Predicting admission test success using SPOC interactions

Pedro Manuel Moreno-Marcos
Universidad Carlos III de Madrid
pemoreno@it.uc3m.es

Tinne De Laet
Katholieke Universiteit Leuven
Tinne.DeLaet@kuleuven.be

Pedro J. Muñoz-Merino
Universidad Carlos III de Madrid
pedmume@it.uc3m.es

Carolien Van Soom
Katholieke Universiteit Leuven
carolien.vansoom@kuleuven.be

Tom Broos
Katholieke Universiteit Leuven
Tom.Broos@kuleuven.be

Katrien Verbert
Katholieke Universiteit Leuven
katrien.verbert@cs.kuleuven.be

Carlos Delgado Kloos
Universidad Carlos III de Madrid
cdk@it.uc3m.es

ABSTRACT: In order to start Medical or Dentistry studies in Flemish universities, prospective students have to pass a central admission test to guarantee they have the proper level of proficiency. To support those learners, a blended program with a SPOC (Small Private Online Course) was designed on Edge edX. The logs from the platform provide a great opportunity to delve into the behavior of learners and to try to predict their success in the test based on students' interactions with the SPOC. This article has the following objectives: (1) analyze the differences of user interactions between learners based on their background, (2) develop and analyze predictive models to forecast who will pass the admission test, (3) discover which variables have more effect on success in this test, and (4) discuss about the generalizability of the solution. The results show that the SPOC learning behavior differs significantly between students with different background; it is not possible to predict success the admission test until the last months; and the average grade using only first attempts stands out as the best predictor.

Keywords: SPOCs, prediction, learners' success, learning analytics, indicators, generalizability

1 INTRODUCTION

In most countries, entry into medical schools is restricted by a high-stake admission test. In Flanders, this test consists of a scientific part with questions on chemistry, physics, mathematics and biology, and an information processing part. The passing rate fluctuates around 20%. The most influential success predictor is the prior educational track, giving students with a science and/or mathematics background (hereafter called “traditional students”) an advantage (Roggemans & Spruyt, 2014). As a result, students train intensively for the admission test to be optimally prepared.

In the digital era, technologies have enabled new ways to provide learning that can support those future students. With the popularity of online learning (and particularly with MOOCs, Massive Open Online Courses) because of its flexibility (Orlando & Howard, 2018), new kind of courses have appeared that use new learning facilities such as quizzes and video interactions. SPOCs (Small Private Online Courses) (Fox, 2013) have emerged as a way to use MOOC technology for specific on-campus training (e.g., for students enrolled in a course). Moreover, note that all these digital platforms not only serve as a repository to upload teaching materials, but they can also get comprehensive traces about learners' interactions, which can be very useful to detect patterns about students' behaviors and to predict trends on advance (e.g., who will pass the course) (Moreno-Marcos, Muñoz-Merino, Alario-Hoyos, Estévez-Ayres, & Delgado Kloos, 2018)

Prediction in education has a special relevance because stakeholders (e.g., teachers and students) can anticipate what will happen in the course, so they can adapt their teaching/learning behavior to improve. Furthermore, predictions can be presented through dashboards to aid sensemaking (Ali, Hatala, Gašević, & Jovanović, 2012), e.g., presenting information about students' success or students at risk (Park & Jo, 2015) to make students self-reflect on their learning. At this point, stakeholder engagement is very important and course builders and instructors should be involved in the design of visualizations, predictions, etc. (without neglecting students). However, although many people are involved, and accurate and meaningful predictions are obtained, a prominent issue is how to make the results generalizable because the course context can considerably affect the results.

