

Aligning Learning Design and Learning Analytics through Instructor Involvement: A MOOC Case Study

Erkan Er
Universidad de Valladolid
erkan@gsic.uva.es

Eduardo Gómez-Sánchez
Universidad de Valladolid
edugom@tel.uva.es

Yannis Dimitriadis
Universidad de Valladolid
yannis@tel.uva.es

Miguel L. Bote-Lorenzo
Universidad de Valladolid
migbot@tel.uva.es

Juan I. Asensio-Pérez
Universidad de Valladolid
juaase@tel.uva.es

Susana Álvarez Álvarez
Universidad de Valladolid
susana.alvarez@uva.es

Abstract. With the emergence of MOOCs, there is a growing interest in prediction research. Most existing predictive models do not consider the context for which they are intended, thus resulting in limited impact. Learning design (LD) can provide a contextual understanding for the design of predictive models in collaboration with the instructors, maximizing their potential for supporting learning. This paper presents the findings of a mixed-methods research that explored the potentials emerging from aligning LD and LA during the design of a predictive analytics solution and from involving the instructors in the design process. The context was a past MOOC, where the learner data and the instructors were accessible for posterior analysis and additional data collection. Through a close collaboration with the instructors, the details of the prediction task were identified, including the learning activity to focus on, the target variable to predict, and the practical constraints to consider. Later, two predictive models were built for the prediction task identified: LD-specific model, in which the features were based on the LD and pedagogical intentions, and a generic model, which was based on cumulative features, not informed by the LD. Although the LD-specific predictive model did not outperform the generic one, some features derived from the LD and pedagogical intentions were predictive. The quantity and the power of such features were associated with the degree to which the students acted as guided by the LD and pedagogical intentions. The leading instructor's opinion about the importance of the learning activities in the LD was compared with the results of the feature importance analysis. This comparison helped identify the parts of the LD that need improvement. That is, the results of the LA informed back the LD, where the instructor was a mediator. The implications for improving the LD are discussed.

Keywords: MOOCs, learning design, learning analytics, predictive analytics, collaborative learning

1. Introduction

In the recent years, the growth of massive open online courses (MOOCs) has attracted much attention on predictive analytics, a popular area of learning analytics (LA). The unprecedented amount of learner data generated in massive online courses is providing great affordances for building accurate predictive models of student engagement and success (Moreno-Marcos, Alario-Hoyos, Muñoz-Merino, & Kloos, 2018). Although numerous predictive models and techniques have been developed, they have had a limited capacity in improving the teaching and learning processes in real-world contexts (Gardner & Brooks, 2018). One main reason for the limited impact is that many of these works do not take into account the pedagogical context (Gašević, Dawson, Rogers, & Gašević, 2016). That is, they are decoupled from the contexts for which they are intended, and therefore they offer insufficient pedagogical affordances (Gašević et al., 2016).

Learning design (LD), which is the blueprint of “the sequence of learning tasks, the resources, and the sequence of teaching methods” (Lockyer, Heathcote, & Dawson, 2013,

p.1441), can provide a contextual understanding that can inform the design of predictive models and enhance their potential use for improving pedagogy. First, the features (or variables) generated to train predictive models can be decided considering the impact of the LD on student activities. For example, instead of the total number of video views, a commonly used generic feature in the literature, LD could be used to interpret the varying importance of each video on student learning and to decide on a relevant feature. Accordingly, several research works have noted the relevance of LD in building predictive models (Gašević et al., 2016; Veeramachani, O'Reilly, & Taylor, 2014). Yet, the role of LD during the design of predictive models remains largely unknown.

Moreover, instructors may play a crucial role in aligning the predictive analytics solutions with LD and their pedagogical intentions. In order for the predictive models to be used in the targeted context, they have to serve a predetermined pedagogical need decided by the course instructors (or instructional designers). Without their involvement, predictive models are highly likely to serve only for experimentation, as in most dropout prediction models in the literature (e.g., Halawa, Greene, & Mitchell, 2014; Kurka, Godoy, & Von Zuben, 2016). Previous research noted that the close collaboration between the researchers and the designers can lead to LA interventions that connects well with particular needs of a learning context (Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2015; Rodríguez-Triana, Prieto, Martínez-Monés, & Asensio-Pérez, 2017).

Thus, LD and LA can offer new opportunities for improving teaching and learning when one informs the other (Eradze, Rodríguez-Triana, & Laanpere, 2017; Mor, Ferguson, & Wasson, 2015; Schmitz, Limbeek, Greller, Sloep, & Drachsler, 2017). However, this synergic relationship has been understudied in the area of predictive analytics. Attending to this gap, this paper reports on a research investigating the opportunities emerging from aligning LD and LA in a particular MOOC context in close collaboration with the whole instructional designer team (of this particular course). The instructional designers played an active role when: (1) determining a relevant prediction task, (2) identifying the constraints imposed by the context, (3) assessing the importance of learning activities included in the LD, and (4) refining the LD of the course by analysing the results. LD and LA were connected in two ways: (1) the LD and the pedagogical intentions were used to guide feature generation, and reciprocally (2) the results of the predictive analysis were used to review and improve the LD for the future run of the same course. This research is guided by the following research questions:

- RQ1: What are the potentials emerging from aligning LA and LD during the design of a predictive analytics solution?
- RQ2: What are the potentials emerging from involving the instructional designers in the design process of the predictive analytics solution?

