Understanding MOOC Stopout Patterns: Course and Assessment-Level Insights

Yunlang Dai
Haverford College
Haverford, PA, United States

Kirk Vanacore
University of Pennsylvania
Philadelphia, PA, United States

Ryan Baker
University of Pennsylvania
Philadelphia, PA, United States

Stefan Slater
University of Pennsylvania
Philadelphia, PA, United States

Abstract

This study investigates stopout patterns in MOOCs to understand course and assessment-level factors that influence student stopout behavior. We expanded previous work on stopout by assessing the exponential decay of assessment-level stopout rates across courses. Results confirm a disproportionate stopout rate on the first graded assessment. We then evaluated which course and assessment level features were associated with stopout on the first assessment. Findings suggest that a higher number of questions and estimated time commitment in the early assessments and more assessments in a course may be associated with a higher proportion of early stopout behavior.

CCS Concepts

• **Applied computing → E-learning**.

Keywords

MOOC, stopout, student engagement, course design, online learning analytics

ACM Reference Format:
Yunlang Dai, Kirk Vanacore, Ryan Baker, and Stefan Slater. 2025. Understanding MOOC Stopout Patterns: Course and Assessment-Level Insights. In *Proceedings of the Twelfth ACM Conference on Learning @ Scale (L@S '25), July 21–23, 2025, Palermo, Italy.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3698205.3733944

1 Introduction

Massive Open Online Courses (MOOCs) have been widely used for delivering educational content to a global audience [14]. Over the past decade, there have been over 19,000 courses and more than 200 million MOOC users around the globe [10]. Despite their accessibility and flexibility, MOOCs face a significant challenge with high dropout rates, which undermines their educational impact [9]. Understanding why students drop out at various stages of a course is essential for improving course design and increasing retention.

To achieve a more granular understanding of dropout behavior, this study focuses on the *assessment-level stopout rate*—the percentage of students who drop out of the class after completing a specific assessment within a MOOC. More specifically, we are interested in exploring general stopout patterns across different courses and identifying common course-related features that may influence these patterns.

In doing so, we address two key research questions:

(1) Does stopout in MOOCs follow patterns of exponential decay similar to those found by Botelho et al. (2019) within middle school math assignments?
(2) What course and assessment-level features are associated with stopout in MOOCs?

2 Background

Dropout has historically been recognized as a significant challenge in Massive Open Online Courses (MOOCs), with completion rates often below 15% [6, 9]. Recent studies confirm that this trend persists, including a 2025 systematic review and a 2024 comparative study highlighting continued low completion rates [1, 3]. Furthermore, studies have found that students tend to dropout of MOOCs within the first few assessments [7]. This behavior is often called stopout [8, 11, 12]. Thus, understanding which MOOC assessment features drive stopout behavior may help MOOC designers build courses that improve student retention.

This has led to considerable investigation into why students stop out, including efforts to understand the timing and context of disengagement, as well as differences in students' prior knowledge [8, 11]. For instance, Taylor et al. [11] demonstrated the effectiveness of predicting student stopout in MOOCs using the first four weeks of historical data. Matcha et al. [8] examined different stopout patterns among learners with varying levels of prior knowledge.

Moreover, assessment-level features have been shown to shape stopout behavior [14]. Early assessments influence students' perception of course difficulty and workload. If these assessments are too complex or demanding, students may disengage prematurely. Botelho et al. [2] observed an exponential pattern of disengagement over time on digital math learning platforms, but this pattern has not yet been thoroughly examined in MOOC contexts.

This study investigates how general course and assessment-level features contribute to stopout patterns, with the goal of informing course design strategies to reduce early stopout.

L@S '25, Palermo, Italy

ACM ISBN 979-8-4007-1291-3/2025/07
https://doi.org/10.1145/3698205.3733944

3 Research Context and Data

3.1 Research Context

The current study examines a collection of MOOCs offered through Coursera by the University of Pennsylvania from 2015 to 2025. This collection includes over 200 courses and has been used by over 5.7 million individual users. The courses cover a diverse range of subjects, including mathematics, business, law, computer science, and English. The flexibility of enrollment and course completion has made MOOCs accessible to a global audience, but student engagement remains highly variable, contributing to significant stopout rates.

