

Comparing the Factors that Predict Completion and Grades among For-Credit and Open/MOOC students in Online Learning

Ma. Victoria Almeda¹, Joshua Zuech², Ryan S. Baker³, Christopher Utz², Greg Higgins²,
Rob Reynolds²

¹ Teachers College, Columbia University

² NextThought

³ University of Pennsylvania

Abstract

Online education continues to become an increasingly prominent part of higher education, but many students struggle in distance courses. For this reason, there has been considerable interest in predicting which students will succeed in online courses, achieving poor grades or dropping out prior to course completion. Effective intervention depends on understanding which students are at-risk in terms of actionable factors, and behavior within an online course is one key potential factor for intervention.

In recent years, many have suggested that Massive Online Open Courses (MOOCs) are a particularly useful place to conduct research into behavior and interventions, given both their size and the relatively low consequences/costs of experimentation. However, it is not yet clear whether the same factors are associated with student success in Open courses such as MOOCs as in For-Credit course -- an important consideration before transferring research results between these two contexts. While there has been considerable research in each context, differences between course design and population limit our ability to know how broadly findings generalize; differences between studies may have nothing to do with whether students are taking a course For-Credit or as a MOOC.

Do learners behave the same way in MOOCs and For-Credit courses? Are the implications for learning different, even for the exact same behaviors? In this paper, we study these issues through developing models that predict student course success from online interactions, in an online learning platform that caters to both distinct student groups (i.e., students who enroll on a For-Credit or a non-credit basis). Our findings indicate that our models perform well enough to predict students' course grades for new students across both of our populations. Furthermore, models trained on one of the two populations were able to generalize to new students in the other student population. We find that features related to comments were good predictors of student grade for both groups. Models generated from this research can now be used by instructors and course designers to identify at-risk students both For-Credit and MOOC learners, towards providing both groups with better support.

Introduction

Online learning has become increasingly prevalent in higher education, and it is important for educators and researchers to address the challenges associated with this learning setting. In recent years, the research on online learning has diverged into two subcategories, with researchers focusing on the efficacy of For-Credit courses and MOOCs. While there are plenty of studies on either For-Credit courses or MOOCs, little has been done investigating how findings from one subcategory can help inform the other. Bridging this gap would be valuable, as Massive Online Open Courses (MOOCs) are a particularly useful place to conduct research into behavior and interventions, given both their size and the relatively low consequences/costs of experimentation. At the same time, the different goals held by MOOC learners, many of whom never intend to complete the course they start (Koller et al., 2013), may result in different patterns of behavior being associated with course completion, a risk to researchers attempting to translate findings between courses.

The present paper asks whether MOOC findings are applicable to For-Credit students as well by studying student success across two distinct populations (i.e., students who enroll on a For-Credit and non-credit basis) who are engaging with the exact same course materials and instructor. By controlling for these key aspects of the learning experience, we can avoid confounds and focus on the differences (or similarities) between For-Credit and MOOC learners, comparing online behaviors and student outcomes between these two student groups. Specifically, we develop classification and regressions models that predict student course completion in an online learning platform for each of our distinct populations. We then study whether the same models and corresponding features apply across contexts.

In the following sections of our paper, we examine recent research on MOOCs and For-Credit courses, emphasizing the gap in the literature about whether findings are comparable to each other across these contexts. The Methods section discusses the datasets we use in our analysis as well as the course platform, and the algorithms applied to predict student success in our For-Credit and Open courses. In the Results section, findings from the classification and regression models for each of the student populations are given, followed by results on the generalizability of these models. Lastly, in the Discussion section, we explore the implications of our findings to help instructors and researchers of different populations identify and support struggling online learners.

Literature Review

Many students struggle in distance and online courses. There is increasing evidence that students enrolled in online courses have higher drop rates (Diaz, 2002; Tyler-Smith, 2008) than in traditional classroom settings where instructors can offer support through face-to-face interactions. The overall low completion rates for online courses raise concerns about eventual student success.

As instructors have less opportunities to directly interact with students enrolled in online courses (Beard, 2002; Zhang, 2006), it becomes more difficult for instructors to determine whether all students are on a path to success. One approach that can support students is to use learning analytics. Several previous projects have used analytics to predict whether a student will pass or fail an online course or program. Early projects focused on demographic data and

assignment performance (Zafra & Ventura, 2009; Arnold et al., 2012); however, these features are predictive, but are not always actionable.