The paper is structured as follows. Section 2 discusses the related literature and identifies the need for research, and Section 3 describes the context and the research design. Then, Section 4 presents the results in three sections: the prediction task and the practical constraints (as determined by the instructor), the predictive models and their accuracies, and the comparison of instructor opinion with the results of the predictions. Section 5 discusses the findings organized around the research questions. The paper concludes with Section 6, which presents the limitation of this research work and suggest future research.

2. Background: Learning Design, Learning Analytics, and MOOCs

A critical issue often echoed in the LA literature is the lack of real-time interventions that align well with instructors' pedagogical needs (Harrer & Göhnert, 2015; Nguyen, Rienties, & Toetnel, 2017; Wise & Vytasek, 2017). As a result, the LA researched has yielded a limited impact so far. In their recent literature review of LA dashboards, Jivet, Scheffel, Drachsler, and

Specht (2017) found that LA solutions are built mainly based on the motivation to take advantage of the trace data available instead of a clear pedagogical intention. Similarly, in their systematic literature review on teaching and LA, Sergis & Sampson (2017) noted only few efforts to convert LA into actionable recommendations for instructors to refine the pedagogy based on the contextual needs. Thus, even though considerable progress has been made in conceptual development and design of LA (Bakharia et al., 2016; Lockyer et al., 2013; Manolis & Karkalasi, 2017), the products generated (e.g., data, visualizations, and tools) often ignore the influence of the pedagogical context (Eradze et al., 2017; Nguyen et al., 2017) and do not align well with instructors' needs for refining their pedagogy and learning activities (Mor et al., 2015). For a greater impact on practice, LA solutions need to be implemented in a way that is sensitive to the needs of learning context for which they are intended (Harrer & Göhnert, 2015; Wise & Vytasek, 2017).

The trace data used in LA are collected through students' interactions with the learning systems (e.g., LMS, MOOC platforms), and these interactions are largely shaped by the LD (Gašević, Mirriahi, Dawson, & Joksimović, 2017). Therefore, understanding the LD is important for the meaningful interpretation of learning outcomes (e.g., visualizations of students' daily course visits) generated based on the trace data (Mangaroska & Giannakos, 2017). For instance, LD would determine the roles of the instructor and students in an online discussion activity and shape their interactions (Lockyer et al., 2013). Instructors may play a central role or may intend to promote peer interactions in discussions depending on the pedagogical intentions. LD of MOOCs may play a similar role in shaping learner' interactions with the MOOC platforms and could help identify relevant features when building predictive models in MOOC settings. However, given that MOOCs are non-formal learning settings with massive numbers and variety of participants (Kizilcec, Piech, & Schneider, 2013), previous findings from formal learning contexts should not be taken for granted. Yet, there is scarce research on the use of LD to inform predictive analytics in MOOCs and to create interventions driven by the context.

Most MOOC research has exploited a posteriori student data to understand the learning behaviour of MOOC participants (Kizilcec et al., 2013; Tseng, Tsao, Yu, Chan, & Lai, 2016) and to identify the variables that can explain learners' engagement and success (Bote-Lorenzo & Gómez-Sánchez, 2017; Er, Bote-Lorenzo, Gómez-Sánchez, Dimitriadis, & Asensio-Pérez, 2017). These research studies have substantial value in establishing the theoretical foundations of how individuals behave, engage, and learn at massive and open learning settings (Brooks & Thompson, 2017; Liu & Koedinger, 2017). However, their impact in the practice of MOOC teaching and learning is still questionable. Accordingly, in their review of the literature on MOOC prediction research, Gardner & Brooks (2017) underlined the need for practical predictive analytics research in real-world MOOC contexts in order to close the gap between LA research and MOOC practice.

Thus, there is gap in the MOOC literature regarding the potentials emerging from aligning predictive analytics with LD. MOOC predictive analytics have been based on experimental studies with minor direct effects in MOOC practice (Gardner & Brooks, 2018). On the other hand, in collaboration with instructors, LD can inform the design of LA solutions and lead to predictive models for designing interventions that target the contextual needs (Mangaroska & Giannakos, 2017).

3. The Current Study

3.1. Context

The context of this research was a MOOC that teaches the translation of terms from English/Spanish to Spanish/English in the domain of finance, offered by a Spanish university in the Canvas Network MOOC platform. The enrolment was closed at the end of the second week.

Total number of enrolled students was 1031. Regarding the gender distribution of the students, the majority was woman (75%). Given that the course was taught in Spanish, a great portion of the students were from Spain (57%) and Latin America (31%). Most participants were between 20-30 years old (61.2%) and holding at least Master (26 %), Bachelor (53%) or High School (16%) degrees.

The lead course instructor is experienced in online teaching, mostly in formal learning settings with relatively fewer students. This MOOC was her first teaching experience in massive and open learning context. Moreover, the instructor's field of expertise is in the domain of Translation and Interpretation. The other two instructional designers were PhD students in the domain of the educational technologies who worked with the leading instructor to design the course.

The course was composed of 7 blocks (or modules or weeks). Each block was available one week after another, and once the block was published, its content remained accessible until the end of the course. Two collaborative activities were designed in this MOOC, involving the extraction of relevant terms from given documents dealing with finance. Groups of six participants were formed in both collaborative activities. The instructional designers of the course consisted of a leading instructor and two teaching assistants, all of which were involved in this research. Further details about the design of the course are provided in Appendix C and D.

