Prediction of Students' Assessment Readiness in Online Learning Environments: The Sequence Matters

Donia Malekian
School of Computing and Information Systems
The University of Melbourne
Melbourne, Victoria, Australia
malekiand@unimelb.edu.au

James Bailey
School of Computing and Information Systems
The University of Melbourne
Melbourne, Victoria, Australia
baileyj@ unimelb.edu.au

Gregor Kennedy
Melbourne Centre for the Study of Higher Education,
The University of Melbourne
Melbourne, Victoria, Australia
gek@ unimelb.edu.au

ABSTRACT

Online learning environments are now pervasive in higher education. While not exclusively the case, in these environments, there is often modest teacher presence, and students are provided with access to a range of learning, assessment, and support materials. This places pressure on their study skills, including self-regulation. In this context, students may access assessment material without being fully prepared. This may result in limited success and, in turn, raise a significant risk of disengagement. Therefore, if the prediction of students' assessment readiness was possible, it could be used to assist educators or online learning environments to postpone assessment tasks until students were deemed "ready". In this study, we employed a range of machine learning techniques with aggregated and sequential representations of students' behaviour in a Massive Open Online Course (MOOC), to predict their readiness for assessment tasks. Based on our results, it was possible to successfully predict students' readiness for assessment tasks, particularly if the sequential aspects of behaviour were represented in the model. Additionally, we used sequential pattern mining to investigate which sequences of behaviour differed between high or low level of performance in assessments. We found that a high level of performance had the most sequences related to viewing and reviewing the lecture materials, whereas a low level of performance had the most sequences related to successive failed submissions for an assessment. Based on the findings, implications for supporting specific behaviours to improve learning in online environments are discussed.

CCS CONCEPTS

• **Applied computing** → Education • **Applied computing** → E-learning • **Applied computing** → Distance learning

LAK '20, March 23–27, 2020, Frankfurt, Germany
© 2020 Association for Computing Machinery. ACM ISBN 978-1-4503-7712-6/20/03...$15.00
https://doi.org/10.1145/3375462.3375468

KEYWORDS

Learning Analytics, LSTM, Assessment Readiness Prediction, Sequential Pattern Mining, MOOCs

ACM Reference format:

D. Malekian, J. Bailey and G. Kennedy. 2020. Prediction of Students' Assessment Readiness in Online Learning Environments: The Sequence Matters. In *Proceedings of the International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 2020 (LAK'20), 10 pages.* https://doi.org/10.1145/3375462.3375468

1 Introduction

Online learning environments (e.g., online courses, blended learning environments, MOOCs) are now pervasive in higher education. Even though there are many advantages of online learning, the scale of these environments, both in mainstream higher education and in new online environments such as MOOCs, present educators with challenges, such as providing the necessary supervision to students and the practical ability of instructors to guide and mentor students through courses. In many digital learning environments, students are often relatively solitary learners, and their success relies more heavily on their abilities and skills in self-directed learning. Moreover, in digital learning environments such as MOOCs, open learning structures are often emphasised in which students are offered unlimited access to different instructional materials, such as content-related resources (e.g., video, lectures), assessments, and collaboration tools [18]. This may well make these learning environments more challenging for students. There is a distinct danger that students may have difficulty adequately covering the learning material and, as such, may subsequently be ill-prepared for both developmental (formative) and summative assessment tasks.

Assessment tasks often form a critical component of digital learning environments, as they allow students to reflect on their own level of understanding and provide a mechanism to determine whether or not students have mastery of the course materials [22]. Educational research has shown that when students are unsuccessful in assessment, this can have a negative impact on their confidence in their ability to succeed, which in turn can lead to disengagement [18]. This argument has also been confirmed in a study based on five MOOCs [31], where students who received significantly lower grades in assessments, were likely to disengage from the course (e.g., drop-out the course, or become auditors). Thus, monitoring and predicting students' performance on assessment tasks could provide an opportunity for developing an adaptive learning environment aware of students' progress and readiness for online assessment tasks, and potentially delay or postpone tasks in cases where students are seen as not ready. Moreover, a deeper understanding of students' behaviour before completing an assessment and knowing whether these behaviours differ for students with a high or low level of performance, could be used to tailor support and interventions that could maximise the likelihood of student success.

The use of learning analytics allows us to model students' behaviour throughout a course in order to *predict* and *describe* their specific characteristics. This area is known as *Student Modelling* and has been long described as a key area of learning analytics [6], particularly in MOOCs environment, which is the focus of our study.

1.1 Student Modelling: Prediction

During recent decades, the emergence of MOOCs and their known challenges (i.e. lack of direct supervision, low completion rates) has led to many studies that have focused on prediction of a particular variable describing students such as their final achievement or drop-out [15, 24]. The way that the prediction task is formulated has a considerable impact on the results. Many choices exist regarding the representation of students' historical data with regards to different granularities and different forms in various contexts. Moreover, numerous decisions could be made regarding the choice of the prediction model.

Early research that has focused on modelling students, has found frequency measures of students participation in MOOCs [2, 21, 24], demographic factors [17], or a combination of these [20, 38] can act as early predictors of student drop-out. Although there are many ways to represent learning behaviour, most studies have represented behaviour based on simple frequency measures of student participation in the course such as the number of sessions, video watched and access to discussion forums [24]. However, a sequential representation of students' learning behaviour might be useful in showing how the process of students' learning over time impacts on other key learning processes and outcomes.