As a result, many researchers have begun to use more granular features of student interaction to predict student achievement in online courses. In one example, Whitmer (2012) found a significant relationship between student usage of a learning management system (LMS) and their achievement. Specifically, they found that participation in online assessment activities such as completing auto-graded quizzes had the largest effect in terms of prediction of final grades. In another example, Romero and colleagues (2013) predicted students' final performance on a course from participation in online discussion forums. They found that students who passed the course actively participated in the forum, posting messages more frequently and writing messages of higher quality as measured by the instructor's evaluation. In a third example, Andergassen et al. (2014) found that the duration of participation in online activities (e.g., the total amount of time spent between the first and last assignments of the course) significantly predicted students' final exam scores. In a fourth example, Baker and colleagues (2015) discovered that students with higher grades were more likely to access the course e-book early in the semester, with access of course materials even as early as before the semester started being predictive of eventual grade.

In parallel to this research within For-Credit courses, there has been increasing work in determining whether students will complete a Massive Online Open Course (MOOC), where completion rates are typically lower than in For-Credit courses (in part because many learners enter MOOCs without the goal of completing at all - Koller et al., 2013), and grades do not matter as much as whether students complete the course with the certificate or not. For example, researchers have investigated whether course completion can be predicted by posting behavior and social network participation on course discussion forums (Yang et al., 2013, Jiang et al., 2014). Students were less likely to drop out when their posts spanned a longer duration between the first and the last post of the current week – suggesting positive impacts with from more continual participation (Yang et al., 2013). Students were also less likely to drop out when they engaged with a greater number of students in online discussions. Similarly, Jiang et al. (2014) predicted whether or not students will earn a certificate of normal completion (basic track) or distinction (scholarly track) from their posting behavior in course discussion forums. Students were more likely to earn a certificate of distinction than normal completion when they actively participated in the forum in the first week. In another example, Morris and colleagues (2015) predicted completion rates based on student demographics. Higher completion rates were associated with higher prior educational attainment and prior online experience.

Of course, in considering this work, it is important to note that most of the work on student success in MOOCs has focused on completion with a certificate. Using completion rates to indicate student success in a MOOC course is by no means a perfect nor comprehensive criterion. In particular, some have argued that using completion as the metric of student success in Open courses obfuscates the various reasons students enroll in a MOOC. Whereas traditional courses take considerable effort to enroll in, cost money, and typically have penalties of various sorts for withdrawing, MOOCs are easy and free to enroll in, and have no penalty for withdrawal. As such, it has been argued that many students take MOOCs with more of a goal of sampling or learning specific material (Clow, 2013). However, recent studies suggest that many students who enroll in a MOOC on a non-credit basis have expectations similar to those seen in a traditional course, demonstrating high and sustained levels of engagement. Specifically, Shrader and colleagues (2017) found that some non-credit learners watch a considerable percentage of

videos and earn high points on quizzes, which are arguably potential indicators of a desire to pass the course. In a study of nine HarvardX courses, pre-course surveys indicate that, on average, 58% intend to finish the course, 25% intend to audit, 3% intend to browse, and the remaining 14% are uncertain of their goal. Even with many learners having goals other than completion, completion data still offer valuable information to instructors and researchers. Completing learners appear to behave differently than other types of disengaged learners, such as “sampling” learners (Clow, 2013) who only access content materials at the beginning of the course. For instance, completing learners post in discussion boards and access in-video assessments significantly more frequently than sampling learners (Kizilcec et al, 2013). Additionally, Wang, Paquette, and Baker (2017) indicate that course completion can lead to eventual career advancement, with completers being more likely to submit scientific papers in the field they were studying as well as being more likely to join scientific societies aligned with the topic of the course.

As can be seen in two bodies of literature (For-Credit and MOOCs), many of the same types of features are used when studying MOOCs and For-Credit online courses, such as discussion forum participation and use of course resources. However, there has been relatively little work directly comparing For-Credit learners and MOOC learners, with each individual study focusing on predicting student achievement within a specific student population – either among students who enroll for a For-Credit or a non-credit (Open/MOOC) basis. Simply comparing between papers is not sufficient to fully compare these types of environments. While there have been several studies conducted within each of these groups, the differences between the courses and learning environments studied have made it difficult to compare between papers. As such, there remains a gap in the literature on how students’ online interactions correspond to learning outcomes in these student populations. Do the same factors predict success in each of these contexts?

This research question becomes particularly scientifically interesting given recent research that suggests there are important differences between the behaviors of students who enroll in For-Credit online courses versus Open online courses/MOOCs. For instance, forum participation appears to be relatively more sustained in For-Credit courses (Nistor & Neubauer, 2010) than in MOOCs (Clow, 2013). As such, there appear to be **varying patterns** of engagement between For-Credit and MOOC learners, but it is not yet conclusive whether these differences are due to the differences between these learners, or for example, other features of these courses or platforms investigated across these studies. Previous studies on edX, one of the largest platforms for Open courses, consistently show low completion rates – averaging fewer than 10% of participants (Breslow et al., 2014; Ho et al., 2014). As such, it would not be surprising to find lower completion rates for Open courses than For-Credit courses. And it is possible that **behaviors** associated with successful performance (e.g. good grades) will differ as well. However, there has not yet been work to compare the potential differences in behavior between For-Credit and Open course learners, and the relationship between these behaviors and student outcomes, when controlling for factors such as platform and course design and topic. Thus, investigating which factors predict student performance in both groups of learners may also be useful to help support both of these distinct student populations, as well as helping us understand how course format influences student behavior and performance.