3.2. Method

This is a mixed-methods study (Greene, 2007) that employs both qualitative and quantitative data to address the research questions. The design of the study includes three stages as delineated in Figure 1. In the first stage (see section 4.1), qualitative data were collected through a face-to-face meeting with the teaching assistants and an online survey filled by the lead instructor (see Appendix A), with the goal of identifying the learning design, pedagogical intentions and the prediction task. The LD of the course (see Appendix C) was derived from the course pages in the Canvas Platform and finalized with the help of the teaching assistants during the face-to-face meeting. The meeting notes were analysed using thematic analysis (Marshall, C. & Rossman, 1999) to identify the pedagogical intentions. The course components in the LD were used as the themes to guide the analysis. This analysis resulted in the pedagogical intentions behind each learning element (see Appendix D) and the learning activity to be targeted with predictive analytics.

Moreover, in the first stage, an online survey was administered to the lead instructor to identify the target variable for the prediction along with the practical constraints. In the survey, the instructor was asked to rank the possible engagement variables to predict about the target activity according to their pedagogical usefulness. The instructor was also asked to indicate when the intended predictions would be needed (i.e., how early before the target activity starts), which was used as a practical constraint to consider in the design of the prediction models.

The second stage (see section 4.2) involved building and running two models within the practical contexts for the prediction task decided in the first stage: 1) a generic model, built with cumulative features (e.g., total number of discussion posts), and 2) a LD-specific model, built with features informed by the LD and pedagogical intentions. Then, the prediction accuracies of these models were assessed and compared. The predictive features were generated from the learner activity logs (as the quantitative data source) retrieved using the API of the Canvas Network MOOC platform. Consent from the MOOC participants for the use of their activity logs was obtained through the welcome survey administered at the initial days of the course. All learner data were anonymized before the analysis. After the anonymization, the data files that contain identifying user information were deleted. Any corrupted logs were discarded.

As further discussed in section 4.2., qualitative data (i.e., LD and pedagogical intentions) were used to inform the analysis of the quantitative data (i.e., feature generation) when building the LD-specific model. That is, the research design involved mixing for sequential exploration

(Leech & Onwuegbuzie, 2009). Note that the personnel who worked with the research data had undergone a training program on Human Subject Research.

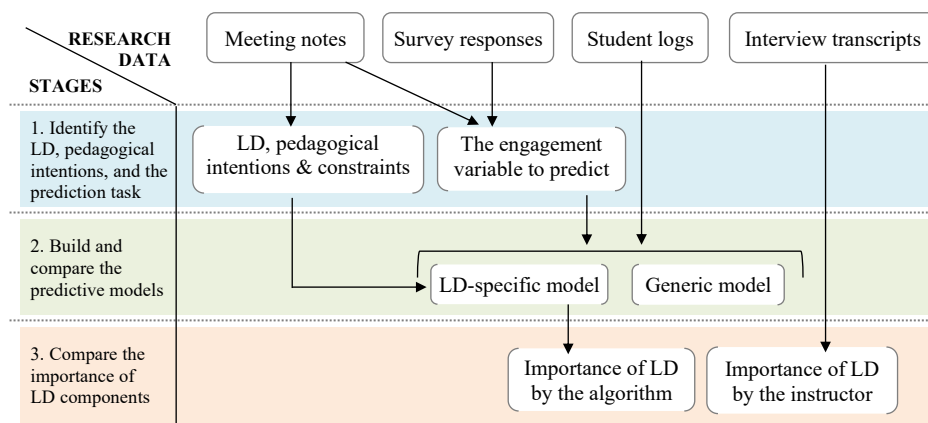


Figure 1. Design of the study

In the last stage (see section 4.3), LD-specific features were analysed based on their predictive powers determined by the machine learning algorithm, which helped identify the important learning activities in the LD. The number of corresponding features and their predictive powers were used to identify the importance of the learning activities. Then, through an online interview (see Appendix B), instructor ranked the same learning activities based on their importance for student learning, which was compared with the algorithmic results by the researcher simultaneously during the interview and the results were shared with the instructor. Instructor reflected on the discrepancies emerging from the comparison and provided possible reasons for the differences. The interview was audio recorded and transcribed for the analysis.

The interview transcript was analysed using a general inductive approach (Thomas, 2006). The general inductive approach is a simple yet effective approach to derive findings from qualitative data shaped with focused questions that require concrete answers. Similarly, the interview was guided by the discussion on specific discrepancies between the instructor and algorithmic ranking. We first formatted the transcription in a common format and then performed a close read to get familiar with the content. During the analysis, we first created the general categories (e.g., problems in conveying the pedagogical intentions to the learners) and then refine them to identify specific themes (e.g., poor assignment descriptions, low accessibility of the syllabus), which helped us capture the core message reported by the instructor. Thus, through the inductive approach, we were able to effectively convert the instructor input into a structured summary format and to establish links with particular course elements and activities (that resulted in discrepancies) for deriving useful implications (see section 5.3).

Logistic regression was used as the classifier algorithm to train the predictive models (i.e., generic and LD-specific) as it has been effective in many classification tasks in the MOOC literature (Kizilcec, Pérez-Sanagustín, & Maldonado, 2016; Whitehill, Mohan, Seaton, Rosen, & Tingley, 2017b). Features were calculated based on the student activity after February 14 since the logs before that day were missing in the database of the Canvas platform. Students with no visits to any course pages were removed from the data. All features were standardized before they were used for training and testing the models. Learners with zero pageviews (n=458) were eliminated to avoid noise in the analysis.

Given the unbalanced distribution of the class labels, the performance of the models were evaluated using area under the curve (AUC) as performance metric (Hanley & McNeil, 1982).