3.2 Data Preparation and Cleaning

We used Penn Coursera MOOC data through the MORF repository database [4], collected from September 2015 to February 2025 on courses whose structure remained consistent during that time period. Additionally, courses with missing values in the engineered features (discussed below) were excluded, resulting in a final dataset of 90 courses with 5,370,635 student enrollments. For our analyses, we focused specifically on required, graded quizzes or tests of the course, which we refer to as assessments, because students need to complete these components to complete the course. In total, our dataset includes 469 graded assessments across all courses, with an average of 6.26 assessments per course. Individual courses had between 3 and 26 graded assessments. The data cleaning and feature engineering scripts used in this study are available online.[1]

3.3 Stopout Definition

The primary response variable in this study is the stopout percentage at the assessment level. Following Taylor et al. [11], we define a student's stopout point as the assessment after which the learner fails to submit any further graded assessments. For instance, if a student's last graded assessment is the third assessment, their stopout point is considered to be the third assessment. The stopout percentage for each assessment was the number of students who dropped out after that assessment divided by the total course enrollment. If no dropout events were observed for an assessment, a stopout percentage was zero. Since a student cannot drop out of a course after completing the last assessment, we exclude last graded assessments in each course from our data.

3.4 Course and Assessment-Level Feature Extraction

To explore how course and assessment design may influence stopout behavior, we extracted a range of features from the dataset at both the course and assessment levels. At the course level, we computed several indicators identified in previous research as predictors of student performance and engagement in MOOCs [13]. These included the total number of assignments (e.g., essays, projects, viewing course content) in the course, the expected course length in days, the number of assessments (quizzes/tests) in a course, and the required number of reviews in peer-graded assignments. We also measured student engagement by counting the number of forum posts and comments. To account for potential submission or grading type influence on engagement rate, we count the total number of submission and grading types. These indicators were averaged across all branches (updates) of each course to create consistent course-level features.

Since our analysis examines assessment-level stopout, we also evaluated which assessment-level features were associated with stopout. At the assessment level, we collected detailed information on the graded assessments within each course. Assessment-level features included the number of questions, the distribution of question types (e.g., multiple choice vs. numeric etc.), and the estimated time commitment to complete the assessment.

[1]https://github.com/davidhhhhhhh/MoocFeatureAnalysis.git

4 Analysis 1: Assessing Stopout Trend

4.1 Method

To evaluate whether stopout behavior in MOOCs follows an exponential decay pattern, we estimated two generalized mixed-effects models with a gamma distribution and a log link. The gamma distribution is suitable for modeling positive, skewed data such as stopout rates, and the log link captures the exponential nature of the relationship between stopout percentage and the sequence of graded assessments.

Similar to Botelho et al.'s [2] models, we fit one model with all assessments and another model excluding the first assessment. Both models included a variable for the ordered assessment sequence within the course, which captures the exponential decay rate. However, our method differs from Botelho et al. [2] in also including course-based random intercepts to account for variability across different courses. We then compared the model fit using the deviance statistic and conditional R^2, which measures the proportion of variance explained by both fixed and random effects.

4.2 Results

Table 1: RMSE comparison with and without the first assessment

	With 1st	Without 1st
all assessments	0.070	0.084
assessment ≥ 2	0.029	0.029

The generalized mixed-effects model demonstrated a strong fit to the data, with a conditional R^2 of 0.49, as shown in Figure 1. The results confirmed an exponential decay pattern in stopout rates across graded assessments, with a 12% decrease in stopout rate for every additional assessment completed (estimate = −0.12, SE = 0.009, p-value $< 2 \times 10^{-16}$). However, the stopout rate at the first graded assessment was disproportionately high and did not follow the general exponential decay trend.

When the first graded assessment was included in the model, the deviance statistic was 349.14, indicating a poorer fit compared to excluding it (deviance = 228.76). We further examined the Root Mean Square Error (RMSE) of the two models on datasets with and without the first assessment. These results are shown in Table 1.

Table 2: Linear regression results for first-assessment stopout across three models: course-level features, assessment-level features, and all features combined

Feature	Model 1: Course Level Coefficient	Model 1: Course Level p-value	Model 2: Assessment Level Coefficient	Model 2: Assessment Level p-value	Model 3: All Features Coefficient	Model 3: All Features p-value
Course Level Features						
Intercept	0.187	<0.001	0.187	<0.001	0.187	<0.001
Number of Assignments	-0.056	0.019	–	–	–	–
Expected Course length (days)	-0.029	0.036	–	–	-0.015	0.183
Number of Assessments	0.062	0.011	–	–	0.032	0.017
Number of Non-Graded Assessments	–	–	–	–	-0.082	0.115
Number of Peer Reviews Required	–	–	–	–	0.071	0.156
First Assessment Features						
Reflection Questions (%)	–	–	-0.025	0.024	–	–
Fill-In Numeric Questions (%)	–	–	0.032	0.018	0.037	0.002
Expected Time Commitment	–	–	–	–	-0.0495	0.060
Number of Questions	–	–	0.069	<0.001	0.073	<0.001
Second Assessment Features						
Expected Time Commitment	–	–	0.024	0.071	0.078	0.002
Number of Questions	–	–	–	–	–	–
Fill-In Numeric Questions (%)	–	–	0.030	0.044	–	–
AIC	-345.48		-380.82		-381.61	
Multiple R^2	0.117		0.450		0.494	

