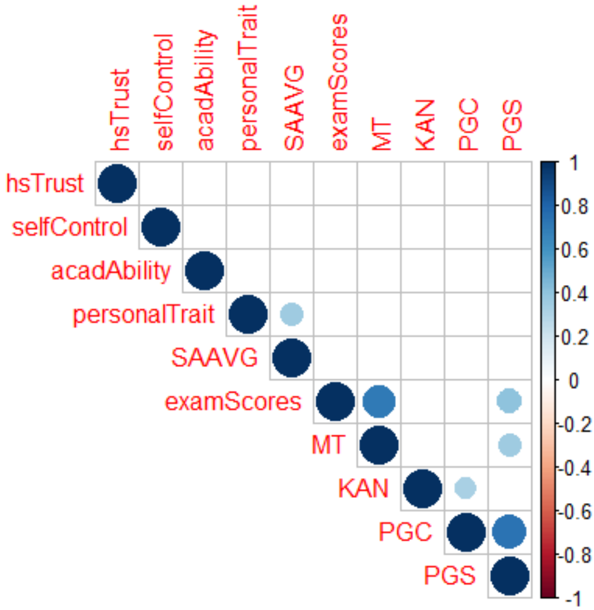

Fig. 3. Correlation matrix of the assessment predictors selected in the best performing models for predicting the final exam score, using Spearman's correlation method for non-parametric rank-based correlation test (from left to right): *self-reported high school trust, self-reported self-control, self-reported academic abilities, self-reported personal traits, average score in self-assessment quizzes, final exam score, midterm exam score, attempted Khanacademy exercises, Peergrade combined score,* and *Peergrade submission score*. The insignificant correlations ($p < 0.01$) are left blank.

The most correlated variables in a data table are highlighted with a blue or red color, signifying the direction of the relationship with the correlation coefficients labelled on the right. The insignificant correlations ($p < 0.01$) are left blank in this figure. Spearman's correlation found six significant correlations that all have a positive relationship (i.e. as the value of one variable increases, so does the value of the other variable). Figure 4 shows the relationship between the seven variables that are significantly correlated (from left to right): *self-reported personal traits, average score in self-assessment quizzes, final exam score, midterm exam score, attempted Khanacademy exercises, Peergrade combined score,* and *Peergrade submission score*.

The distribution of each variable is shown on the diagonal in Fig. 4. On the bottom of the diagonal, the bivariate scatterplots show a fitted line that indicates a monotonic or non-monotonic relationship. The top of the diagonal shows the value of the correlation and the significance level as stars. As expected, the two highest correlations in this matrix are naturally related: The *midterm exam score* contributes to the *final exam score*; and the *Peergrade combined score* is based on *Peergrade submission score* and *Peergrade feedback score*, where the latter was not a significant predictor and thus deselected from the correlation analysis. Furthermore, the *Peergrade* *submission score* is also significantly correlated to the *midterm exam score* and the *final exam score*.