The AUC score of a classifier refers to the likelihood of ranking a randomly chosen positive example higher than a randomly chosen negative example. The categorization of model performance based on AUC scores is: .9-1: excellent, .8-.9: very good, .7-.8: good, .6-.7: fair, and .5-.6: bad (Bekkar, Djemaa, & Alitouche, 2013). The research on human behavior prediction considers AUC values of .7 and higher reasonably accurate (Rice & Harris, 2005). Previous MOOC research have regarded the models with such accuracy (i.e., $AUC > .7$) as robust models (Boyer & Veeramachaneni, 2016). Even though the model performances will be evaluated based on the AUC scores, precision, recall, and f1-scores were also presented to provide a more comprehensive view of the model performances.

Feature selection was performed using randomized logistic regression (RLR). The use of RLR for feature selection is also known as stability selection (SS) (Meinshausen & Bühlmann, 2010). The purpose of using SS in this study was to be able to interpret the feature importance rather than to select features. RLR assigns a score to each feature indicating their predictive power. SS has been used in MOOC prediction research (Taylor, Veeramachaneni, & O'Reilly, 2014; Veeramachaneni et al., 2014). The Scikit-Learn (Pedregosa et al., 2012) implementations of logistic regression and RLR were used.

4. Results

The results are organized by the stages of the research (see Figure 1).

4.1. Stage 1: The prediction task and its practical requirements

The first stage involved identifying prediction task and its practical requirements as determined based on the qualitative data collected from the teaching staff. According to the meeting notes with the teaching assistants, the collaborative learning activities were identified as critical for student learning for several reasons. The teaching assistants noted that the collaborative learning activities enabled students (1) “to acquire prerequisite skills necessary for translating terms in the business domain”, and (2) “to directly interact with peers”, thus allowing them “to learn from peers’ knowledge and experience” as well as “to feel more connected with other learners”. Therefore, collaborative learning activities were decided for a possible intervention through predictive analytics. Among the two activities (one in the fourth week and another in the sixth week), for the scope of this study, we focused on the first one.

To determine the information to be produced by predictive analytics for intervening the activity, through an online survey the instructor ranked four specific pieces of information based on their pedagogical usefulness. According to the rankings of the instructor (see Table 1), knowing *whether the number of students who participate in the group discussion will be higher than a threshold (i.e., 3)* was considered the most useful. As the rationale behind this decision, the instructor noted that “by utilizing this information regarding students’ participation levels, a certain level of interaction in groups can be assured and the workload per group members would be manageable.”

Table 1

The instructor’s ranking on the usefulness of information needed for the collaborative activity

Ranking	Information
1	Whether the number of students who participate in the group discussion will be higher than a threshold. (e.g., at least 3 students make a post in a group)
2	Whether the number of messages posted in the group discussion will be higher than a threshold (e.g., having at least 5 discussion entries)
3	Whether the collaborative group will submit their assignment or not.
4	Whether a student will post in the group discussion or not.

The prediction task was aimed at identifying if a student would participate in group discussions (i.e., post at least one entry) or not, as these predictions would be needed to obtain the number of active participants in groups as demanded by the instructor. In this prediction task, students with at least one entry are identified as contributors, and students with zero entries are identified as non-contributors. Please note that this decision on the target variable to predict was determined by what was considered pedagogically more useful from the instructor's perspective. A different classification (e.g., very active contributors, active contributors, low active contributors and not contributors) could be preferred in a different course depending on pedagogical intentions in the context.

For the predictions to be usable for intervening in real time, they would need to be available when they are still considered actionable by the instructor. For this purpose, through the online survey, the leading instructor's preference was obtained. The instructor indicated that she would need the predictions by the 2nd of March, which was then used as the cut-off point for calculating the features. The underlying reason for the instructor's decision was (1) to allow students to make some progress on the collaborative activity as they had probably different availability and schedule and (2) to have enough time (around 10 days) before the activity deadline to be able to intervene using the predictions. Please note that the decision of having the predictions 10 days before the target activity may not apply to different contexts. We recommend allowing instructors to decide on the best date according to contextual factors and pedagogical intentions.

4.2. Stage 2: Building and comparing the classification models (LD-driven vs generic model)

The feature engineering for the LD-driven model and the generic model are provided below. The performances of the models are also compared.

Feature engineering for LD-driven model (mixing for sequential exploration)

The LD of the MOOC (see Appendix C) and pedagogical intentions (see Appendix D) were used to inform the feature generation (i.e., variables about student activities) that can possibly lead to an accurate model predicting participation in group discussions. Please note that all the features described below were derived by the research team, directly informed by the LD and the pedagogical intentions.

Regarding the potential influence of the pedagogical intentions on students' activities several features were derived. In particular, given the intent of the review videos to promote the self-evaluation and recap as necessary, the subsequent student activity could be an indicator of student engagement. For example, after watching the review video a student might identify a gap in his own understanding and may want to revisit the related lecture content. Accordingly, the research team identified 4 features based on the sequence of activities in which visiting a review-video page is anterior, resulting in 13 sequence features in total. Moreover, learning resources and activities in a block could be important for student learning in subsequent blocks. For example, the glossary document created in Block 1 was intended to be a useful resource for students during the whole course. Accordingly, three sequence features were derived based on students' visits to previous blocks. These features indicate the number of visits from Block 3 to preceding blocks (i.e., Block 2, Block 1, and Block 0), and from Block 2 to preceding blocks (i.e., Block 1 and Block 0). Thus, informed by the instructors' pedagogical intentions, in total 15 sequence features were generated by the research team. The sequence of learner actions was previously noted in the MOOC literature as important to student learning (Boroujeni & Dillenbourg, 2018)

Moreover, features were generated based on each particular content item in the LD. The first set of variables includes page-view features that convey student engagement in each specific

learning component. Links to PDF files (if exists within a content page) were considered separately, since they may signal higher student interest and engagement in the corresponding activity or the content. In total, 48 page-view features were generated. Last, 17 features regarding students' active participation in the course were generated, including one or all of the following activities depending on the block: taking a quiz, posting in a discussion forum, and submitting an assignment. Please note that previous MOOC research has showed the predictive capacity of such learning activities but using cumulative features (Bote-Lorenzo & Gómez-Sánchez, 2018; Veeramachaneni et al., 2014; Whitehill et al., 2017b), as in this work to build the generic model.