In this paper, we therefore compare online behaviors from two groups learners with different intentions, enrolling on For-Credit and non-credit basis, who are using the same course and platform concurrently. Specifically, we develop models based on student interaction and

participation to predict final course grades, for each population. In doing so, we identify features of student interaction and participation which are intended to be interpretable, both to support instructors in eventual intervention based on these features and for scientific interpretation of our findings. We then examine whether these automated models can generalize from one distinct student population to the other, in order to understand whether the same patterns of interaction are predictive of student success in both populations.

Methods

Datasets

We analyze these issues in the context of data from NextThought, an interactive, online learning platform used to deliver online courses for a range of universities. The NextThought platform emphasizes community and social interaction around course materials, with features such as the ability to create threaded conversations and group-visible notes contextualized within pieces of content such as videos and readings. Data on online interactions were collected from students who enrolled in an online Humanities course at a large public university during the spring and summer terms of 2015. The course was enrolled in by both 143 For-Credit students and 90 Open students (i.e., those enrolled in the courses on a non-credit basis). These two groups of students took the course concurrently, sharing the same resources and discussion threads.

The interaction logs of the NextThought platform were distilled to produce the following 9 features, grouped into five types. The same features were used to describe the behavior of both For-Credit students and Open students:

- Readings: number of times a student viewed readings
- Forum: number of times a student viewed a forum
- Video: number of times a student viewed a video
- Comments: number of comments a student created, number of top-level comments a student created (i.e., not replies to other student's comments), number of comments created in response to a student's original comment, and number of comments that were replies to other comments
- Social interaction: number of distinct students who replied to a given student and number of distinct students a given student has replied to

All of these features were included in our predictive analysis of students' final course performance. Students' final average grades were computed based on course requirements for each term. The syllabus specified performance criteria, which were set up by the instructor to be different between the two student populations. The For-Credit students were told that an average grade of 0.70 was equivalent to a C letter grade. The Open students were told that completing 200 points out of 350 points would result in receiving a badge. [We followed the instructor's decision and used different standards of success for the two populations, as applying the same standards would imply applying a different standard to one of the groups than that group would normally have been outside the context of this study. MOOCs typically allow learners to drop multiple quizzes and have lower cut-offs for obtaining a certificate than For-Credit courses.](#) . In order to authentically represent whether each student matched the instructor's expectation for successful performance, the instructor's chosen criteria were used as cut-offs in our models to determine whether students passed or failed the course: 0.70 for For-Credit students and 0.5714

(200 out of 350 points) for Open students. Using these differing standards of success represents the authentic expectations set by the instructor for each group, and reflects what typically occurs in most online courses, where MOOCs have lower standards for obtaining a certificate than For-Credit courses have for passing.

Data Analysis

We use a predictive analytics/educational data mining paradigm (Romero & Ventura, 2007; Baker & Siemens, 2014) to interpret the performance of these two groups of students. Using this paradigm enables us to combine features of online interactions in more complex patterns and evaluate these patterns in terms of whether they are likely to generalize to new students, rather than in terms of whether findings falsify hypotheses within the current student population (Baker, 2015).

Within this paradigm, we develop a pair of models, one for each population, to predict whether students will pass or fail the course, according to the criterion set by the instructor. Each pass/fail prediction model was built using the J48 implementation of the C4.5 algorithm (Quinlan, 1993). In a J48 decision tree, a set of if-then choices is based on features of student interaction, culminating in a final probability of whether the student passed or failed the course. We assessed the J48 models using two metrics: Cohen's (1960) Kappa and AUC ROC (Hanley & McNeil, 1982). Cohen's Kappa assesses the degree to which our models are better than chance at predicting students' final course performance. A Kappa of 0 indicates that the model performs at chance, and a Kappa of 1 indicates that the model performs perfectly. For example, a Kappa of 0.31 would indicate that the model is 31% better than chance. AUC ROC is the probability that the algorithm will correctly identify whether a student will pass or fail in the online course. It is mathematically equivalent to the Wilcoxon statistic. A model with AUC ROC of 0.5 performs at chance, and a model with AUC ROC of 1.0 performs perfectly. Note that AUC ROC was computed at the student-level.

We built our models in RapidMiner 5.3 (Mierswa et al., 2006). All of our models were developed using 10-fold cross-validation at the student-level (e.g. models are repeatedly trained on 9 groups of students and tests on the 10th group of students), in order to assess how accurate the models and findings will be for new students.