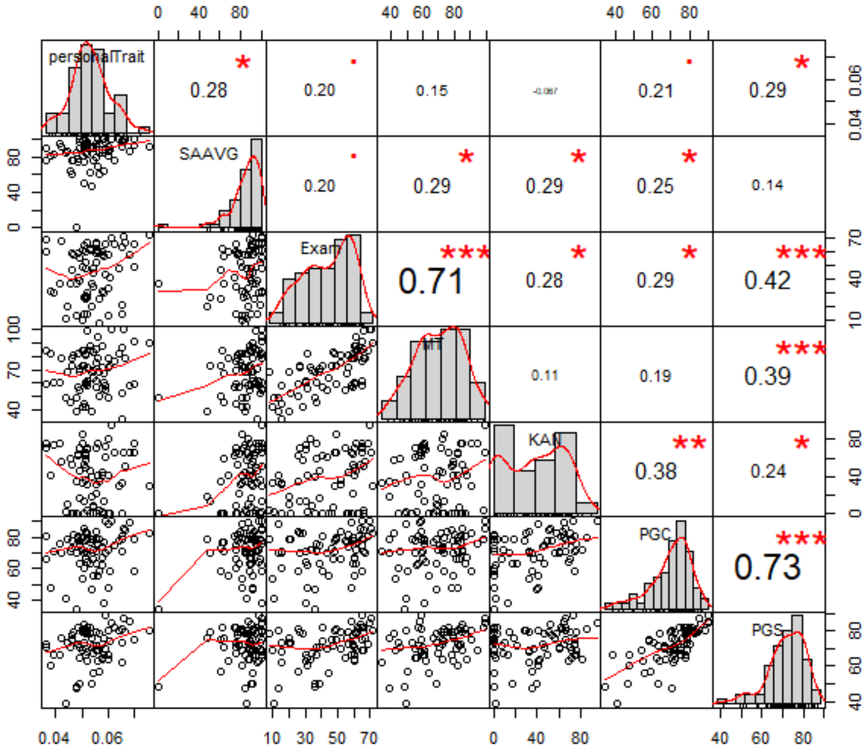

Fig. 4. Correlation matrix of the significant correlations in Fig. 3 (from left to right): *self-reported personal traits, average score in self-assessment quizzes, final exam score, midterm exam score, attempted Khanacademy exercises, Peergrade combined score*, and *Peergrade submission score*. The distribution of each variable is shown on the diagonal. On the bottom of the diagonal, the bivariate scatterplots show a fitted line that indicates a monotonic or non-monotonic relationship. The top of the diagonal shows the value of the correlation and the significance level as stars (*p-values: *** less than 0, ** less than 0.001, * less than 0.01*).

6 Discussion

Our first analysis with best-subset-regression and our second analysis with lasso both identified the midterm exam scores as a strong predictor of the final exam score. This echoed findings by Meier et al. [27] in which in-class exams were better predictors of overall course performance than homework assignments, for a recent literature review see [21]. Including a subset of assessment predictors significantly improved the models in both analyses (*Best-subset-regression:* $F = 6.49, p < 0.001$*; Lasso:* $F = 4.74, p < 0.01$). The subsets included retention factors from the SVQ of which self-reported *high school trust, personal traits,* and *self-control* were significant predictors of final exam scores across the analyses and predictor selection methods. In addition, the methods selected *attempts in Khanacademy exercises* and *Peergrade combined score*. Thus, our results indicate that while student performance on the *midterm exam* was positively correlated with performance on the *final exam*, having a learning ecosystem, which consisted of several appropriate and diverse assessments, as demonstrated by model *BSR5* and *LASSO5*, significantly improved the prediction of final exam scores.

The *midterm exam score* as the strong predictor does not necessarily mean that none of the other predictors are important for student performance. It is likely that if they were good predictors of the final exam score, they would also be good predictors of the midterm exam score. An explanation of high performance of the *midterm exam score* can be that it was designed to be part of the students' final grade, not the *final exam score*. Therefore, including the *midterm exam score* in the model might obscure the importance of other predictors. Repeating the analysis without the midterm exam found the *Peergrade submission score* as the best single predictor, both in the best-subset-regression and lasso. Although this predictor is positively correlated with the *final exam score*, it has a somewhat non-monotonic relationship (or non-linear correlation), having the orthogonal shape on Fig. 4., which can question the reliability of this result.

The predictive models when no midterm exam score available were also improved significantly from the base models (*Best-subset-regression:* $F = 4.69$, $p < 0.003$; *Lasso:* $F = 3.20$, $p = 0.02$). However, these model performances never reach the same performance level as in *BSR5* in terms of R-squared values or in *LASSO5* in terms of AIC and BIC. Another limitation of the study is that we risk modelling noise in the data when including more predictors to the base model as seen in the cross-validation error on Fig. 2. Similar to Casey's findings [26], the relationship between course performance and the assessment predictors is complex and might be influenced by additional factors. This may include how the students performed and were motivated in other courses about which we had no data. Additionally, we might get varied results due to the diversity in the population sample, as found in [2]. These varied results can be due to the lack of control for student non-equivalence [28], e.g., we could have grouped the assessment predictors in scores before and after the midterm exam. Including further behavioural data e.g. from Moodle activity logs (c.f. [26] e.g. on and off campus use of the LMS) might improve predicting course performance further.

While our results suggest otherwise, the SVQ could potentially include retention factors, locate struggling students early in the education, and extend our understanding of the Media Tech students from [2]. While we have no clear explanation of why known retention factors had little influence on the final exam score, it can be related to the factors not working well with either the Media Tech students, the first semester of study, and/or predicting the course performance.

While the analyses found different assessment predictors (e.g. how often students attempted self-guided course assignments), we deem a model without the midterm exam highly relevant, due to the associated cost of preparing and running such exams compared to LTI instruments, such as teacher facilitated interactive online activities that automate grades and feedback. Therefore, the model from the second analysis (*LASSO10*) might be most relevant for predicting the course performance, as data from self-assessment quizzes, peer reviewing activities, and interactive online exercises helped predict exam performance and identified struggling students.