Table 2
Number of LD-specific features per each block

	Block 0	Block 1	Block 2	Block 3
<i>Features regarding students' sequence activities</i>				
Number of sequences of review-video and lecture content visits	1	1	1	1
Number of sequences of review-video and discussion forum visits	1	1	1	1
Number of sequences of review-video and quiz visits	1	1	0	0
Number of sequences of review-video and assignment visits	0	1	1	1
<i>Features regarding students' page-view activities</i>				
Number of page-view features	6	9	12	12
<i>Features regarding students' active participation</i>				
Number of discussion posts	2	1	1	1
Number of quiz attempts	0	1	1	0
Time spent in quizzes	0	1	1	0
Quiz scores	0	1	1	0
Assignment-submission status	0	0	2	2

In total, 80 features were generated. The names of these features and their distribution across different blocks are provided in Table 2. These features were calculated based on the student activity logs beginning from the course start until the cut-off point (March 2), resulting in 419 students. Histogram plots along with some descriptive statistics of all features are provided in the Appendix E. Feature names are composed of two parts: the first part is the name of the associated item in the LD, and the second part is the suffix to describe the feature. This notation is used to better communicate the association of the features generated with specific components of the LD. The explanations for the suffixes are given in Table 3.

Table 3
The explanations for the suffixes used in feature names

Suffix	Meaning
*_PAGEVIEW	Page-view count
*_DISCENTRY	Entry count in a discussion
*_ASSIGNSUBM	If assignment is submitted or not
*_QUIZATTEMPT	Number of quiz attempts
*_QUIZSCORE	Quiz score
*_QUIZTIMESPENT	Minutes spent on a quiz
*_RVID_DISC	Number of times visiting a discussion after a review-video page
*_RVID_LECTURE	Number of times visiting a lecture content page after a review-video page
*_RVID_QUIZ	Number of times visiting a quiz page after a review-video page
*_RVID_ASSIGNMENT	Number of times visiting an assignment page a review-video page

*_ACROSSBLOCKS Number of visits to previous blocks

Feature engineering for the generic model

In the feature engineering for the generic model, no specific information about the LD and the pedagogical intentions were used. These features are widely used in the predictive analytics research. Features and their descriptions are provided in Table 4.

Table 4
Descriptions of generic the features

Feature name	Description
TTL_INFOPAGE_VIEW	Number of times students visited introduction pages
TTL_LECCONT_VIEW	Number of times students visited lecture-content pages
TTL_DISC_VIEW	Number of times students visited discussion forums
TTL_QUIZ_VIEW	Number of times students visited quiz pages
TTL_RVID_VIEW	Number of times students visited review-video pages
TTL_ASSIGN_VIEW	Number of times students visited assignment-related pages
TTL_PAGE_VIEW	Number of times students visited any course page
TTL_QUIZ_ATTEMPT	Number of quiz attempts
TTL_QUIZ_TIMESPENT	Total time spent on quizzes
TTL_QUIZ_SCORE	Total quiz score
TTL_ASSIGN_SBM	Number of assignment submissions
TTL_DISCENTRY	Number of discussion posts

Comparing the classification models

To test the performance of the classification models, 10-fold CV was used. The accuracy of the predictions for both models are provided in Table 5. The results show that the classification models perform quite accurate (AUC scores are 0.858 and 0.822 for LD-driven and generic models, respectively). Thus, the generic model built with cumulative features has very similar predictive capacity to that of LD-drive model that is more particular to the context.

Table 5
The performance scores of prediction accuracies for LD-Driven and generic models

	AUC	Precision	Recall	F1-score
LD-Driven Model	0.858	0.880	0.770	0.821
Generic Model	0.822	0.862	0.737	0.794

4.3. Stage 3: Instructor opinion on the importance of learning activities and comparison with feature importance analysis (mixing for concurrent triangulation)

The importance of learning activities as ranked by the instructor was compared with those determined based on the feature importance as automatically determined by the SS approach. The results are provided in Table 6. In this table, the first column indicates the name of the LD component, which are shown in Appendix C. The number of predictive features in each category are provided in the second column. These numbers are accompanied with average of the percentages of the selected features. As discussed in section 4.2, this percentage represents the predictive power. For example, 100% would indicate a very high predictive capacity as the feature would have been selected within all inspected feature subsets. In the third column, the rankings of the instructor are provided.

According to the results, in general, the instructor rankings on the importance of various LD components did not match with the rankings determined by the SS approach. To begin with,

there was a disagreement on the importance of blocks (or modules) except Block 3. This disagreement was stronger regarding the importance of discussions and content pages. Although the instructor considered these components critical to student learning, low engagement in these components led to very minor predictive power of the features associated with them. Similarly, the instructor considered both quizzes equally important; however, according to the results of SS, quiz activities were barely predictive.