A recent direction of research has focused on building machine learning models based on students' historical data over time in order to predict their learning outcome in exercises as they complete MOOC [27, 30]. Among the existing studies, Pardos et al. [29], employed Bayesian algorithms to model students' knowledge over time based on the past history of quiz responses. They showed it is possible to predict with good accuracy whether or not each student has acquired the related knowledge during the course according to the result of prediction that is whether or not the quiz would be answered correctly. In a similar work by Yudelson et al. [41], a variant additive factors algorithm was used to model students' knowledge. They determined acquirable knowledge using a domain model, which is the vocabulary of skills for the Java programming language. They demonstrated that their model is accurate for both modelling student knowledge and suggesting the concepts that students can address in their code. In another study, Piech et al. [30], used Recurrent Neural Networks (RNNs) to model students' knowledge based on their performance on previously answered quizzes. Each time a student began a quiz, they predicted whether or not the quiz would be answered correctly. The historic sequences of correct and incorrect quiz responses determined the student's gained knowledge. They created a graph of conditional influence between quiz concepts that can be used for curriculum design. Further, the effectiveness of RNNs was confirmed in a very similar study [27]. They utilised an RNN with a similar procedure as [30], but for students with different learning abilities separately. Overall, the focus of these approaches is on modelling students based on the concepts that they learned and the result of prediction was considered as an indication of students' knowledge development and mastery on a topic. However, success in assessments is not just about the content. It also requires appropriate skills to access and work with the most relevant content. The current focus of educational research in this area is on strategies that enable effective learning. What remains understudied is evaluating how effective students' preparation behaviours are when it comes to their success on assessment tasks. Such analysis may be helpful to automatically delay or postpone tasks when students are not adequately prepared.

Another stream of research has applied and compared various predictive models to improve educational prediction. For instance, Al-Shabandar et al. [1], utilised various machine learning algorithms to model students' online behaviour based on the frequency measures of their participation in a MOOC and found the random forest to be the best prediction model. A similar study by Chen et al. [12], compared various classifiers and found XGBoost as the most reliable model. While comparing and selecting an appropriate algorithm is a challenging task, switching between algorithms usually results in a minimal change (for more discussion see [5]), whereas a proper representation of data can lead to a considerable difference and can be more challenging. To the best of our knowledge, comparing the effect of modelling students based on the sequential and aggregated representations of the same activities in their behaviour within MOOCs has not yet been investigated. Such analyses could reveal the importance of considering the temporal nature of students' behaviour for the application of student modelling.

Prediction of students' learning outcome could result in a valuable set of information for academic advisors, who can then communicate directly with students. However, these approaches usually do not provide actionable metrics regarding students' online data traces. Thus, in addition to prediction, identifying important factors capable of predicting students' learning outcome offers an excellent opportunity to support students by providing them with early feedback.

1.2 Student Modelling: Structure and Relationship Discovery

There has been considerable focus on the use of learning analytics approaches to profile students' behaviours and determine factors that differentiate them based on different levels of learning outcome. Recent studies have focused on factors in which educational data mining and educational psychology are combined. In this area, the positive effect of specific learning behaviours on students' achievement has been widely studied using learning analytics approaches [19, 40], particularly in MOOC environments.

The results of these studies mainly demonstrated that there were distinct student behaviours during learning, and there was a connection between those behaviours and students' learning outcome in the course. For instance, the positive effect of regular study and negative impact of the number of late submissions [40], direct association of re-reading pages and video watching in response to an incorrect assessment submission [19], or direct association of the number of video viewed and assessment submissions [7] and many others were found on learning outcome. Each factor in the mentioned studies was considered as a reflection of using specific learning behaviour. For instance, the regular study and late submissions were defined as a measure of time-management [40], re-visiting the previously visited pages as revision strategy [19], and the frequency measures of video hits and quiz attempts were found to be correlated with students' value beliefs and mastery approaches respectively [7].

The empirical investigations and modelling of students' learning behaviour in MOOCs have mostly focused on predefined aggregated factors and evaluating their effect on a students' learning outcome. However, utilising a data-driven approach for extracting the distinct students' learning behaviours (measures of them) might reveal novel behaviours which were not identified by domain experts. Additionally, considering the sequential representation of students' behaviour might disclose the effect of the process students used during learning. For instance, Brinton et al. [11], mined patterns from students' sequences of video-watching behaviour, with the aim of finding meaningful strategies rather than predefining them. They identified several video-watching characteristics correlated with a high chance of success in a quiz, such as repeatedly playing and pausing the video as an indicator of reflection strategy, and revising the previously watched video related to revision strategy. Sequential pattern mining can also be applied through other methods such as hidden markov model. As an example of this case, Coffrin et al. [14], investigated students' transitions between course materials in a MOOC and were able to visualize patterns of students' engagement based on their achievement. In these studies, mining patterns, instead of predefining them, could lead to the identification of previously unknown solutions to defined problems. In this paper, we utilise a sequential pattern mining approach to discover patterns from sequences of behaviours used by students to prepare for assessments.

1.3 Research Aim

The first goal of this study is to utilise learning analytic techniques to build a students' readiness profile (prediction model) to predict their performance in assessment tasks. Monitoring of students' knowledge based on the sequences of concepts they have learned – as determined by their performance in assessments – is well-studied in the literature. Our investigation focuses on the kinds of study behaviours students adopt in an online course as the foundation of a student model (rather than the concepts that have been learned). To create an accurate prediction model, we profile students' behaviours in two distinct forms, including aggregated and sequential. We employ a range of machine learning algorithms to determine whether a particular profile would be a stronger representative of students' readiness for assessment tasks.

The second goal of this study is to explore whether the most recent activities prior to submission are more influential on assessment results compared to incorporating the more historic activities applied by students. This analysis could reveal the importance of considering the overall process used by students throughout the course rather than the most recent activities prior to assessment submissions.

The last goal of this study is to explore students' assessment preparation behaviours using sequential pattern mining. We aim to explore and expose underlying factors that might influence students' assessment results. These factors could not only illustrate the different ways in which students prepare for assessment tasks but could also identify behaviours that need support to improve students' future preparation.

Overall, in this study, we model students' assessment preparation behaviour based on their fine-grain overt behaviour in a MOOC to investigate the following research questions (RQs):

RQ1. Can we develop useful prediction models for forecasting students' readiness for assessment tasks? Does it matter if the sequential nature of students' activities is considered in the model rather than aggregated measures?