Particularly, the course context and course design have special relevance in online or blended courses where learners are more at risk to procrastinate and need good self-regulation skills for success (You, 2016). That is the case for the SPOC KU Leuven developed to support last year high-school students to prepare for the chemistry component of the admission test. In that course, any student can enroll to access videos, theoretical background and exercises to prepare for the admission test. In particular for students from non-traditional study programs, the SPOC format would allow them to study at their own pace. However, it is not clear how the learning behavior in a SPOC to prepare for a high-stake admission test can influence success of the student and how results of the SPOC can be generalizable. In this context, this work aims to address the following objectives:

- Analyze the difference on grades and platform behavior between learners depending on their secondary school background
- Analyze the moment in which we can anticipate accurately if students will pass the admission test

- Identify the variables that have more influence on the predictive models to forecast success in the admission test
- Discuss about how to achieve the generalizability of the results presented in the previous objectives

2 RELATED WORK

In literature, there is an increasing interest in developing predictive models in education. Some of the most typical cases are related to forecasting dropout (e.g., Aguiar, Chawla, Brockman, Ambose, & Goodrich, 2014) and student success (e.g., Ashenafi, Riccardi, & Ronchetti, 2015). Particularly in MOOCs, which have similar format to SPOCs although their contexts and characteristics of learners are different, Moreno-Marcos, Alario-Hoyos, Muñoz-Merino, and Delgado Kloos (2018) carried out a literature review on prediction. They found that dropout is the most-used outcome variable (e.g., Jian & Li, 2018), followed by final or assignment scores (e.g., Brinton & Chiang, 2018) and certificate earners (e.g., Ruipérez-Valiente, Muñoz-Merino, and Delgado Kloos, 2018). They also stated that that there are many possible prediction features (although those related to platform use stand out) and indicated that new ones could be introduced (e.g., self-regulated learning variables, as used by Maldonado-Mahauad et al., 2018, to forecast success).

Among the most prominent variables to predict are test scores. For example, Okubo, Yamashita, Shimada, and Ogata (2017) used a Recurrent Neural Network (RNN) to predict the grade (between A-F) in a university course and compared the predictive power in the 15 weeks of the course. Fewer contributions focus on SPOCs. Yu (2018) used combined linear regression and deep neural network (DNN) to predict the final score of a computer science course. Moreover, Ruipérez-Valiente et al. (2018) predicted learning gains in a 0-course for freshmen students. This article presents a similar kind of study, although the logs and context (e.g. course duration and objective, pedagogy, etc.) are different. Finally, regarding state exams, Feng, Heffernan, and Koedinger (2006) developed a regression model to forecast grades in the exam based on interactions with an Intelligent Tutoring System (ITS). More recently, Fancsali et al. (2018) also predicted a math state exam from logs of their ITS (MATHia), such as solving time, knowledge components (KC) mastered, etc.

This paper presents a study that analyzes how admission test success can be predicted from the learning behavior in a SPOC and which variables affect the prediction. That contributes to the analysis of learning behavior in SPOCs and how it relates to student success. One of the differences with previous research is the identification of learning factors that are important in relation to the educational background (i.e. between learners whose background is appropriate or not for a certain bachelor) and success. Moreover, we innovate with new variables (e.g., variables related to the run of consecutive actions, pauses in videos, whether a student asks for the answer). In addition, the context is different (e.g., sequence of activities, pedagogy, etc.), there are reflections about the best moment to predict and models are not developed only at the end. Finally, we also include reflections about the generalizability of the solution, which are often neglected in articles about prediction.

3 METHODOLOGY

3.1 Case study and data collection

The study was carried out in a SPOC about chemistry, which was developed in Edge edX as a joint project of the Faculty of Science and the Faculty of Medicine at KU Leuven. The SPOC consists of 11 modules including 66 videos and 121 exercises, which cover the required contents for the chemistry component of the medicine admission test in Flanders. This entrance exam contains several tests, although this SPOC was focused only on chemistry. The SPOC was part of a blended learning support program: online modules were released gradually every fortnight (from September to May) and alternated with three face-to-face interactive sessions that used a flipped classroom approach, with the intention to stimulate SPOC learners to spread their learning activities over the year. Nevertheless, in practice many students enrolled late and they studied at their own pace. The target users were students in the last year of secondary school (in the academic year 2016-2017) who wanted to enter Medicine in any university in Flanders and paid a registration fee for the blended learning program. A total of 1,062 students accessed the course, although only 680 completed at least one exercise, and only 750 had interactions with videos.