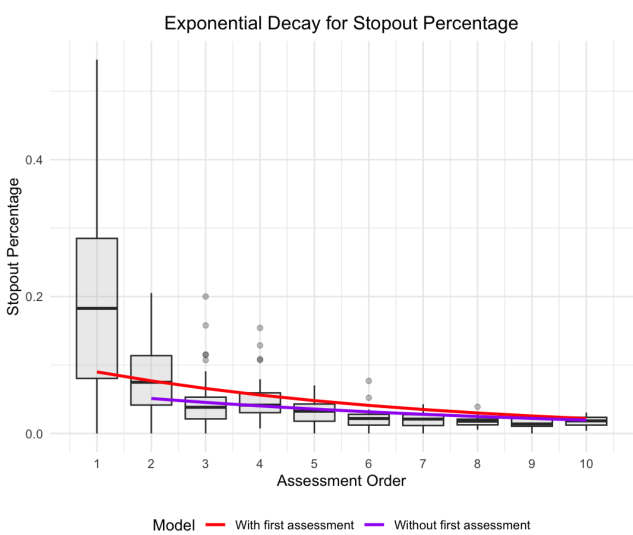

Figure 1: Exponential decay in stopout percentage for the first ten assessments in order, with/out the first assessment.

For all assessments excluding the first, the model trained without the first assessment fits slightly better than the model trained with it. However, when the first assessment is included, the RMSE for both models increases substantially. This suggests that stopout behavior at the first assessment is driven by a different underlying mechanism compared to subsequent assessments, aligning with the findings of Botelho et al. [2] in a distinct learning context.

5 Analysis 2: First Assessment Stopout Rate

Building on the findings from Analysis 1, we explored what features are associated with the disproportionately high stopout rate in the first graded assessment. The average stopout rate at the first graded assessment across all courses was 18.71%, compared to an average of 7.27% for all other graded assessments. This large discrepancy suggests that early disengagement may be influenced by both course design and assessment characteristics.

5.1 Method

To model the predictors of first-assessment stopout, we standardized all continuous variables and estimated three multivariate linear regression models where the dependent variable was the stopout percentage at the first graded assessment.

The first model includes only the course level features, and the second model includes assessment level features on both first assessment (which students stop on first assessment complete) and the second assessment (which students stop on first assessment did not complete). The third model contains both course and assessment-level features to explain the first assessment stopout rate.

For assessment-level features, we used the percentage of total question types within each assessment: multiple-choice questions (72.9% on average), checkbox questions (14.0%), numeric questions

(6.2%), and reflective questions (2.5%). The percentage of multiple-choice questions was excluded from the model to prevent collinearity.

For all three models, we applied stepwise Akaike Information Criterion (AIC) selection using the MASS library in R to maximize model fit while minimizing overfitting. Statistical significance of predictors was evaluated using t-tests, and overall model fit was assessed using the multiple R^2 value.

5.2 Results

Table 2 presents the final coefficients from three linear regression models that predict first-assessment stopout rate. Model 3, which includes both course-level and assessment-level features, achieved the highest multiple R^2 (0.494), compared to Model 1 (course-level only, $R^2 = 0.117$) and Model 2 (assessment-level only, $R^2 = 0.450$). In what follows, we discuss the statistically significant features (i.e., those with $p < 0.05$) for each model.

In Model 1, three course-level features emerged as significant predictors of first-assessment stopout. First, the number of assignments was negatively associated with stopout ($\beta = -0.056$, $p = 0.019$), suggesting that students in courses offering a greater number of assignments to practice are less likely to leave the course early. Second, the course designer's expected time commitment (in days) was also negatively related to stopout ($\beta = -0.029$, $p = 0.036$), indicating that longer courses tend to retain learners through the first assessment. Third, the number of assessments was positively associated with stopout ($\beta = 0.062$, $p = 0.011$), suggesting that courses with more assessments may deter students from progressing beyond the first assessment.

For Model 2, assessment-level features for both the first and second assessments significantly predicted stopout. A higher proportion of numeric fill-in questions in the first graded assessment was positively associated with stopout ($\beta = 0.032$, $p = 0.018$), and a similar effect was observed for the second graded assessment ($\beta = 0.030$, $p = 0.044$). These results imply that more numeric fill-in questions in assessments increases the likelihood of early stopout. In contrast, a higher percentage of reflection questions in the first assessment was negatively associated with stopout ($\beta = -0.025$, $p = 0.024$), indicating that a greater emphasis on reflection encourages students to stay in the course. Additionally, the total number of questions in the second graded assessment was positively related to stopout ($\beta = 0.069$, $p < 0.001$), suggesting that lengthier second assessments may further deter students from continuing the course.