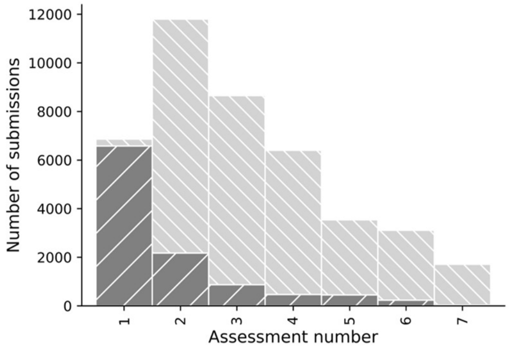
RQ2. What is the impact of considering the most recent activities prior to submission in the models compared to incorporating more historic activities applied by students?

RQ3. Are there differences in assessment preparation behaviours that lead to high or low performance?

2 Approach

2.1 Study Context

This study focuses on two offerings of a MOOC called Discrete Optimization (Disc Opt.) provided by the University of Melbourne in 2013 and 2014. This MOOC is a graduate-level course consisting of nine weeks of material presented in an open curriculum structure. It includes seven summative assessments, each of which contains a bank of questions. Students need to earn a passing grade (0.7/1) in each assessment to complete the course. The student's final outcome for an assessment is based on their best submission score for that assessment. The content of assessments is slightly different across both offerings. A summary of the two offerings and students' participation in assessment tasks are presented in Table 1. In the first offering of this course, of the 51,306 students who initially enrolled, around 12% (6,664) engaged in the assessments. From those, 39% (2610) dropped out of the course immediately after a failure in an assessment. We cannot be sure about the reason(s) for this abrupt disengagement, but one



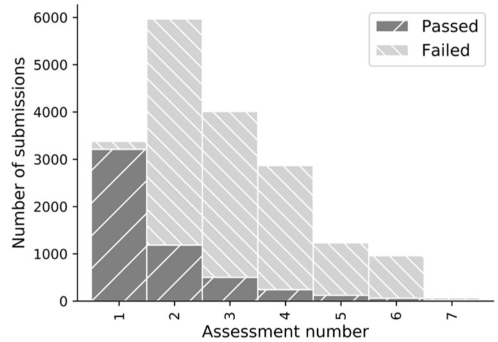


Figure 2: Number of passed and failed submissions in the a) 1st offering, and b) 2nd offering of Disc Opt.

Table 1: Summary of the two offerings of Disc Opt.

	1st offering	2nd offering
Number of students enrolled	51,306	33,975
Number of assessments	7	7
Number of students engaged in assessments	6,664	3,258
Number of students dropping-out after assessment failure	2610	332
Number of assessment submissions	42,094	18,495
Number of failed assessment submissions	31,285	13,159
Number of students completed	795	322

possible explanation is that failure experiences influenced some students to doubt their abilities to succeed, which then resulted in their disengagement [17, 30].

In this course, students were permitted to make an unlimited number of submissions for each assessment. In this paper, we have restricted our analysis to the first ten student's submissions per assessment. Moreover, for the successful submissions, we have limited our analysis on those that finalised the assessments' outcome. This means that if a student had multiple successful submissions on an assessment, only the first successful one was considered. Statistics summarising the total number of passed and failed submissions in each assessment of the two MOOCs are shown in Figure 2. We see that in both offerings, most submissions were passed for the first assessment (as it was considered to be an easy task). In contrast, for all the other assessments, less than half of the submissions were successful.

2.2 Data Analysis

A standard MOOC usually contains three different instructional events for students: content-related resources (e.g., video lectures, slides), assessments (e.g. assignments, and quizzes), and collaboration tools (e.g. discussion forums) [10]. The focus of our study is on students' fine-grain overt learning activities in response to these instructional events, particularly those that help students achieve course outcomes and reflect particular kinds of approaches to online study. This involves accessing and downloading lecture, re-accessing and re-downloading the previously accessed lecture, accessing the discussion forums, and submitting assessments. A sample representation of a student's list of ordered activities is shown in Figure 1. We refer to this figure throughout the paper as we develop our sequential analysis. For simplicity, the activity names are mapped to IDs as provided in the figure caption.

Following sub-sections describe the analysis undertaken to assist on answering our research questions.

2.2.1 RQ1. Assessment Readiness Prediction: Does the Sequence Matter. Using predictive models, we built a student profile based on their past history of learning activities, to determine which students were not ready for their assessment task. Once developed, this profile was then used to predict students' success for *each submission* they make on an assessment task. We cast this problem as a binary classification task, where a classifier was developed based on a student's past learning activities before submitting an assessment in two forms of aggregated and sequential, and the class label was whether or not a student's submission passes.

2.2.1.1 Aggregated Representation. Each submission s was described by a d-dimensional vector of features ($s \in \mathbb{R}^d$)

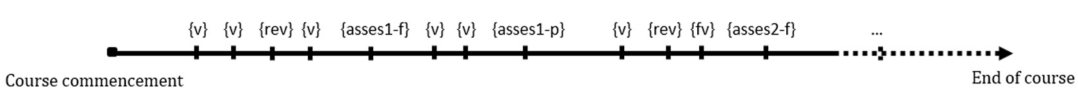


Figure 1: A sample representation of a student's ordered list of learning activities in a MOOC. v= lecture viewed or downloaded, rev= lecture reviewed or redownloaded, fv= forum viewed, asses#-p= passed submission on a specific assessment#, asses#-f= failed submission on a specific assessment#

Table 2: Sequences for each *assessment* and their respective outcome based on the example provided in Figure 1. casses-f= failed submission on the same assessment

Assessment#	Sequence before completing each assessment (assessment outcome)	Finalised Sequence	Assessment Outcome
1	v v rev v asses1-f v v (asses1-p)	v v rev v casses-f v v	passed
2	v rev fv (asses2-f)	v rev fv	failed

conducted by a student (as presented in Table 3), started from the beginning of the course and ended before that submission. We aimed to build our prediction profile irrespective of the specific assessment. Thus, for each submission, we divided the previous submissions into two groups of; submissions on the same assessment, and on the other assessments with their respective results as represented in Table 3.