For the analysis of data, two main sources were used. The first one includes the tracking logs from Edge edX (edX, 2018). Particularly, the following events have been considered: (1) *problem_check*, (2) *problem_show*, (3) *play_video*, (4) *pause_video*, (5) *seek_video* and (6) *stop_video*. The second source consists of the information about the self-reported results of 133 students of the science part of the admission test (which contains chemistry, physics, mathematics, and biology). The limited number of students completing the survey is a clear limitation of the study. As the sample is limited, all learners who have at least one access to the platform and completed the survey are included in the study.

3.2 Variables and techniques

Once the events from the tracking logs are filtered, high-level variables are derived to be used in the prediction models. Particularly, indicators are classified depending on their relationship with accesses to the platform, videos, and exercises. The list of considered features is shown in Table 1. The dependent variable is the binary result of passing/failing the test.

Predictive models have been created using the library *caret*[1] of R, and four of the most common algorithms have been considered: Random Forest (RF), Generalized Linear Model (GLM), Support Vector Machines (SVM) and Decision Trees (DT). With these models, results are obtained using 10-fold cross-validation and 10 repetitions. AUC (Area Under the Curve) is used to evaluate the quality of the prediction as this metric is widely used, generally appropriate for student behavior classification problems (Pelánek, 2015), and avoids some problems that other metrics face (e.g. accuracy) in imbalanced datasets (Jeni, Cohn, & De La Torre, 2013).

[1] http://topepo.github.io/caret/index.html

Table 1: Features used in the study.

ID	Variable	Description
Variables related to accesses to the platform		
1	streak_acc	Longest consecutive run of accesses to the platform
2	ndays	Number of days the student has accessed to the platform
3	avg_con	Average number of consecutive days that the student accesses the platform
4	per_pc	Percentage of accesses from a PC (and not from a mobile, tablet, etc.)
5	per_wk	Percentage of accesses during weekend
6	per_night	Percentage of accesses during evening/night
Variables related to interactions with videos		
7	per_vtotal	Viewed percentage of total video time
8	per_compl	Percentage of completed videos
9	per_open	Percentage of opened videos
10	avg_rep	Average number of repetitions per video
11	avg_pause	Average number of pauses per video
Variables related to interactions with exercises		
12	per_attempt	Percentage of attempted exercises over the total
13	avg_grade	Average grade of formative exercises (only using the first attempts)
14	avg_attempt	Average number of attempts in the exercises attempted
15	per_correct	Percent of correct exercises over attempted exercises (using all attempts)
16	CFA	Number of 100% Correct exercises in the First Attempt
17	streak_ex	Longest consecutive run of correct exercises
18	nshow	Number of times the user asks for the solution of an exercise (without submitting an answer)

4 RESULTS

This section is divided into four parts, which address each of the first four objectives that were introduced in Section 1.

4.1 Differences between learners based on secondary school background

In this section, we analyze the differences of students based on their educational background. The medicine admission test can be taken by any student finishing secondary school, but students from educational tracks with sciences and math (traditional students, TR) are better prepared for the test compared to students who do not have this background (non-traditional students, NTR). The aim of this section is to analyze if there are significant differences in the learning behavior depending on the educational background. To do that, data was separated in four groups: (1) traditional students who pass (TP, n=92), (2) traditional students who fail (TF, n=22), (3) non-traditional students who pass (NTP, n=6), and (4) non-traditional students who fail (NTF, n=10). In addition, we measured the difference of students with respect to all the variables of Table 1. As not many learners completed the survey, the number of cases of some groups is limited. Therefore, we used the Mann-Whitney test to compare the groups. Table 2 shows the results when comparing different groups and the mean of each variable for each group.