Beyond simply predicting whether a student will pass or fail, we also built models to predict numerical course grades using linear regression with greedy feature selection, using RapidMiner 5.3. This algorithm repeatedly removes the worst-performing attribute until the Akaike (1973) information criterion (AIC) no longer improves, creating simple models with reduced risk of over-fitting (Das, 2008). All predictor variables were standardized using z scores to increase the interpretability of the resulting coefficients; note that this standardization procedure does not influence model goodness or predictive power. Model goodness is assessed using cross-validated correlations (Pearson's r) between the model and data, and root mean square error (RMSE) values. Positive cross-validated correlation indicates the relationship is consistent between the training and test datasets (but has no implication about the relationship's direction); negative cross-validated correlation indicates that the models obtained from the training dataset are worse than chance when applied to new students.

Beyond testing the models for new students in their original populations, we can test the generalizability of the classification models between the Open and For-Credit students, building the model on one group and evaluating it on the other. In other words, models trained on the For-Credit student population were tested on the Open student population, and vice versa; goodness

metrics were then calculated to determine how each of the models was generalizing to the new population. These tests allow us to compare the degree to which each model (and the patterns they capture) generalize across the two populations studied here. This also enables us to understand how similar the relationships being modeled are between data sets. In our case, model generalizability tests are important for determining whether the same features are predictive of student success across our two student populations, regardless of any potential pre-existing group differences. For example, if a model trained on one population works on the other population, we can conclude that the factors which lead to course completion are similar between the two populations. Testing generalizability is better evidence than simply looking at whether the models have the same features, as often the exact models found are only slightly better than several alternate models, which may have different features.

Results

Descriptives

We found that the average proportion of final course grades was considerably higher for students who enrolled for the online course on a For-Credit basis ($M = 71.48$, $SD = 0.30$), than students who enrolled on a non-credit basis ($M = 31.61\%$, $SD = 0.32$).

Models predicting student pass/fail

In this section, we examine the overall predictive power of the J48 decision trees predicting whether a student would successfully pass the course, for both groups of students (For-Credit and Open). We analyze the model predictive power for new students, within-group and across-groups. We will interpret the features within these models in closer detail in a later section.

The J48 model for Open students achieved a cross-validated AUC ROC of 0.674 and a cross-validated Kappa of 0.288. For students who enrolled For-Credit, the J48 model achieved a cross-validated AUC ROC of 0.685 and a cross-validated Kappa of 0.470. These results show that both models are successful at predicting whether a student passes or fails, with the For-Credit model appearing to perform slightly better than the model developed for the Open student population. The finding that the For-Credit student model performed better than the Open student model is perhaps not unsurprising given that Open courses likely attract a more diverse collection of students with a wider variety of reasons for enrolling in classes (Clow, 2013). These results suggest that our models are capable of predicting student achievement for new students within each of our two distinct populations.

Across both of our student populations, the J48 models included the following features: the number of comments, the number of replies, the number of forum views, and the number of video views. As shown in Figure 1, the Open student model also included additional features not incorporated in the For-Credit student model: the number of top-level comments, number of comments created in response to a student's original comment, the number of reading views, and the number of distinct students who replied to a given student. Additionally, as shown in Figure 2, the For-Credit model included the number of distinct students a given student has replied to, which was not incorporated in the Open student model.

```

Number of top-level comments a student created <= 3
| Number of times a student viewed readings <= 37: PASS (100%)
| Number of times a student viewed readings > 37
| | Number of comments a student created = 0
| | | Number of times a student viewed a video <= 7: PASS (100%)
| | | Number of times a student viewed a video > 7
| | | | Number of times a student viewed a forum <= 10: FAIL (100%)
| | | | Number of times a student viewed a forum > 10: PASS (75%)
| | Number of comments a student created > 0
| | | Number of distinct students who replied to student = 0: PASS (92.86%)
| | | Number of distinct students who replied to student > 0: FAIL (75%)
Number of top-level comments a student created > 3
| Number of comments created in response to a student's original comment <= 18: FAIL
(83.33%)
| Number of comments created in response to a student's original comment > 18
| | Number of comments that were replies to other comments <= 9: PASS (100%)
| | Number of comments that were replies to other comments > 9: FAIL (100%)

```

Figure 1. J48 decision tree predicting whether Open-students pass or fail the course.