Table 3: Aggregated profile for each student's submission

Aggregated profile
Number of lectures viewed or downloaded
Number of lectures reviewed or redownloaded
Number of forums viewed
Number of failed submissions on the same assessment
Number of passed submissions on the other assessments
Number of failed submissions on the other assessments

2.2.1.2 Sequential Representation. Each submission was described by a variable sequence of t ordered activities ($X_1, \ldots, X_{t-1}, X_t$) conducted by student started from the beginning of the course and ended before that submission. ($X_t \in$ Learning activities). Table 4 presents the extracted sequences for the example provided in Figure 1 for illustration purpose. As can be seen, each submission is assigned a sequence that starts from the beginning of the course and ends before that submission. Similar to the aggregated profile, we made a distinction between submissions on the "same assessment" and "other assessments". As can be seen in Table 4, we mapped them to "casses" and "oasses" respectively in the "Finalised Sequence" column.

We first evaluated the effectiveness of the aggregated representation of students' learning activities using a non-sequential machine learning classifier. Then the prediction power of a sequential machine learning model was examined to investigate the impact of modelling the sequential nature of students' behaviour. We used the first offering of the MOOC to train the model. Using this model, at each student's submission in the second offering of the MOOC, we predicted the probability of success to determine the effectiveness of our models.

2.2.1.3 Algorithms. We utilised Neural Network (NN) models for the evaluation purpose. NNs are popular for learning complex relationships from data in domains such as medicine [9, 36], security [37], loan applications [35], speech recognition [42], and natural language processing [23]. There are, thus far, a limited number of studies employing NNs for tasks in learning analytics. As an example, Okubo et al. [28], confirm the effectiveness of NNs for predicting students' final grade compared to regression-based models. In this paper, we explored the utility of the **aggregated** profile for assessment readiness prediction, employing a basic three-layer structure NN with all layers fully connected - a **Multiple Layer Perceptron (MLP)**. This is a simple proof of concept structure, and more complex structures (e.g. with more layers) are possible. the output was a prediction model learned for all the submissions. To investigate the impact of the **sequential** nature of students' activities on the prediction task, we used a sequential prediction model based on a popular variant of Recurrent Neural Networks - **Long Short-Term Memory (LSTM)**. An LSTM is capable of fitting patterns from data across both long- and short time scales. Our model used the sequential form of data as the input. The output was a prediction model learned for all the submissions.

For this analysis, N is the number of assessments in a course, and we had n submissions for each assessment $s_{i,}$ $1 \le i \le N$. Each assessment i was defined by a set of submissions, $s_{i,} = \{s_{i1}, s_{i2}, \ldots, s_{in}\}$. For each submission s_{ij}, a class label was defined as passed (y = 1) or failed (y = -1). Thus each assessment was given by $(s_i, y_i) = [(s_{i1}, y_{i1}), (s_{i2}, y_{i2}), \ldots, (s_{in}, y_{in})]$. The binary classification algorithm generated a probability estimate for each class by fitting a single model on all assessments' submissions, $\{(s_{11}, y_{11}), (s_{12}, y_{12}), \ldots, (s_{21}, y_{21}), (s_{22}, y_{22}), \ldots\}$.

2.2.1.4 Experimental Configurations. For both MLP and LSTM, we utilized binary cross-entropy as a loss function and trained models with mini-batch stochastic gradient descent. For determining the best parameters, grid search hyperparameter

Table 4: Sequences before each *submission* and their outcome (class label) based on the example provided in Figure 1. casses-f= failed submission on the same assessment, oasses-f= failed submission on the other assessments, oasses-p= passed submission on the other assessments

Assessment#	Submission#	Sequence before each submission (submission outcome)	Finalised Sequence	Submission outcome
1	1	v v rev v (asses1-f)	v v rev v	failed
1	2	v v rev v asses1-f v v (asses1-p)	v v rev v casses-f v v	passed
2	1	v v rev v asses1-f v v asses1-p v rev fv (asses2-f)	v v rev v oasses-f v v oasses-p v rev fv	failed

optimization [8] was used. In this technique, a set of parameter values for mini-batch size, the number of hidden dimensions (for MLP) and memory units (for LSTM) was defined. Then the model was trained using all combinations of those parameters and the solution that minimized the cost function was selected [8]. As there existed relationships between submissions, we considered the temporal order in which submissions were made when doing the cross-validation in parameter optimisation. For this purpose, we used the walk-forward cross-validation approach [26], in which multiple split points were selected in the ordered list of observations for determining the train and test data. We used early stopping on the number of epochs to avoid overfitting and unnecessary computations. For doing so, 20% of the training data at each fold of parameter optimisation was used as a validation set, and then the training in each fold stopped when the validation error started to increase. We implemented both networks using the Keras framework with TensorFlow™ back-end.

2.2.2 RQ2. The Most Recent Activities Prior to Submissions and The Assessment Outcomes. To answer the second research question, we built our models (MLP and LSTM) based on the variable length of recent activities before *submissions.* Having these models, we then compared their prediction performance to examine whether the most recent activities prior to submission would be more predictive in each of the algorithms rather than including the more historical activities.

2.2.3 RQ3. Differences in Successful and Unsuccessful Assessment Preparation Behaviours. To address this research question, we investigated students' behaviour and their approach to preparing for assessment tasks for each of the MOOC offerings. We assessed, through a sequence mining approach, whether students exhibited systematic behavioural patterns before *accomplishing assessment* tasks (before and in between submissions). Then we investigated whether these patterns differed for students with a high and low level of task outcome, and whether they can be mapped on to interpretable learning behaviours. Overall, our aim was to relate students' assessment preparation behaviour to their assessment outcomes. We defined *assessment outcome* and *behaviour* as below.

Assessment Outcome. This corresponds to the final result of an assessment (typically passed or failed) for a student. As mentioned before, for each assessment, students needed a mark of at least 70% to pass. Therefore, in our analysis, the pass mark for each assessment was considered to be 70% of the total assessment score.