7 Conclusion

The instruments of the learning ecosystem presented in this paper provide initial findings in support of additional strategies for targeting struggling students in a PBL environment. While the results leave much room for improvement, they nonetheless demonstrate that regular student feedback through self-regulated knowledge assessments and proper evaluations of student behavior and psychology may be essential factors in reducing rates of PBL student failure. Moreover, technological learning tools through, e.g., Moodle, Peergrade or Khan Academy, might serve as useful tools for ensuring academic success of students. Such benefits are particularly needed at universities where more and more degree programs are becoming interdisciplinary and courses are being taught by different instructors at separate campus locations. Guaranteeing the quality of education in these situations is essential.

8 Future work

In the future, we hope to implement the significant assessments of our learning ecosystem into a system for identifying struggling students prior to the midterm exam of a given course and incorporate additional sources of relevant information such as Moodle course activity. The variety of significant assessment predictors in our models emphasize the need for a learning ecosystem that is both targeted and wide-ranging, as shown, for example, in Fig. 1. With this model, it might be possible in future semesters to target struggling students even before the start of a course by identifying those who reported low self-reported scores on high school trust or self-control. In compliance with LTI, an automated weekly analysis could invite students in need to group tutoring sessions based on how often students either attempted Khan Academy exercises or completed self-assessment quizzes.

Acknowledgments. We would like to thank the anonymous reviewers for their very thoughtful and constructive feedback that helped shaping the final version of this paper. This work was partially funded by PBL dev 2.0 project at Aalborg University and the EU Erasmus+ DigiCulture project grant no. 2018-1-RO01-KA204-049368.

References

1. Dekker G.W., Pechenizkiy M., Vleeshouwers J.M.: Predicting students drop out : a case study Proc. 2nd Int. Conf. Educ. Data Min. EDM 2009 July 1-3 2009 Cordoba Spain, pp. 41–50 (2009)
2. Bøgelund P., Justesen K., Kolmos A., Bylov S.M.: Undersøgelse af frafald og fastholdelse ved medialogi og andre uddannelser ved det Teknisk-Naturvidenskabelige Fakultet 2015-2016: Arbejdsrapport Nr. 1, Aalborg Universitet, (2016)
3. Dweck C.S.: Self-theories: Their Role in Motivation, Personality, and Development, Psychology Press, (2000)

4. Duckworth A.L., Quinn P.D.: Development and Validation of the Short Grit Scale (Grit–S) J. Pers. Assess., 91, pp. 166–174 (2009)
5. Glynn J., Sauer P., E. Miller T.: A Logistic Regression Model For The Enhancement Of Student Retention: The Identification Of At-Risk Freshmen Int. Bus. Econ. Res. J. IBER, 1, pp. 79–86 (2011)
6. Tsukayama E., Duckworth A.L., Kim B.: Domain-specific impulsivity in school-age children Dev. Sci., 16, pp. 879–893 (2013)
7. Christensen B.C., Bemman B., Knoche H., Gade R.: Identifying Students Struggling in Courses by Analyzing Exam Grades, Self-reported Measures and Study Activities in Knoche, H., Popescu, E., and Cartelli, A. (eds.) The Interplay of Data, Technology, Place and People for Smart Learning. pp. 167–176. Springer International Publishing (2019)
8. Higher Education Research Institute: CIRP Freshman Survey – HERI, https://heri.ucla.edu/cirp-freshman-survey/, (2017)
9. Cabrera A.F., Nora A., Castaneda M.B.: College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention J. High. Educ., 64, pp. 123–139 (1993)
10. Herzog S.: Measuring Determinants of Student Return VS. Dropout/Stopout VS. Transfer: A First-to-Second Year Analysis of New Freshmen Res. High. Educ., 46, pp. 883–928 (2005)
11. Lassibille G., Gómez L.N.: Why do higher education students drop out? Evidence from Spain Educ. Econ., 16, pp. 89–105 (2008)
12. Touron J.: The determination of factors related to academic achievement in the university: Implications for the selection and counselling of students High. Educ., 12, pp. 399–410 (1983)
13. Seymour E., Hewitt N.M.: Talking about leaving: why undergraduates leave the sciences, Westview Press, Boulder (Colo.), (1997)
14. Mayer R.E.: Principles for Managing Essential Processing in Multimedia Learning: Segmenting, Pretraining, and Modality Principles The Cambridge handbook of multimedia learning. pp. 169–182. Cambridge University Press, New York, NY, US (2005)
15. Guerrero W.: Flipped Classroom and Problem-based Learning in Higher Education. Latin-American Context pp. 5
16. Green L.S., Banas J.R., Perkins R.A. eds: The Flipped College Classroom: Conceptualized and Re-Conceptualized, Springer International Publishing, (2017)
17. Bloom B.S., Krathwohl D.R., Masia B.B.: Taxonomy of educational objectives: the classification of educational goals, Longman, (1984)
18. Banas J.R., Velez-Solic A., Banas J.R., Velez-Solic A.: Designing Effective Online Instructor Training and Professional Development, https://www-igi-global-com.zorac.aub.aau.dk/gateway/chapter/105275?ct=-8586434867164189976
19. Evans C., Palacios L.: Interactive Self-Assessment Questions within a Virtual Environment Int. J. E-Adopt. IJEA, 3, pp. 1–10 (2011)