Percentages reflect the model's degree of certainty in an assessment. This tree can be read as follows (for example): If a student makes 3 or fewer top-level comments AND views readings 37 or fewer times, the student is predicted to pass 100% of the time. But if a student makes 3 or fewer top-level comments AND views readings more than 37 times AND creates 0 comments AND views videos more than 7 times AND views forums 10 times or fewer, the student is predicted to fail 100% of the time.

```

Number of comments a student created <= 2: PASS (94.74%)
Number of comments a student created > 2
| Number of distinct students a given student has replied to <= 2
| | Number of comments that were replies to other comments = 0: FAIL (88.89%)
| | Number of comments that were replies to other comments > 0
| | | Number of times a student viewed a forum <= 14: PASS (100%)
| | | Number of times a student viewed a forum > 14
| | | | Number of times a student viewed a video <= 44: FAIL (13.89%)
| | | | Number of times a student viewed a video > 44: PASS (100%)
| | Number of distinct students a given student has replied to > 2: FAIL (90.52%)

```

Figure 2. J48 decision tree predicting whether For-Credit students pass or fail the course.

Model generalizability. The J48 model trained on the Open population performed considerably better than chance when tested upon new For-Credit students, with a AUC ROC of 0.745 and a Kappa of 0.495. The J48 model trained on the For-Credit population achieved well above chance when applied upon new Open students, with a AUC ROC of 0.629 and a Kappa of 0.400. These numbers, on average, were reasonably close to the numbers the models achieved on their original data sets. Overall, these results suggest that the classification models

trained on a single student population generalize well to the other student population, and that there are similarities between the factors associated with passing in the two populations.

Linear Regression models

After modeling whether students passed or failed their courses, we generated models to predict numerical grades, both because these are useful in its own right, and because the components of these models are typically easier to interpret than the internals of decision trees.

Individual-feature models. Our first step is to generate linear regression models for each data feature, taken individually, in order to understand how each feature correlates to student grade. We do this separately for each group of students (For-Credit and Open). Table 1 shows the relationship of each of the individual features with student achievement among the Open students. For this group of students, each of our features was found to significantly and positively predict students' proportions of final grades. In other words, every behavior was positively correlated with student grade.

Table 2 provides a summary of the relationship of each individual feature with student grades in the For-student population. For these students, as with the students taking the course on an Open basis, each of our features has a positive and statistically significant relationship with student's proportion of final grades.

Table 1

Individual feature correlations to student grades among Open students. Coefficients show relationship direction – all relationships were positive.

Features	Coefficient	Cross-validated r
Number of times a student viewed readings	0.143***	0.296
Number of times a student viewed a forum	0.185***	0.266
Number of times a student viewed videos	0.119***	0.308
Number of comments a student created	0.167***	0.293
Number of top-level comments a student created	0.207***	0.255
Number of comments that were replies to other comments	0.122***	0.314
Number of comments created in response to a student's original comment	0.126***	0.304

Number of distinct students who replied to a given student	0.150***	0.293
Number of distinct students a given student has replied to	0.103**	0.312

Note. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

Table 2

Individual feature correlations to student grades among For-Credit students. Coefficients show relationship direction – all relationships were positive.

Features	Coefficient	Cross-validated r
Number of times a student viewed readings	0.104***	0.314
Number of times a student viewed a forum	0.085***	0.187
Number of times a student viewed videos	0.093***	0.584
Number of comments a student created	0.180***	0.430
Number of top-level comments a student created	0.207***	0.678
Number of comments that were replies to other comments	0.135***	0.430
Number of comments created in response to a student's original comment	0.118***	0.345
Number of distinct students who replied to a given student	0.128***	0.394
Number of distinct students a given student has replied to	0.132***	0.418

Note. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

Multiple-feature models. When we combined these features into a single model, the resultant model for students who enrolled on an Open basis achieved a cross-validated correlation of 0.592, with a RMSE of 0.263. This value of cross-validated correlation was 0.28 higher than the best individual-feature model. Given this level of cross-validated correlation, the model is likely to generalize to new students better than chance.

Table 3 provides a summary of the best-fitting linear regression model for students who enrolled in online courses without credit.

Table 3

Final linear regression model predicting average proportions of grades in the Open student population

Features
0.101 * Number of times a student viewed a forum +
0.124 * Number of distinct students who replied to a given student +
0.459 * Number of comments a student created +
-0.183 * Number of comments created in response to a student's original comment +
-0.336 * Number of comments that were replies to other comments +
0.316

As seen in Table 3, a multi-feature model predicting final grades for the Open students incorporated several features, each of which was statistically significant with $p < 0.0001$. The number of times a student viewed the forum and the number of comments a student created were positively correlated with the student's grade, when controlling for other variables in the model. This aligns with the results from the individual-feature models. However, the number of replies and the number of comments created in response to a student's original comment to it flipped direction when controlling for other variables in the full model. This suggests that once we control for how much students view the forum (and interact with it) overall, replying to comments is not characteristic of successful students. This may in part be due to the presence of students who posted many irrelevant comments.