Table 6

LD components, number of features selected, and instructor ranking

LD Components	Number of features	Instructor ranking
<i>Blocks (1: the least important, 4: the most important, and 0: not relevant)</i>		
Block 00	1 (35%)	0
Block 0	0	4
Block 1	10 (20%)	2
Block 2	14 (33.32%)	1
Block 3	6 (8.83%)	3
<i>Assignments (1: the least important, 4: the most important, and 0: not relevant)</i>		
IA1	6 (27.75%)	4
IA2	7 (39.43%)	1
IA3	2 (9.0%)	3
IA4	2 (23.5%)	2
<i>Discussions (1: the least important, 6: the most important, and 0: not relevant)</i>		
BLK00_4	1 (35%)	5
BLK00_5	0	6
BLK0_2	0	4
BLK1_2	0	4
BLK2_2	1 (1%)	4
BLK3_2	0	4
<i>Resource pages (1: the least important, 9: the most important, and 0: not relevant)</i>		
BLK3_1	1 (0.5%)	6
BLK3_5	0	6
BLK1_41	1 (6.5%)	9
BLK0_1	0	6
BLK0_3	0	6
BLK1_1	0	6
BLK1_5	1 (9%)	6
BLK2_1	2 (40.2%)	6
BLK2_5	0	6
<i>Quizzes (1: the least important, 2: the most important, and 0: not relevant)</i>		
BLK00_4	0	1
BLK00_5	2 (5.5%)	2
<i>Views after review-video (1: the least important, 5: the most important, and 0: not relevant)</i>		
*_RVID_INTRODUCTORY	0	0
*_RVID_DISC	0	0
*_RVID_LECTURE	0	5
*_RVID_QUIZ	0	4
*_RVID_ASSIGNMENT	0	0
<i>Views across blocks (1: the least important, 2: the most important, and 0: not relevant)</i>		
BLK2_ACROSSBLOCKS	1 (46.5%)	0
BLK3_ACROSSBLOCKS	1	2

Regarding the assignments, there was a higher agreement between what the instructor considered important and what was turned out to be important by the SS approach, except IA2, which was the most predictive in this prediction task but ranked to be least important by the instructor. Last, among the sequence features, which were generated based on the pedagogical intentions, none of the review-video sequence features were found to be predictive, although

instructor considered two of them important. Results were similar regarding the block sequence features.

5. Discussion

The discussion on the findings is organized around the research questions as follows.

5.1. What are the potentials emerging from involving the instructional designers in the design process of the predictive analytics solution?

The instructional designers played an active role during the design of the predictive analytics solution in the current MOOC context. Their involvement was critical in that the predictive analytics solution was created towards the learning needs in the context. Collaborative activity was identified by the instructors as the target learning activity. The research team identified some relevant information items that could be produced by predictive analytics and that could offer an opportunity for intervening the collaborative activity. In this step, the researchers' experience in predictive analytics was essential to guide the instructor in her decision to identify the information that could be useful for an intervention while ensuring that this information could be produced by predictive analytics. Thus, similar to the efforts of involving the instructors in the loop in formal learning settings (Rodríguez-Triana et al., 2015, 2017), collaboration with MOOC instructors offers a great value for creating predictive analytics solutions tailored to particular needs of the MOOC context.

Later when building the prediction models, instructor participation allowed the research team to identify features that are specific to the LD of the course and that are informed by the pedagogical intentions. With these features, an LD-driven predictive model was created, which has produced relatively higher accuracy in comparison with the generic model. The instructor involvement also continued when interpreting the importance of various learning activities included in the LD. Instructor opinion was compared with the importance of the learning activities based on the feature importance analysis. Generally, the instructor's opinion on the importance of course components was contradictory to that identified by the SS approach, although some congruences were noted as well. This was because what the instructor has planned pedagogically was not fully realized in practice in students' behaviours. However, this comparison provided hints on the flaws or weaknesses in the LD of the course and allowed the instructor to make informed decisions to improve the pedagogy of the course. In the subsection 5.3, the implications for the LD of the course are described.

Based on these findings, we recommend the active engagement of whole instructional design team when creating predictive analytics solutions. The active participation of all can result in a deeper understanding of the learning design and yield a more rigorous feature selection process.

5.2. What are the potentials emerging from aligning LA and LD during the design of the predictive analytics solution?

This research also noted important findings regarding the role of LD and pedagogical intentions in building predictive models in MOOC contexts. One might expect that the LD-driven model would outperform the generic model as it carries features that are more refined representations of student engagement in the course. This finding somehow contradicts with previous research noting the influence of LD on students' interactions with learning systems in formal learning settings (Gašević et al., 2017). This contradiction could be associated with the characteristics of the MOOC learning contexts. MOOCs are non-formal learning settings where a high variation exists among its massive number of participants (Kizilcec et al., 2013). Generally, the intentions to earn the final course certificate are the main drivers for active MOOC learners who intend to

complete learning tasks and assignments (Sharma, Jermann, & Dillenbourg, 2015). In the context of this study, assignments, which needed to be completed for receiving the course certificate, seem to have a similar effect. In both models, features regarding students' assignment activities (e.g., submissions, visits to assignment description pages, downloads of the PDF resources associated with assignments) were the most predictive. Indeed, LD-driven predictive model produced results that were accurate enough for practical use. However, two cumulative features (one regarding the assignment submissions and one regarding the visits to assignment-related pages) were sufficiently predictive in the generic model, leading to results that were as accurate as the LD-driven model.