Table 2: Statistical comparison between traditional and non-traditional students. P1: p-value TP-TF, P2: p-value TR-NTR, P3: p-value TP-NTP, P4: p-value TF-NTF

Variable	TR	NTR	TP	NTP	TF	NTF	P1	P2	P3	P4
streak_acc	1.87	3.31	2.04	3.17	1.14	3.40	0.01	$<10^{-2}$	0.04	0.01
ndays	12.18	18.44	13.76	22.17	5.55	16.20	$<10^{-4}$	0.02	0.04	$<10^{-2}$
avg_con	0.49	0.76	0.51	0.70	0.37	0.80	0.04	0.05	0.59	0.03
per_pc	0.88	0.94	0.91	0.97	0.76	0.93	0.14	0.75	0.66	0.26
per_wk	0.31	0.29	0.32	0.34	0.25	0.26	0.01	0.84	0.81	0.13
per_night	0.07	0.05	0.07	0.01	0.09	0.07	0.69	0.50	0.43	0.92
per_vtotal	0.57	0.78	0.62	0.88	0.34	0.71	$<10^{-2}$	0.01	0.05	0.01
per_compl	0.45	0.64	0.50	0.71	0.28	0.59	$<10^{-2}$	0.02	0.07	0.02
per_open	0.61	0.79	0.67	0.89	0.37	0.74	$<10^{-2}$	0.04	0.12	0.02
avg_rep	1.24	1.52	1.39	1.83	0.62	1.33	$<10^{-3}$	0.03	0.04	0.01
avg_pause	5.94	11.21	6.37	12.60	4.16	10.38	$<10^{-2}$	0.01	0.05	0.01
per_attempt	0.49	0.72	0.54	0.80	0.27	0.67	$<10^{-3}$	$<10^{-2}$	0.01	0.01
avg_grade	0.48	0.50	0.54	0.53	0.24	0.47	$<10^{-6}$	0.61	0.78	0.03
avg_attempt	1.42	1.81	1.57	1.66	0.78	1.90	$<10^{-3}$	0.03	0.97	$<10^{-2}$
per_correct	0.72	0.85	0.80	0.83	0.37	0.86	$<10^{-4}$	0.18	0.73	$<10^{-2}$
CFA	31.29	43.63	35.14	52.33	15.18	38.4	$<10^{-4}$	0.04	0.05	0.01
streak_ex	5.38	7.25	6.10	7.67	2.36	7.00	$<10^{-5}$	0.03	0.21	$<10^{-2}$
nshow	40.17	69.19	43.22	79.33	27.41	63.10	$<10^{-2}$	$<10^{-2}$	0.01	0.01

* p-values under 0.05 (confidence level) are colored in blue

Results show that there is statistical significant difference in most of the variables (excepting *perc_pc* and *per_night*) between TP and TF (no comparison has been done between NTP and NTF because of the few number of cases), which suggests that the learning behavior in the SPOC can influence success. Similar results are obtained when comparing all TR and NTR, with the exception of the variables related to user habits too (*perc_pc*, *per_wk* and *per_night*). Note that no statistical difference in some variables related to exercises achievement (e.g., average grade) were found for students who pass (TP vs. NTP). This entails that if the performance in the SPOC is similar, both groups can manage to pass. Nevertheless, NTP are more active on the SPOC as they access more often and watch more videos. Indeed, the SPOC format has the advantage that NTR, who have less background knowledge, can study at their own pace. Regarding the students who fail, there is statistical difference in most of the variables. NTR put more effort on the SPOC, and in some cases, they work harder than TP, as they access and watch more videos on average than TP. Their background seems to be a strong disadvantage however given the low number of NTP. To sum up, there are many differences in the behavior between TR and NTR and these groups should be treated separately to avoid bias in the models.

4.2 Anticipation of grades and results of predictive models

This section is focused on how success in the sciences part of the admission test can be predicted and more importantly, how early it can be anticipated. For that purpose, seven dates were selected (T_i) corresponding to crucial deadlines in the blended learning program (specific dates are in Table 3). T1, T2, and T4 correspond to the face-to-face interactive sessions that were organized to discuss problems on specific topics of the SPOC. At T3, traditional lectures were organized on topics that were not part of the SPOC, but that were crucial for the exam. The first session of the admission test was organized at T5, and the second at T6 (there were two sessions of the test to give a second chance to students who failed the exam). T7 includes all the interactions in the SPOC. With these dates, predictive models were trained for TR (NTR are excluded because they are very different from TR, and there are few students to develop models with representative samples, although it will be interesting to develop them if more NTR students appear in future editions) from the beginning of the course (September 7th) to each T_i. Table 3 shows the results of the models.