For students who enrolled on a For-Credit basis, we again combined the original set of features into a single model using greedy feature selection. The resultant model of For-Credit student achievement achieved a cross-validated correlation of 0.681 with a RMSE of 0.221, performing slightly better than the full model for Open students. However, this multi-feature model does not perform much better than the best individual model, which achieved a cross-validated correlation of 0.678. That single feature alone, the number of top-level comments the student created, is very predictive, making it difficult for a combined model to perform much better. In fact, only one other feature, the number of times the student viewed readings, enters into the combined model.

Table 4 shows the best-fitting linear regression model for students who enrolled in online courses for credit.

Table 4

Final Linear regression model predicting average proportions of grades in For-Credit student population

Features
0.036 * Number of times a student viewed readings +
0.195 * Number of top-level comments a student created +
0.715

Model generalization. We next apply each model to the other data set, to understand whether the patterns which predict better grades in For-Credit students also predict better grades in Open students, and vice-versa. When we do so, we find that the multi-feature models trained on the Open student population performed well when tested on the population of For-Credit students (correlation on a new dataset = 0.573, RMSE = 0.472). Similar to this, our multi-feature model trained on the For-Credit student population also performed well when tested on the population of Open students (correlation on a new dataset = 0.638, RMSE = 0.471). The correlations remain respectable in both cases; however, there is substantial degradation in the RMSE values. This degradation is likely due to the overall higher average grades in the For-Credit student population relative to the average grades of the Open student population. As there were different cut-offs for passing in the two populations, it stands to reason that the students would set different goals and therefore the same behaviors would be associated with different absolute grades (even as they were associated with the same relative grades).

Discussion and Conclusions

In this paper, we studied which factors in student interaction during an online course predict their success, using classification models to predict whether students succeed or fail, and regression models to predict students' numerical grades. One of our core goals was to investigate whether the same features predict student success within students taking the course on a For-Credit basis and students taking the course on an Open (e.g., non-credit/MOOC) basis. While some of these features have emerged in previous studies, we believe that demonstrating their generality across For-Credit and Open students, when controlling for other aspects of course design and the instructor, makes a valuable contribution.

When taken individually, we found that each of our features were significant predictors of numerical grades within each student group and across both student populations. In other words, the same features were predictive in the two populations. One key finding is that features related to comments were significant predictors of final grades across Open and For-Credit students. This result corroborates previous findings that show a significant relationship between students' online comments and final course grades (Wang & Newlin., 2000; Wang et al., 2001). More specifically, our results also reveal that top-level comments are associated with better course grades - even when controlling for other variables in the model. This finding generalizes

across both Open and For-Credit learners. With the considerable amount of information in discussion forums, it becomes difficult for instructors of both Open and For-Credit courses to track and determine which specific types of comments best support students in their online learning (e.g. Mazzolini & Maddison, 2005). But if instructors know to focus on top-level comments, they can focus their efforts there. As such, studying the content of top-level comments may be a promising approach to evaluate what are engaging students within non-credit and For-Credit courses. This result is also a first step towards identifying which aspects of the online learning experience are general across these two settings.

Within the Open student population, both classification and regression models included the following features: number of times a student viewed a forum, number of distinct students who replied to a given student, number of comments a student created, number of comments created in response to a student's original comment, and number of comments that that were replies to other comment. In addition to these, the classification model for Open students also incorporated the following features that were not included in the For-Credit classification model: the number of times a student viewed readings, the number of times a student viewed videos, and the number of top-level comments a student created.

Within the For-Credit student group, classification and regression models differed from each other. For-Credit classification models incorporated the number of comments, number of distinct students a given student has replied to, number of forum views, number of replies, and number of video views. By contrast, only the number of reading views and the number of top-level comments were included in the regression model. This finding suggests that the factors important for classifying whether For-Credit students pass or fail were not necessarily the same as the predictive factors for finer-grained outcomes of student achievement. In particular, features related to videos, forums, comments and social interaction are predictive of students' success or failure, while features related to readings and a certain type of comment are predictive of students' proportions of final grades.

All models performed well at predicting student grades/pass-fail status, within both populations. Their performance was slightly better for the For-Credit students, for both types of models. The ability to predict student achievement from automated models of student online interactions has potential implications for educational interventions. As our findings indicate that our predictive models of students' success or failure can be applied to new students, this creates the possibility of early remediation for students who are struggling (Arnold & Pistili, 2012). By identifying struggling students early on, instructors can monitor their students' needs and provide scaffolding, before these students fall too far behind. However, as the models were not perfect, there is still a possibility of classifying students in the wrong category, suggesting that these models should be used with "fail-soft interventions", where there are limited risks associated with a student incorrectly receiving an intervention. An example of this is seen in Course Signals, an early warning system that triggers interventions designed by instructors to help students at risk of failing the course. While the interventions are recommended by an automated algorithm, they are mediated by instructors who make the final decision on whether to send a message and how to tailor it based on what they specifically know about the student. In general, initial results show that the intervention was associated with higher student performance and retention, and that students who receive the intervention feel that instructors care about them (Pistili & Arnold, 2010; Arnold & Pistili, 2012). A similar approach is adopted by Jayaprakash and colleagues (2014), who send instructors notifications about at-risk students and scaffold the instructors in sending struggling students emails regarding consultation hours, web-based