Nonetheless, influence of the LD and pedagogical intentions was noted in the prediction model even though it was limited. Specific features about student interaction with different components of the LD (whether it is a content page, discussion forum, or quiz) offered varying predictive powers, which would not be captured otherwise with the generic approach, as followed in most MOOC research to date (Taylor et al., 2014; Whitehill et al., 2017a). For example, some content-pageview features were more predictive than the others. This finding suggests that certain content pages were more important to student learning than others. Such differences would not be captured by a single generic feature that cumulates all content pageviews. Aligned with previous research findings (Gašević et al., 2016; Lockyer et al., 2013), these results suggest that different elements of the LD might have varying levels of influence on student behaviour.

5.3. Implications for the learning design of the MOOC

Based on the analysis of the interview transcripts, several implications were derived for improving the LD of the MOOC. First, participation in forum discussions was not found to be predictive, which was associated with the fact that students did not use them. Instructor plans to use some gamification strategies such as badges to increase student motivation and engagement in the discussions (e.g., earning a badge when posting a useful question). Moreover, to enable students to benefit from the resource-sharing forum more effectively, instructor suggested that the resources shared by students could be compiled and presented in a separate page (or a Google Document).

Similarly, the review videos were rarely visited by students, thus suggesting their limited impact on student learning although they were considered critical by the instructor. Based on this finding, in the next run of the course, the instructor plans to place them before the assignments or quizzes in each block and insert a link to review videos in the descriptions of the assignments. Moreover, the results of the predictive analytics indicated that visits between blocks may be an important indicator of student engagement in the current context. To exploit the potentials of such learner behaviour, the instructor plans to place direct links to related and useful learning resources (e.g., review videos, Google Document of the glossary terms) across different blocks.

Regarding the limited capacity of the pedagogical intentions on shaping the learning behaviour as desired, the instructor noted that the intent of the learning tasks and resources might not have been conveyed clearly to students. The instructor recommended for the next run of the course that the intentions behind all activities and their importance for student learning should be added to activity description pages and that such details should be highlighted in the course syllabus. It was also noted that the syllabus should be placed in the course home page (besides Block 0), and it should be presented to students as a Web page instead of a PDF file as this would make it more accessible.

6. Limitations and Future Research

This research has several limitations that open opportunities for future research. First, although the proposed prediction models produced accurate predictions as demanded by the instructor, these predictions were not actually put into practice to create an intervention in the MOOC context studied. In a follow-up study, we plan to collaborate with the course instructor to replicate this work in the next run of the same MOOC and use the predictions to intervene the collaborative learning activity in real time. A possible intervention could be to support the group formation task for the collaborative tasks. Related with this future work, in collaboration with the teaching staff of the same MOOC, we plan to refine the LD of the course based on the revisions suggested in this paper, and then examine if the instructor's intentions would be better reflected in the learning activities of students. After this refinement, it would be interesting to see if the performance of the LD-driven prediction model improves significantly and surpasses the generic model.

Furthermore, instructor involvement in the loop was limited at some points. For example, the instructor was asked only to reflect on the importance of various course components; however, the teaching staff could be directly involved in the process of the feature selection. That is, instead of researchers choosing the predictive features based on the LD and the pedagogical decisions, instructors could themselves suggest features to predict the target student behaviour, which is already noted in the literature (Veeramachaneni et al., 2014). In our follow-up work, we plan to involve the instructors in feature selection through a comprehensive guidance. Enabling them to participate in such critical decisions could offer further implications regarding their role in the process of building the predictive model.

References

- Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., ... Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 329–338). Edinburgh, Scotland. <https://doi.org/10.1145/2883851.2883944>
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, 3(10), 27–39.
- Boroujeni, M. S., & Dillenbourg, P. (2018). Discovery and temporal analysis of latent study patterns from MOOC interaction sequences. In *Proceedings of Eighth International Conference on Learning Analytics and Knowledge* (pp. 206–215). Sydney, Australia. <https://doi.org/10.1145/3170358.3170388>
- Bote-Lorenzo, M. L., & Gómez-Sánchez, E. (2017). Predicting the decrease of engagement indicators in a MOOC. In *Proceedings of Seventh International Conference on Learning Analytics and Knowledge* (pp. 143–147). Vancouver, Canada. <https://doi.org/10.1145/3027385.3027387>
- Bote-Lorenzo, M. L., & Gómez-Sánchez, E. (2018). An approach to build in situ models for the prediction of the decrease of academic engagement indicators in Massive Open Online Courses. *Journal of Universal Computer Science*, 24(8), 1052–1071.
- Boyer, S., & Veeramachaneni, K. (2016). Robust predictive models on MOOCs: Transferring knowledge across courses. In *Proceedings of the Ninth International Conference on Educational Data Mining* (pp. 298–305). Raleigh, NC, USA.
- Brooks, C., & Thompson, C. (2017). Predictive Modelling in Teaching and Learning. In C. Lang, A. Wise, & D. Gasevic (Eds.), *Handbook of Learning Analytics* (pp. 61–68). <https://doi.org/10.18608/hla17.005>
- Er, E., Bote-Lorenzo, M. L., Gómez-Sánchez, E., Dimitriadis, Y., & Asensio-Pérez, J. I. (2017). Predicting student participation in peer reviews in MOOCs. In *Proceedings of*