Table 3: Results of the predictive models (in AUC).

Period finish	T1 22/10	T2 14/01	T3 07/04	T4 06/05	T5 05/07	T6 30/08	T7
RF	0.46	0.45	0.70	0.78	0.84	0.87	0.87
GLM	0.59	0.71	0.72	0.73	0.74	0.77	0.77
SVM	0.55	0.51	0.72	0.73	0.84	0.85	0.85
DT	0.50	0.50	0.70	0.71	0.78	0.80	0.80

Results show that at the beginning of the course, the predictive power is poor. With an AUC threshold of 0.8 (as used by Moreno-Marcos et al., 2018), the predictive power of the model is only considered good from T5, the first session of the exam. A possible reason is the low activity at the beginning of the SPOC (57.45% of interactions occur after T4). If medium predictive performance (AUC=0.7) is acceptable (there is always a trade-off between anticipation and predictive power), the prediction from T3 can be considered. In that case, at least 31.03% of interactions are included, which is much more than the 13.43% in T2, which is not enough to predict. The low level of activity may also indicate that the SPOC does not really work in the synchronous way it was planned. That may affect the prediction because the activity is not uniform among students during the course. This can be important to reflect about the methodology. If face-to-face sessions with flipped classroom are organized, it would be advisable to enhance its relevance to ensure more people attends and are engaged from early stages.

In terms of the algorithms, the best model from T3 onwards is RF, which achieves an AUC of 0.87 at the end of the course. While differences are not big in some periods, this algorithm seems to be more consistent in this scenario. However, if the continuous grade was predicted and the RMSE (Root Mean Square Error) was used, SVM would be better (0.110 vs. 0.119), although both SVM and RF also perform better than the others.

4.3 Influence of variables on predictive models

After evaluating the predictive power of the models, the next challenge is to determine the variables that contribute most to the prediction, as this identifies the activities that are important for success. From the best model (RF in T7), the importance of the variables has been evaluated using the *Mean Decrease Gini*, which is often used to evaluate importance in RF (Louppe, Wehenkel, Sutera, & Geurts, 2013).

The results in Table 4 indicate that the average grade of exercises using only the first attempt (*avg_grade*) is the most important variable. This is reasonable as correct answers at first attempt indicate successful processing of the learning material, and after several attempts, the correctness of the answer can be affected by chance. Next, the number of days the user accesses (*ndays*) and the number of times the user asks for the solution (*nshow*) stand out. The last variable represents that students who demand and read the explanation of answers are more likely to pass. Regarding the variables about streaks, results show that long consecutive runs of correct exercises (*streak_ex*) have strong effect on success, unlike long consecutive runs of accesses to the platform (*streak_acc*). Finally, regarding video interactions, the variables that have more effect on success are the percentage of videos opened (*per_open*) and the number of times learners repeat the videos (*avg_rep*).

Table 4: Variable importance (VI) and correlation of variables of all students (CA) and traditional students (CT).

Variable	VI	CA	CT	Variable	VI	CA	CT
streak_acc	0.35	0.01	0.14	avg_rep	1.55	0.21	0.26
ndays	2.67	0.18	0.26	avg_pause	1.41	-0.05	0.07
avg_con	0.64	-0.06	0.04	per_attempt	1.33	0.20	0.30
per_pc	1.08	0.09	0.12	avg_grade	8.42	0.41	0.44
per_wk	1.13	0.15	0.14	avg_attempt	1.41	0.24	0.38
per_night	1.96	0.06	0.08	per_correct	2.14	0.35	0.47
per_vtotal	1.11	0.14	0.21	CFA	1.15	0.29	0.35
per_compl	0.62	0.12	0.19	streak_ex	2.12	0.32	0.40
per_open	2.10	0.17	0.24	nshow	2.34	0.10	0.21