20. Queiros R., Leal J.: Using the Learning Tools Interoperability Framework for LMS Integration in Service Oriented Architectures Presented at the January 1 (2011)
21. Hellas A., Ihantola P., Petersen A., Ajanovski V.V., Gutica M., Hynninen T., Knutas A., Leinonen J., Messom C., Liao S.N.: Predicting Academic Performance: A Systematic Literature Review Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. pp. 175–199. ACM, New York, NY, USA (2018)
22. Severance C., Hardin J., Whyte A.: The coming functionality mash-up in Personal Learning Environments Interact. Learn. Environ., 16, pp. 47–62 (2008)
23. Vygotsky L.: Zone of proximal development. Mind in society: The development of higher psychological processes, 52-91 (1987)
24. Bergmann J., Sams A.: Flipping for Mastery Educ. Leadersh., 71, pp. 24–29 (2014)
25. Casey K., Gibson J.P.: (m)Oodles of data : mining Moodle to understand student behaviour ICEP 2010 : International Conference on Engaging Pedagogy. pp. 61–71. , Maynooth, Ireland (2011)
26. Touya K., Fakir M.: Mining students' learning behavior in moodle system 7, pp. 12–26 (2014)
27. Meier Y., Xu J., Atan O., Schaar M. van der: Predicting Grades IEEE Trans. Signal Process., 64, pp. 959–972 (2016)
28. Theobald R., Freeman S.: Is It the Intervention or the Students? Using Linear Regression to Control for Student Characteristics in Undergraduate STEM Education Research CBE Life Sci. Educ., 13, pp. 41–48 (2014)
29. Zimmermann J., Brodersen K.H., Heinimann H.R., Buhmann J.M.: A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance J. Educ. Data Min., 7, pp. 151–176 (2015)
30. Barrows H.S.: A taxonomy of problem-based learning methods Med. Educ., 20, pp. 481–486 (1986)
31. Chang M., Li Y. eds: Smart Learning Environments, Springer-Verlag, Berlin Heidelberg, (2015)
32. Giovannella C.: Smart Learning Eco-systems: "fashion" or "beef" ? J. E-Learn. Knowl. Soc., 10, pp. 13–21 (2014)
33. Giovannella C., Rehm M.: A critical approach to ICT to support participatory development of people centered smart learning ecosystems and territories Aarhus Ser. Hum. Centered Comput., 1, pp. 2 (2015)
34. Retention Management System Plus Samples | Survey and Report Samples | Ruffalo Noel Levitz, https://www.ruffalonl.com/complete-enrollment-management/student-success/rnl-retention-management-system-plus/samples
35. Yeager D.S., Henderson M.D., D'Mello S., Paunesku D., Walton G.M., Spitzer B.J., Duckworth A.L.: Boring but Important: A Self-Transcendent Purpose for Learning Fosters Academic Self-Regulation J. Pers. Soc. Psychol., 107, pp. 559–580 (2014)

36. Yeager D.S., Purdie-Vaughns V., Garcia J., Apfel N., Brzustoski P., Master A., Hessert W.T., Williams M.E., Cohen G.L.: Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. J. Exp. Psychol. Gen., 143, pp. 804–824 (2014)
37. Cohen G.L., Garcia J., Purdie-Vaughns V., Apfel N., Brzustoski P.: Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap Science, 324, pp. 400–403 (2009)
38. Walton G.M., Cohen G.L.: A question of belonging: race, social fit, and achievement J. Pers. Soc. Psychol., 92, pp. 82–96 (2007)
39. Lu Z.Q.J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction J. R. Stat. Soc. Ser. A Stat. Soc., 173, pp. 693–694 (2010)