- the Second European MOOCs Stakeholder Summit 2017*. Madrid, Spain.
- Eradze, M., Rodríguez-Triana, M. J., & Laanpere, M. (2017). Semantically annotated lesson observation data in learning analytics datasets: A reference model. *Interaction Design and Architecture(s) Journal*, *33*, 75–91.
- Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127–203. <https://doi.org/10.1007/s11257-018-9203-z>
- Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education*, *28*, 68–84. <https://doi.org/10.1016/j.iheduc.2015.10.002>
- Gašević, D., Mirriahi, N., Dawson, S., & Joksimović, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. *Computers in Human Behavior*, *67*, 207–220. <https://doi.org/10.1016/j.chb.2016.10.026>
- Greene, J. C. (2007). *Mixed Methods in Social Inquiry*. San Francisco, CA: Jossey-Bass.
- Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. In *Proceedings of the Second European MOOC Stakeholder Summit* (pp. 58–65). Lausanne, Switzerland.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36.
- Harrer, A., & Göhnert, T. (2015). Integrated representations and small data: Towards contextualized and embedded analytics tools for learners. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 406–407). Poughkeepsie, NY, USA. <https://doi.org/10.1145/2723576.2723665>
- Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *Proceedings of the Twelfth European Conference on Technology-Enhanced Learning* (pp. 82–96). <https://doi.org/10.1007/978-3-319-66610-5>
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, *104*, 18–33. <https://doi.org/http://dx.doi.org/10.1016/j.compedu.2016.10.001>
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of Third International Conference on Learning Analytics and Knowledge* (pp. 170–179). Leuven, Belgium. <https://doi.org/10.1145/2460296.2460330>
- Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity*, *43*(2), 265–275.
- Liu, R., & Koedinger, K. R. (2017). Going beyond better data prediction to create explanatory models of educational data. In C. Lang, A. Wise, & D. Gasevic (Eds.), *Handbook of Learning Analytics* (pp. 69–76). <https://doi.org/10.18608/hla17.006>
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439–1459. <https://doi.org/10.1177/0002764213479367>
- Mangaroska, K., & Giannakos, M. (2017). Learning analytics for learning design: Towards evidence-driven decisions to enhance learning. In *Proceedings of the Twelfth European Conference on Technology Enhanced Learning* (pp. 428–433). <https://doi.org/10.1007/978-3-319-66610-5>
- Manolis, M., & Karkalasi, S. (2017). Reflective analytics for interactive e-books. *Interaction Design and Architecture(s) Journal*, *33*, 33–53. Retrieved from http://www.mifav.uniroma2.it/inevent/events/idea2010/doc/33_2.pdf
- Marshall, C., & Rossman, G. B. (1999). *Designing qualitative research* (3rd ed.). Thousand Oaks, CA, US.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society*, *72*(4), 417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>

- Mor, Y., Ferguson, R., & Wasson, B. (2015). Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology*, 46(2), 221–229. <https://doi.org/10.1111/bjet.12273>
- 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). <https://doi.org/10.1109/TLT.2018.2856808>
- Nguyen, Q., Rienties, B., & Toetenel, L. (2017). Mixing and matching learning design and learning analytics. In *Proceedings of Forth Conference on Learning and Collaboration Technologies* (pp. 1–15). <https://doi.org/10.1007/978-3-319-58515-4>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2012). Scikit-learn: Machine learning in Python, 12, 2825–2830. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, 29(5), 615–620. <https://doi.org/10.1007/s10979-005-6832-7>
- Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*, 46(2), 330–343. <https://doi.org/10.1111/bjet.12198>
- Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., & Asensio-Pérez, J. I. (2017). The teacher in the loop: Customizing multimodal Learning Analytics for blended learning. In *Cross Multimodal Learning Analytics Workshop*. Tallinn.
- Schmitz, M., Limbeek, E. Van, Greller, W., Sloep, P., & Drachsler, H. (2017). Opportunities and Challenges in Using Learning Analytics in Learning Design. In *Proceedings of the Twelfth European Conference on Technology Enhanced Learning* (pp. 209–223). Tallinn. <https://doi.org/10.1007/978-3-319-66610-5>
- Sergis, S., & Sampson, D. G. (2017). Teaching and learning analytics to support teacher inquiry: A systematic literature review. In A. Peña-Ayala (Ed.), *Learning Analytics: Fundamentals, Applications, and Trends. Studies in Systems, Decision and Control* (Vol. 94, pp. 25–63). Springer, Cham. <https://doi.org/10.1007/978-3-319-52977-6>
- Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Identifying styles and paths toward success in MOOCs. In *Proceedings of the Eighth International Conference on Educational Data Mining* (pp. 408–411). Madrid, Spain.
- Taylor, C., Veeramachaneni, K., & O'Reilly, U.-M. (2014). Likely to stop? Predicting dropout in massive open online courses. *ArXiv*. Retrieved from <http://arxiv.org/abs/1408.3382>
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246. <https://doi.org/10.1177/1098214005283748>
- Tseng, S.-F., Tsao, Y.-W., Yu, L.-C., Chan, C.-L., & Lai, K. R. (2016). Who will pass? Analyzing learner behaviors in MOOCs. *Research & Practice in Technology Enhanced Learning*, 11(8), 1–11. <https://doi.org/10.1186/s41039-016-0033-5>
- Veeramachaneni, K., O'Reilly, U.-M., & Taylor, C. (2014). Towards feature engineering at scale for data from massive open online courses. *ArXiv*.
- Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017a). Delving deeper into MOOC student dropout prediction. *ArXiv*, 21–27. <https://doi.org/10.1145/1235>
- Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017b). MOOC dropout prediction: How to measure accuracy? In *Proceedings of the Fourth ACM Conference on Learning@Scale* (pp. 161–164). <https://doi.org/10.1145/3051457.3053974>
- Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 151–160). Society for Learning Analytics Research. <https://doi.org/10.18608/hla17.013>