5 DISCUSSION ABOUT THE GENERALIZABILITY

In Section 4, results of the prediction analysis in a SPOC were presented. Nevertheless, one important question is how results from this research can be generalized and extrapolated to different courses. Although results are valid for the analyzed SPOC, it is difficult to export the models because of the importance of the course context, which needs to be considered for the predictive models (Gašević, Dawson, Rogers & Gasevic, 2016). It may be possible to generalize the results in a very similar course (blended course with similar thematic), but results might change even in another run of the same SPOC if the context changes. For example, if more/less face-to-face sessions were organized, students might behave different and thus results may change. Similarly, if materials were all released at the beginning of the course or if students were required to do certain activities to continue after some deadlines, behaviors would also change and the interpretations of the results as well. Ocumpaugh, Baker and Gowda (2014) already experienced this problem when they developed EDM (Educational Data Mining) models to detect affective states with different populations and they analyzed whether their models were valid when changing the group of students.

Because of that, we believe there is no one-size-fits-all model to be used for all scenarios. Instead, existing models need to be taken and adapted to the specific context. This means that the scalability of the solution is about reuse and adaptation. For example, if we have different courses from the same source (e.g., several courses from edX), it is possible to use the same or similar algorithms to collect the indicators and train the models, but specific data of the course should be used, and the interpretations of the results should be done based on the methodology and pedagogical background of the course. If the course has some specific features, perhaps some new indicators could be included as part of the adaptation. This way, each model would be specific for each context. Moreover, this approach also opens the door to a possible framework to guide learning analytics researchers and developers in the process of adaption of the models. While the context is different, there are several common steps in the adaptation, such as the reutilization of indicators, and future work should be focused on analyzing this process.

Related to this, there is a question about the validity of the research results. Even if the results can vary depending on the context, results provide insight in effective learning behaviors and when combining results from different scenarios, it may be possible to reach global conclusions about how students learn and what behaviors have relevant effect on their success. If we consider the case study presented in this paper, one finding has been the differences in the behavior based on the background. While it is possible to find another course where educational background may not be so important, e.g., a possible introduction course to something where all learners start from scratch, this conclusion raises the importance of the background, and particularly in admission tests (same context), and suggests considering it when adapting the models to other contexts.

6 CONCLUSIONS

In this paper, an analysis of SPOC data, including predictions for success on a high-stake admission test, has been done. One interesting finding was that there are strong differences in the behavior of students depending on their background. Moreover, prediction models only behaved reasonably well in the last three months, which were also the months with more than half of the activity. Among the variables, the average grade using first attempts stands out, although other behaviors such as accessing to the platform regularly, asking for the solution of exercises and repeating videos had also a positive relationship with success. The discussion of the generalizability also points out that the course context is very important and that makes models need to be adapted to be reused in each scenario. This also opens the door to the definition of a framework to guide people involved in learning analytics in how to adapt and reuse their models.

With regard to the limitations of the study, it is noteworthy that the dataset was limited due to the lack of information about the admission test results. Moreover, that information was self-reported data and, although it appeared reliable, it could only partially be verified (62% of the cases). In future work, it would be interesting to include data about more cohorts to improve the dataset (and be able to develop models for non-traditional students). Furthermore, it would be interesting to design and evaluate some visualizations based on the prediction results to provide SPOC learners with useful interventions based on their interactions. Finally, it would be also interesting to develop a framework about how generalizability can be achieved and reflect about other factors, such as the stakeholder involvement, which are also important to guarantee the sustainability of the learning analytics solution.

ACKNOWLEDGEMENTS

This work has been co-funded by the Erasmus+ Programme of the European Union, through the project LALA (586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP), by the Madrid Regional Government, through the eMadrid Excellence Network (P2018/TCS-4307), and by the Spanish Ministry of Science, Innovation and Universities with the project Smartlet (TIN2017-85179-C3-1-R). The latter is financed by the State Research Agency in Spain (AEI) and the European Regional Development Fund (FEDER). It has also been supported by the Spanish Ministry of Science, Innovation and Universities, under an FPU fellowship (FPU016/00526). This publication reflects the views only of the authors, and funders cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., & Goodrich, V. (2014, March). Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. In *Proceedings of the 4th International Conference on Learning Analytics And Knowledge* (pp. 103-112). ACM.