Behaviour. Similar to the previous analysis, we focused on the sequential representation of students' learning activities. Such representation could reveal the process students used to be prepared for the task, rather than simple frequency measures examined in most studies. We defined behaviour as student's activities before the submission that finalised the outcome for an assessment task. In this analysis, we aim to explore students' behaviour before completing assessments. In fact, each student that engaged with an assessment (with one or more submissions) was assigned one sequence for that assessment. The beginning of the sequence was considered to be before the first submission for the same assessment and started with the latest submission for another assessment. If it was the first assessment attempted through all the course, the sequence started from the beginning of the course. Table 2 presents the extracted sequences for the example provided in Figure 1 for illustration purpose. As can be seen, a sequence started with the first activity after the other assessment submission and ended before the submission that finalised the outcome for the assessment. Since we aim to gain insight about the preparation behaviour irrespective of the specific assessment, for each sequence, we mapped the submissions on the "same assessment" and "other assessments" to "casses" and "oasses" respectively as shown in the "Finalised Sequence" column of Table 2.

Accordingly, we categorized these sequences into two groups: 1) sequences that ended with a passed submission (Passed group), and 2) sequences that ended with a failed submission (Failed group).

We employed a sequential pattern extraction method [4] to identify the distinct behavioural patterns for the sequences. Sequential pattern mining is an analytical technique to identify all common subsequences of n-item in a sequence dataset that the proportion of their occurrence to the total number of sequences (called support) is more than a threshold. In other words, for each subsequence of n-item, support is generated, which determines the proportion of sequences that have enacted that subsequence at least once.

In summary, for each of the course offerings, we extracted all the *subsequence of n-item* (we call them *n-item* for the simplicity) from every sequence. For instance the subsequences of 3-item from the finalized sequences presented in Table 2, are {<v v rev> <v rev v> <rev v casses-f> <v casses-f v> <casses-f v v> <v rev fv>}. Each *unique* n-item was then subject to the following analysis:

1. The support of each n-item within each group of sequences (Passed/Failed) was computed. This allows comparison across groups as the groups might have different numbers of sequences.
2. The most frequent n-items in at least one of the groups were retained as patterns for further investigation.
3. The statistical significance of each frequent pattern was computed, using a chi-square test [39] on the 2*2 contingency table (exists/not exists versus Passed/Failed). The chi-square test provides a measure of the statistical significance of the association between each pattern and the assessment outcome. A Bonferroni correction for multiple testing was applied to the result of the chi-square test to avoid Type I error.
4. The patterns that were highly associated with outcome were further explored to determine the direction of association with the assessment result. This was performed based on comparing the fraction of each pattern in the Failed and Passed groups of sequences. In this way, the top patterns associated with the successful outcome were considered to be those that were more frequent (higher support) in the Passed group compared to the Failed group, ordered by their significance of association with the assessment outcome (based on smallest p-values).

5. The top patterns in each group that were highly associated with learning outcome were then explored to determine whether they can be related to interpretable learning behaviours.

2.2.3.1 Comparison of the two MOOCs in Terms of Their Influential Patterns. We also compared the influential patterns between the two MOOCs based on their significant association with the assessment outcome. This comparison could reveal if the two MOOCs were similar in terms of their prominent patterns. For this purpose, we utilised the Spearman Rank-order correlation [16] that shows the association between the ranked value of two variables. In our case, the correlation between the patterns' rank in the two MOOCs was examined.

3 Results

The analytic approach described above was used with two offerings of a MOOC and is presented in three parts in this section.

3.1 RQ1. Assessment Readiness Prediction: Does the Sequence Matter

The first set of analyses report on the ability of our models to predict assessment readiness. We report our results using the two metrics of accuracy and area under the ROC curve (AUC), a widely used metric in machine learning which is informative for the scenario of imbalance between classes [25]. We trained our models based on the first offering of the MOOC (with three-fold cross-validation for parameter optimisation) and evaluated them based on the second offering of the course.

The outcome of the predictive performance of MLP and LSTM is shown in Table 5. We also investigated the utility of the aggregated profile based on a baseline model of Logistic Regression (LR) to make sure the result we obtained using MLP was not misspecified. The AUC and accuracy results of both MLP and LSTM algorithms demonstrate a high success rate in predicting student readiness for the second offering (AUC score > 85.0% and accuracy>83.7%). A comparison of the two models illustrates that LSTM outperformed MLP in both AUC and accuracy. This suggests that taking the sequential nature of data into account was more beneficial.

Table 5: Comparison of LR, MLP, and LSTM performance based on AUC and accuracy.

	LR	MLP	LSTM
Accuracy	83.3	83.7	**87.1**
AUC	84.1	85.0	**86.4**

3.2 RQ2. The Most Recent Activities Prior To Submissions and The Assessment Outcomes

We examined whether building our models based on the various numbers of recent activities before submission would make any difference in their prediction power. We limited this analysis to at most 77 recent activities. This number was the median number of activities done before submissions. Figure 3 presents the result of this analysis based on AUC. In this figure, the x-axis corresponds to the upper limit of the number of recent activities prior to the submissions that the models were built on. As can be seen, all algorithms worked better when considering more activities. Furthermore, LSTM worked better than MLP in almost all cases. This could mean that the overall processes students applied through the course had more impact on the assessment outcomes rather than only the recent activities prior to submissions.