resources on online tutoring tools, or additional practice exercises for the course materials -- resulting in higher final grades than students in a control group across multiple universities. Each of these systems mitigate the risk of incorrect classification by putting the control over intervention in the instructor's hands, and by scaffolding intervention strategies with low risk of harm even when given to students who ultimately do not need intervention.

In the specific case of this course, our results show a positive relationship between top-level comments and student outcomes. NextThought affords instructors multiple opportunities for facilitating the student's creation of top-level posts. For example, by posting conversation prompts within the discussion margins of the course content, instructors can provide a model for top-level comments that students can replicate. These prompts help raise the visibility of course materials that are worthy of reflection and comments. In addition, instructors can encourage students to produce top-level comments by incorporating thought questions within non-graded reading assignments. These questions, accompanied by an invitation to post reflections in the discussion margins, provide a framework for generating top-level student comments within the course forums. Finally, when posting at the top-level, students can also be encouraged to consider making comments rather than posting questions, as questions encourage responses to the original post instead of promoting other top-level comments.

More broadly, these results also have potential implications for supporting student success in universities that offer courses like this one. Given our findings, it may be valuable for course instructors to message inactive students and encourage them to start new discussion threads within forums. Sending course-related reminders may be a useful instructional strategy, as students generally respond positively to these types of emails and messages (cf. Arnold & Pistili, 2014). Based on work by Belcher et al. (2015), another way instructors can encourage students to become thread starters is to praise them for posting a top-level comment and to summarize the content of these posts within the thread. These strategies could potentially motivate students to initiate top-level comments and think more critically about them, as they know their top-level posts will be attended to by the instructor. These instructor behaviors have also been found to promote higher levels of critical thinking within peer interactions (Belcher et al., 2015).

Models predicting student success and grade used many of the same features across both For-Credit and Open students. Beyond that, a model built for each group of students functioned effectively for the other group of students. As such, this result indicates that the same factors do predict course completion even in different groups of students who have enrolled either on a For-Credit or non-credit basis. While these two studies likely differ in their motivations, the same factors are associated with completion in both. The relationship between interaction behaviors and success is similar for these two distinct populations of students, in terms of patterns and combinations of features as well as individual features. This finding, in turn, has implications in synthesizing literature on MOOCs and For-Credit courses, that have largely been treated by researchers as disparate bodies of work. Our result suggests that many of the expanding body of findings on success in MOOC courses may be relevant to For-Credit online courses, a useful finding given the much larger samples of students available in many Open courses (if not the specific course studied here).

In considering these findings, it is essential to note their limitations. One key limitation is that the research presented here was conducted in a single online course, with both For-Credit and Open learners, a somewhat unusual arrangement (though characteristic of the NextThought platform).

The sample of students was typical for For-Credit courses but was smaller than the sample seen in many xMOOCs offered by the largest MOOC vendors such as edX, Coursera, and FutureLearn. However, not all MOOCs are xMOOCs. In particular, the original use of the term MOOC referred to connectivist MOOCs (cMOOCs), whose design is more based on student-to-student communication and definition of questions and selection of resources (McAuley et al., 2010). While the course studied here is not a cMOOC per se, it shares with cMOOCs the strong focus on student-to-student communication. It is common for cMOOCs to have sizes more comparable to what is seen in this study, rather than the massive sizes seen in xMOOCs from the largest providers. However, this specific atypicality does limit our ability to generalize our findings somewhat. As in the case with most MOOC studies, findings from a specific course are often difficult to generalize beyond the local context. For instance, while For-Credit and Open learners were treated as discrete sets of learners, not all online courses will necessarily apply the same categorization. With the rise of MicroMasters, students who initially enroll in a MOOC can also pursue and eventually earn a master's degree. As such, future work is recommended to test whether these models remain valid when extended to a wider sample of learners, as well as to broader range of MOOCs. Despite these limitations, the findings from our automated models provide information on which factors of student online interactions promote achievement and learning. Additionally, our study shows which findings validated for different groups of students within the current course, and which can be replicated across additional courses. By better understanding the factors in student interaction (whether in For-Credit or Open courses) that are associated with greater success, we can design interventions that instructors can use to nudge students to better learning trajectories and better outcomes.

Acknowledgements

We would like to thank the course instructor, and the other members of the NextThought team.