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, *58*(1), 470-489.

Ashenafi, M. M., Riccardi, G., & Ronchetti, M. (2015, October). Predicting students' final exam scores from their course activities. In *Proceedings of the 45th IEEE Frontiers in Education Conference* (pp. 1-9). IEEE.

Brinton, C. G., & Chiang, M. (2015, April). MOOC performance prediction via clickstream data and social learning networks. In *Proceedings of the 34th IEEE International Conference on Computer Communications* (pp. 2299-2307). IEEE.

edX. (2018). *EdX Research Guide.* Retrieved from https://media.readthedocs.org/pdf/devdata/latest/devdata.pdf

Fancsali, S. E., Zheng, G., Tan, Y., Ritter, S., Berman, S. R., & Galyardt, A. (2018, March). Using embedded formative assessment to predict state summative test scores. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 161-170). ACM.

Feng, M., Heffernan, N. T., & Koedinger, K. R. (2006, June). Predicting state test scores better with intelligent tutoring systems: developing metrics to measure assistance required. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 31-40). Springer, Berlin, Heidelberg.

Fox, A. (2013). From MOOCs to SPOCs. *Communications of the ACM*, *56*(12), 38-40.

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

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013, September). Facing Imbalanced Data--Recommendations for the Use of Performance Metrics. In *Proceedings of the 5th International Conference on Affective Computing and Intelligent Interaction* (pp. 245-251). IEEE.

Jiang, F., & Li, W. (2017). Who Will Be the Next to Drop Out? Anticipating Dropouts in MOOCs with Multi-View Features. *International Journal of Performability Engineering*, *13*(2), 201-210.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In *Proceedings of the 26th International Conference on Neural Information Processing Systems* (pp. 431-439).

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Delgado Kloos, C. (2018, September). Predicting Learners' Success in a Selfpaced MOOC Through Sequence Patterns of Self-regulated Learning. In *Proceedings of the 13th European Conference on Technology Enhanced Learning* (pp. 355-369). Springer, Cham.

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2018). Prediction in MOOCs: A review and future research directions. *IEEE Transactions on Learning Technologies* (In Press).

Moreno-Marcos, P. M., Muñoz-Merino, P. J., Alario-Hoyos, C., Estévez-Ayres, I., & Delgado Kloos, C. (2018). Analysing the predictive power for anticipating assignment grades in a massive open online course. *Behaviour & Information Technology*, *37*(10-11), 1021-1036.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, *45*(3), 487-501.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017, March). A neural network approach for students' performance prediction. In *Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 598-599). ACM.

Orlando, M., & Howard, L. (2018). Setting the Stage for Success in an Online Learning Environment. In *Emerging Self-Directed Learning Strategies in the Digital Age* (pp. 1-9). IGI Global.

Park, Y., & Jo, I. H. (2015). Development of the Learning Analytics Dashboard to Support Students' Learning Performance. *Journal of Universal Computer Science*, *21*(1), 110-133.

Pelánek, R. (2015). Metrics for evaluation of student models. *Journal of Educational Data Mining*, *7*(2), 1-19.

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Delgado Kloos, C. (2018). Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators. *Expert Systems*, e12298.

Roggemans, L., & Spruyt, B. (2014). Toelatingsproef (tand) arts: een sociografische schets van de deelnemers en geslaagden. Brussel: Onderzoeksgroep TOR, Vakgroep Sociologie, Vrije Universiteit Brussel (140 blz.)-TOR, 29.

Xu, B., & Yang, D. (2016). Motivation classification and grade prediction for MOOCs learners. *Computational intelligence and neuroscience*, *2016*, 4.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, *29*, 23-30.

Yu, C. (2018). SPOC-MFLP: A Multi-feature Learning Prediction Model for SPOC Students Using Machine Learning. *Journal of Applied Science and Engineering 21*(2), 279-290.