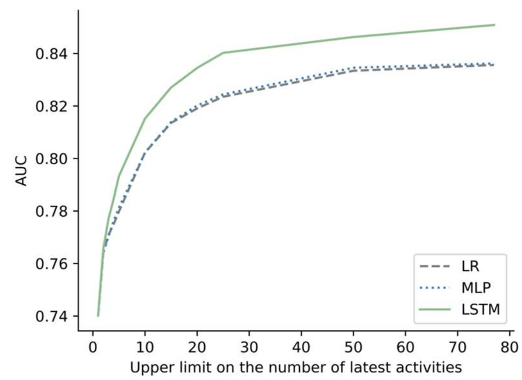


Figure 3: Comparison of LR, MLP, and LSTM based on AUC performance for the variable length of activities before submissions

3.3 RQ3. Differences in Successful and Unsuccessful Preparation Behaviours

It was possible to collect 15,748 and 7,544 student sequences for all the assessments of the first and the second offerings of the course respectively. Using sequential pattern extraction method, we extracted a large number of subsequences (n-items) of various length ($n \in \{1, 2, .., 10\}$) and support for each group of sequences in each of the offerings. Then for each offering, only the subsequences having more than 20% support in at least one of the groups were retained as potential patterns in order to contrast them across groups with different levels of the assessment outcome. We considered the maximum subsequence length (n) as 10 because there was no unique 10-item subsequence that had a support value of at least 20% in each of the two groups.

From all the subsequences, 95 and 84 number of n-items were found as frequent patterns from the first and second offerings of the MOOC respectively. Figure 4 reports the top-10 patterns with the most significant association (p-value <0.0005) with each of the passed and failed assessment outcome for each of the MOOCs, utilising chi-squares. The patterns are represented in the form of state transition diagrams to be able to present as much pattern as possible. The general representation of a pattern is illustrated in Figure 4(a). S and E indicate where a pattern started and ended. Each state illustrates an activity. The notation above each self-transition illustrates the [lower limit - upper limit] on the number of times that activity can be repeated. Thus, the general pattern starts with activity A, that can be repeated for n to m times; then

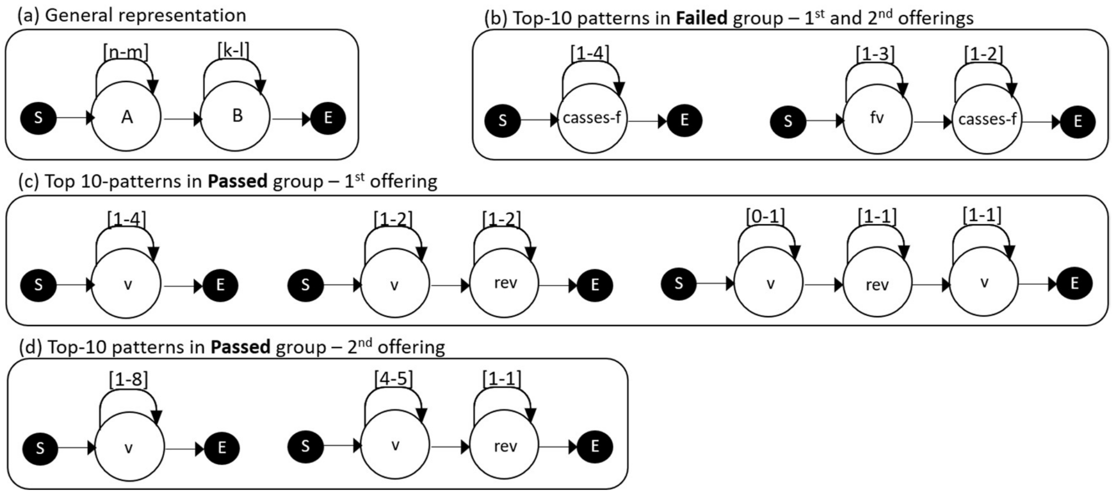


Figure 4: The top-10 significant patterns identified across each of the Failed and Passed groups of sequences for the 1st and 2nd offerings of Disc Opt. S and E indicate where a pattern starts and ends. Each state illustrates an activity. The notation above each self-transition illustrate the [lower limit – upper limit] on the number of times that activity can be repeated. v= lecture viewed or downloaded, rev= lecture reviewed or redownloaded, casses-f= failed submission on the same assessment, fv= forum viewed

we have activity B, which can be repeated for k to l times. For instance if we have [n,m,k,l] as [1,2,1,2], the pattern set would be equal to {(AB), (AAB), (ABB), (AABB)}.

Results revealed that sequences with failed outcome in both MOOCs had exactly the same top-10 patterns as presented in Figure 4(b). This includes multiple attempts in an assessment without doing any activity in between as well as exploring the discussion forum before multiple failed assessment submissions. Whereas, sequences with passed outcome in both offerings had the most sequences related to accessing the lecture material for one or multiple times and reviewing the already accessed content (Figure 4(c&d)). In this group of sequences (Passed group), for both offerings, there is a pattern showing that numerous lecture materials observed before submissions (The leftmost pattern for both offerings). Furthermore, there exists a pattern in which already viewed lecture content was reviewed after viewing the lectures multiple times (the second pattern from the left for both offerings). In the first offering, there exists a pattern in which some contents were accessed after reviewing the material (The rightmost pattern in the first offering).

3.3.1 Comparison of The Two MOOCs in Terms of Their Influential Patterns. Ranking in each MOOC was obtained by assigning a rank of 1 to the n-item with the highest association with the learning outcome, 2 to the next highest and so on. The ranking was calculated based on all the n-items rather than the frequent ones (patterns). The reason that we considered all n-items for the ranking was that there were few patterns in one of the offerings which were not frequent in the other. By ranking all the n-items, we obtained rank for every existing pattern even if it was not prevalent in one of the offerings. The result of this analysis revealed a strong positive correlation (correlation coefficient of 0.83) between the rank of significant patterns of the two MOOCs. This means the two MOOCs shared similar influential patterns.

4 Discussion

In this paper, we used a range of learning analytics techniques to examine whether we can predict students' readiness for assessments and to explore students' preparation behaviour for accomplishing these tasks.