References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Pertaran, B.N., Csaaki, F. (Eds.), *International Symposium on Information Theory*, 2nd ed. Akademiai Kiado, Budapest, Hungary, pp. 267–281.
- Andergassen, M., Guerra, V., Ledermüller, K., & Neumann, G. (2013). Development of a browser-based mobile audience response system for large classrooms. *International Journal of Mobile and Blended Learning (IJMBL)*, 5(1), 58-76
- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267-270.
- Baker, R., Lindrum, D., Lindrum, M.J., Perkowski, D. (2015) Analyzing Early At-Risk Factors in Higher Education e-Learning Courses. In *Proceedings of the 8th International Conference on Educational Data Mining*, 150-155.

- Beard, L.A., Harper, C. (2002) Student Perceptions of Online versus on Campus Instruction. *Education*, 122(4).
- Belcher, A., Hall, B. M., Kelley, K., & Pressey, K. L. (2015). An Analysis of Faculty Promotion of Critical Thinking and Peer Interaction within Threaded Discussions. *Online Learning*, 19(4).
- Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8, 13-25.
- Clow, D. (2013). MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 185-189.
- Cohen, J (2009). A coefficient of agreement from nominal scales. *Educational and Psychological Measurement*, 20 (1), 37-46.
- Das, A., & Kempe, D. (2008, May). Algorithms for subset selection in linear regression. In *Proceedings of the fortieth annual ACM symposium on Theory of computing* (pp. 45-54). ACM.
- Diaz, D.P. (2002) Online Drop Rates Revisited. The Technology Source. May/June. Retrieved on October 16, 2015, from http://technologysource.org/article/online_drop_rate_revisited/
- Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) Curve. *Radiology*, 143, 29-36.
- Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The first year of Open online courses (HarvardX and MITx Working Paper No. 1). doi:10.2139/ssrn.2381263
- Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6-47.
- Jiang, S., Warschauer, M., Williams, A. E., O'Dowd, D., & Schenke, K. (2014, July). Predicting MOOC performance with week 1 behavior. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 273-275).
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). Deconstructing disengagement: analyzing learner subpopulations in massive Open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179). ACM.
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive Open online courses: In depth. *Educause Review*, 48(3), 62-63.

- Mazzolini, M., & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computers & Education*, 49(2), 193-213.
- McAulay, A., Stewart, B., Siemens, G., Cormier, D. (2010) The MOOC model for digital practice. Retrieved online on 3/30/2017 from http://www.elearnspace.org/Articles/MOOC_Final.pdf
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 935-940.
- Morris, N. P., Hotchkiss, S., & Swinnerton, B. (2015). Can demographic information predict MOOC learner outcomes?. *Proceedings of the European MOOC Stakeholder Summit*, 199-207.
- Nistor, N., & Neubauer, K. (2010). From participation to dropout: Quantitative participation patterns in online university courses. *Computers & Education*, 55(2), 663-672.
- Pistilli, M. D., & Arnold, K. E. (2010). In practice: Purdue Signals: Mining real-time academic data to enhance student success. *About Campus*, 15(3), 22-24.
- Quinlan, R. (1993), C4.5: Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann Publishers.
- Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68, 458-472.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1), 135-146.
- Shrader, S., Wu, M., Owens-Nicholson, D., & Santa Ana, K. (2016). Massive Open online courses (MOOCs): Participant activity, demographics, and satisfaction. *Online Learning*, 20(2).
- Tyler-Smith, K. (2006). Early attrition among first time eLearners: A review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking eLearning programmes. *Journal of Online learning and Teaching*, 2(2), 73-85.
- Wang, Y., Paquette, L., & Baker, R. (2017). Workshop on integrated learning analytics of MOOC post-course development. *Proceedings of the Seventh International Learning Knowledge Conferences*, 506-507.
- Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology Web-based classes. *Journal of educational psychology*, 92(1), 137.

- Wang, A. Y., Newlin, M. H., & Tucker, T. L. (2001). A discourse analysis of online classroom chats: Predictors of cyber-student performance. *Teaching of Psychology, 28*(3), 222-226.
- Whitmer, J. (2012) Logging on to improve achievement: Evaluating the relationship between use of the learning management system, student characteristics, and academic achievement in a hybrid large enrollment undergraduate course. Unpublished Doctoral Dissertation, UC Davis.
- Yang, D., Sinha, T., Adamson, D., & Rose, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive Open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop*, 11,14.
- Zafra, A., & Ventura, S. (2009). Predicting Student Grades in Learning Management Systems with Multiple Instance Genetic Programming. Paper presented at the *Proceedings of the 2nd International Conference on Educational Data Mining*, Cordoba, Spain.
- Zhang, J., & Walls, R. (2006). Instructors' self-perceived pedagogical principle implementation in the online environment. *The Quarterly Review of Distance Education, 7*(4), 413-426.