For the first research question, we found that it was possible to predict student's readiness for assessment tasks based on modelling their fine-grain overt behaviour with a MOOC in two forms of aggregated and sequential. We built and evaluated our models based on the same course but different cohorts. Based on our results, students' sequence of task preparation activities was a more powerful predictor of assessment performance than the overall aggregated behaviour. This suggests that taking the sequential nature of students' interactions into account was more beneficial than the frequency measures. Monitoring and predicting students' performance on future tasks could provide an opportunity to support students in their learning path. Such analyses can be used to monitor students' progress and readiness over time, to delay or postpone tasks that students are not yet ready for. The result of our temporal analysis revealed the importance of taking the temporal nature of behaviour into account to have better models and understandings of student behaviours.

The second research question suggested that considering the more historical activities prior to submission led to the higher prediction power for the models. This supports the importance of the overall process students applied in the subject as a significant indicator of performance rather than only the most recent activities before submissions.

For the third research question, sequence mining was used to provide insight into significant differences in the assessment preparation behaviour for each of low and high-performance groups in assessments. Our results suggested viewing and reviewing the lecture content was an important factor for success, which is not surprising. This result is consistent with educational findings that reviewing learning material is beneficial in student learning [34]. As the study by Rohrer [34] contends, reviewing strategies can improve memory retention and lead to an improvement in learning. Other studies have also found that reviewing content is effective in algebra word problem solving [13], and advantageous in students' memory retention of reading content [33]. However, an interesting insight is that a considerable number of students tried to improve their performance by attempting the assessments multiple times without doing any activity in between their attempts. There was a significant negative association between these patterns and learning outcome. This may imply that simply making multiple submissions without applying any clear strategy does not lead to success. This result also confirms educational findings that too many consecutive errors undermine students' learning performance [32]. According to Ashcraft and Kirk [3], when an error is related to lack of knowledge and skill, it can lead to anxiety and negatively affect a student's learning. The patterns reported in this paper, appear to be a valuable source of information to better understand students' behaviour before completing assessment tasks and could be used to inform educators about the behaviours in need of support. Moreover, adaptive learning environments could be designed to steer students away from such behaviour, for instance, by placing a time constraint between successive attempts or limiting the possible number of attempts on an assessment.

Overall, a clear implication of the research presented in this paper is that the temporal dimension in modelling students' behaviour is important, and there may be a need to account for it in learning analytics research. In addition, the findings of pattern extraction analysis provide insights into how a framework for both educators and adaptive environments could be developed to identify which behaviours require support so that relevant interventions could be provided. However, this study has limitations commonly found in most MOOC studies. The first limitation relates to the generalizability of the analysis and interpretations of the patterns found. This study was based on different cohorts of one MOOC. Considering other MOOCs with different learning designs could provide better insight into students' assessment preparation behaviour and its impact on performance. Additionally, the MOOC that was the focus of this study was an optimization course, and the assessments required applied activities that expected students to put efforts in to achieve a good outcome. Our approach might not be as effective for other assessment types, such as when students are able to play with the system or guess the answers without effort. Another limitation is not examining the learners' personal factors such as skills, prior knowledge, demographic factors, goal and motivation throughout the course, as it was beyond the scope of this study.

5 Conclusion

In this paper, we used machine learning techniques to predict students' readiness for assessment tasks and target those at risk of disengagement as a result of assessment failure. This analysis could be useful in monitoring students' progress and postponing tasks that students are not yet ready for. Our results show the effectiveness of our prediction models in estimating student readiness, especially if we formulate the prediction problems using sequential models and focusing on the overall behaviour of students before submission (rather than the most recent activities). We also explored how students approached their assessments and what strategies they applied to be prepared for this task. The patterns we discovered suggest there were several behaviours which impacted on students' performance, like reviewing the learned content, and multiple consecutive failed attempts without mastering the topic. These results align with previous educational research. For future research, it would be interesting to determine whether these patterns and results are applicable to other MOOCs and online learning environments and to conduct a study assessing their effectiveness for feedback via a randomized controlled study.

REFERENCES

[1] Al-Shabandar, R. et al. 2017. Machine learning approaches to predict learning outcomes in Massive open online courses. In *Proc. IJCNN*, 713–720. https://doi.org/10.1109/IJCNN.2017.7965922

[2] Allen, G. et al. 2019. Using Educational Data Mining to Identify and Analyze Student Learning Strategies. In *Proc. HICSS*, 13-21. http://doi.org/10.24251/HICSS.2019.004

[3] Ashcraft, M.H. and Kirk, E.P. 2001. The relationships among working memory, math anxiety, and performance. *Journal of experimental psychology: General*. 130, 2 (2001), 224–237. http://dx.doi.org/10.1037/0096-3445.130.2.224

[4] Baker, R. and Yacef, K. 2009. The state of educational data mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*. 1, 1 (2009), 3–17. https://jedm.educationaldatamining.org

[5] Baker, R.S. and Inventado, P.S. 2014. Educational data mining and learning analytics. In J. A. Larusson, & B. White (Eds.), *Learning Analytics: from Research to Practice.* 61-75. New York, NY: Springer. https://doi.org/10.1002/9781118956588.ch16

[6] Bakhshinategh, B. et al. 2017. Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*. 23, 1 (2017), 537-553. https://doi.org/10.1007/s10639-017-9616-z

[7] de Barba, P.G. et al. 2016. The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning*. 32, 3 (2016), 218–231. https://doi.org/10.1111/jcal.12130

[8] Bergstra, J. and Bengio, Y. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*. 13, 1 (2012), 281–305. https://dl.acm.org/citation.cfm?id=2188395

[9] Bertolaccini, L. et al. 2017. An overview of the use of artificial neural networks in lung cancer research. *Journal of Thoracic Disease*. 9, 4 (2017), 924–931. https://doi.org/10.21037/jtd.2017.03.157

[10] Brinton, C.G. et al. 2015. Individualization for education at scale: MIIC design and preliminary evaluation. *IEEE Transactions on Learning Technologies*. 8, 1 (2015), 136–148. https://doi.org/10.1109/TLT.2014.2370635

[11] Brinton, C.G. et al. 2016. Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance. *IEEE Transactions on Signal Processing*. 64, 14 (2016), 3677–3692. https://doi.org/10.1109/TSP.2016.2546228

[12] Chen, W. et al. 2018. Early detection prediction of learning outcomes in online short-courses via learning behaviors. *IEEE Transactions on Learning Technologies.* 12, 1 (2018), 44–58. http://doi.org/10.1109/TLT.2018.2793193

[13] Christianson, K. et al. 2012. Practice makes (nearly) perfect: Solving 'students-and-professors'-type algebra word problems. *Applied Cognitive Psychology.* 26, 5 (2012), 810–822. https://doi.org/10.1002/acp.2863

[14] Coffrin, C. et al. 2014. Visualizing Patterns of Student Engagement and Performance in MOOCs. In *Proc. LAK,* https://doi.org/10.1145/2567574.2567586

[15] Corbett, A.T. and Anderson, J.R. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction. 4, 4 (1994), 253–278. https://doi.org/10.1007/BF01099821

[16] Daniel, W.W. 1990. Spearman rank correlation coefficient. *Applied Nonparametric Statistics* (2nd ed.). 358–365. Boston, MA: PWS-Kent. https://lib.ugent.be/en/catalog/rug01:000283035

[17] DeBoer, J. et al. 2013. Diversity in MOOC students' backgrounds and behaviors in relationship to performance in 6.002 x. In *Proc. LINC,* 16-19, https://linc.mit.edu/linc2013/proceedings/Session2

[18] Dembo, M.H. and Seli, H. 2012. *Motivation and learning strategies for college success: A focus on self-regulated learning (4th ed.).* USA: Routledge, New York. https://routledgetextbooks.com/textbooks/_author/dembo-9780415894203/

[19] Diana, N. et al. 2016. Extracting measures of active learning and student self-regulated learning strategies from MOOC data. In *Proc. EDM*, 583–585. http://www.educationaldatamining.org/EDM2016/proceedings

[20] Fernandes, E. et al. 2019. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research.* 94, (2019), 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

[21] Gardner, J. et al. 2019. Evaluating the fairness of predictive student models through slicing analysis. In *Proc. LAK*, 225–234. https://doi.org/10.1145/3303772.3303791

[22] Gibbs, G. and Simpson, C. 2005. Conditions under which assessment supports students' learning. *Learning and teaching in higher education.* 1 (2005), 3–31. https://doi.org/10.1007/978-3-8348-9837-1

[23] Goldberg, Y. 2017. Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies.* 10, 1 (2017), 1–309. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

[24] He, J. et al. 2015. Identifying at-risk students in massive open online courses. In *Proc. AAAI*, 1749–1755. http://dl.acm.org/citation.cfm?id=2886521.2886563

[25] Ling, C.X. et al. 2003. AUC: a statistically consistent and more discriminating measure than accuracy. In *Proc. IJCAI*, 519–524. http://dl.acm.org/citation.cfm?id=1630659.1630736

[26] Lopez de Prado, M. 2018. *Advances in Financial Machine Learning*. Hoboken, NJ: John Wiley & Sons. https://www.wiley.com/9781119482086

[27] Minn, S. et al. 2018. Deep knowledge tracing and dynamic student classification for knowledge tracing. *In Proc. ICDM*, 1182–1187. https://doi.org/10.1109/icdm.2018.00156

[28] Okubo, F. et al. 2017. A neural network approach for students' performance prediction. In *Proc. LAK*, 598–599. https://doi.org/10.1145/3027385.3029479

[29] Pardos, Z. et al. 2013. Adapting bayesian knowledge tracing to a massive open online course in edx. In *Proc. EDM*, 137–144. https://doi.org/10.1.1.668.6548

[30] Piech, C. et al. 2015. Deep knowledge tracing. In *Proc. NIPS*, 505–513. https://papers.nips.cc/paper/5654-deep-knowledge-tracing

[31] Poquet, O. et al. 2015. Reconsidering Retention in MOOCs: the Relevance of Formal Assessment and Pedagogy. In Proc. *EMOOCs*, (2015). https://doi.org/10.13140/RG.2.1.1881.3286

[32] Rodrigo, M.M.T. and Beck, J.B. 2014. Understanding wheel spinning in the context of affective factors. *International Journal of People-Oriented Programming (IJPOP).* 3, 2 (2014), 75–90. https://doi.org/10.4018/IJPOP.2014070104

[33] Roediger, H.L. and Karpicke, J.D. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science.* 17, 3 (2006), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

[34] Rohrer, D. 2015. Student instruction should be distributed over long time periods. *Educational Psychology Review.* 27, 4 (2015), 635–643. https://doi.org/10.1007/s10648-015-9332-4

[35] Saha, S. and Waheed, S. 2017. Credit risk of bank customers can be predicted from customer's attribute using neural network. *International Journal of Computer Applications.* 161, 3 (2017), 39–43. https://doi.org/10.5120/ijca2017913170

[36] Shaikhina, T. and Khovanova, N.A. 2017. Handling limited datasets with neural networks in medical applications: A small-data approach. *Artificial Intelligence in Medicine.* 75, (2017), 51–63. https://doi.org/10.1016/j.artmed.2016.12.003

[37] Shone, N. et al. 2018. A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence.* 2, 1 (2018), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

[38] Umer, R. et al. 2017. On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning*. 10, 2 (2017), 160–176. https://doi.org/10.1108/JRIT-09-2017-0022

[39] William G, C. 1952. The chi-square test of goodness of fit. *The Annals of Mathematical Statistics.* 23, 3 (1952), 315–345. https://doi.org/10.1214/aoms/1177729380

[40] 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. https://doi.org/10.1016/j.iheduc.2015.11.003

[41] Yudelson, M. et al. 2014. Investigating automated student modeling in a Java MOOC. In *Proc. EDM.* http://educationaldatamining.org/EDM2014

[42] Zhang, Y. et al. 2017. Very deep convolutional networks for end-to-end speech recognition. In *Proc. ICASSP*, 4845–4849. https://doi.org/10.1109/ICASSP.2017.7953077