In the hybrid Model 3, both course-level and assessment-level features play a role in predicting stopout. At the course level, the number of assessments remains positively associated with stopout ($\beta = 0.032$, $p = 0.0165$), indicating that more assessment-intensive courses may discourage students from continuing. At the assessment level, the number of questions in the first graded assessment is positively related to stopout ($\beta = 0.073$, $p < 0.0001$) and a higher proportion of numeric fill-in questions in the first assessment is linked to increased stopout ($\beta = 0.037$, $p = 0.0023$). This suggests that students may find numeric questions more challenging than multiple-choice questions (the baseline). Additionally, the course designer's expected time commitment for the second graded assessment significantly influenced stopout ($\beta = 0.078$, $p = 0.001$).

6 Discussion and Conclusion

The findings suggest that stopout in MOOCs generally follows an exponential decay pattern across graded assessments, but the disproportionately high stopout rate at the first graded assessment indicates that early disengagement follows a different underlying mechanism. This extends the findings from Botelho et al. [2], who found the exponential decay pattern in an intelligent tutoring system, to MOOCs. At the course level, higher assessment counts were linked to increased stopout rates at the first graded assessment, suggesting that more intensive course structures may overwhelm students early on. This is content with Xing et al. [13]'s finding that students were more likely to drop out of courses with fewer assignments and more quizzes.

This paper extends the evaluation of course features to focus on assessment-level features. We find that a higher proportion of fill-in numeric questions, more questions, and greater time commitments expected for subsequent assessments appear to deter students from progressing past the first graded assessment. Students tend to stop out when they are expected to provide numeric fill-in answers, although this may be due to differential dropout rates by content type. This finding may also be caused by the greater difficulty of fill-in problems compared with multiple-choice questions [5]. Notably, the expected time spent on the first assessment was negatively correlated with stopout, while the expected time on the second assessment was positively associated with stopout. This may reflect that students lose motivation to complete long assessments as the course progresses.

These findings highlight the importance of early course design and assessment strategy in improving student retention. Overall, lengthening the course completion time and increasing the number of non-graded assignments may help students stay in the course for longer. Furthermore, shorter assessments, specifically early on in the course, may also delay stopout. Increasing opportunities for reflection may also be helpful.

7 Limitations

Although this study presents evidence that course level and assessment level features may influence stopout behavior, there are some key limitations. First, this correlational study does not fully account for some potential confounders, such as course topics. Future work should use experimental or quails-experimental methods to evaluate which features cause students to drop out. Second, the current study focuses on assessments (quizzes/tests) in MOOCs, but students may also stop out due to the features of assignments (e.g., essays, forum discussions) and course content. These possibilities should also be explored in future work. Finally, while this project uses data from almost a hundred courses, all of these were offered through one learning management system (Coursera), which may have specific design elements that influence the iteration between course features and dropout. Future work should look across learning management systems to test the generalizability of these findings.

Acknowledgments

We would like to acknowledge support from NSF#DRL-1931419.

References

[1] Z. Betaitia, A. Chefrour, and S. Drissi. 2025. Dropout Rates in MOOC Research: A Bibliometric Analysis. *Journal of Learning for Development* 12, 1 (2025), 76–92. doi:10.56059/jl4d.v12i1.1597

[2] A. F. Botelho, A. Varatharaj, E. G. Van Inwegen, and N. T. Heffernan. 2019. Refusing to try: Characterizing early stopout on student assignments. In *Proceedings of the 9th International Learning Analytics & Knowledge Conference (LAK19)*.

[3] B. Celik and K. Cagiltay. 2024. Uncovering MOOC completion: A comparative study of completion rates from different perspectives. *Open Praxis* 16, 3 (2024), 445–456.

[4] J. Gardner, C. Brooks, J. M. L. Andres, and R. Baker. 2018. MORF: A Framework for Predictive Modeling and Replication at Scale with Privacy-Restricted MOOC Data. *arXiv preprint* arXiv:1801.05236 (2018).

[5] Ashish Gurung, Kirk Vanacore, Andrew A Mcreynolds, Korinn S Ostrow, Eamon Worden, Adam C Sales, and Neil T Heffernan. 2024. Multiple Choice vs. Fill-In Problems: The Trade-off Between Scalability and Learning. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*. 507–517.

[6] K. Jordan. 2015. Massive open online course completion rates revisited: Assessment, length and attrition. *International Review of Research in Open and Distributed Learning* 16, 3 (2015).

[7] Katy Jordan. 2015. Massive open online course completion rates revisited: Assessment, length and attrition. *International Review of Research in Open and Distributed Learning* 16, 3 (2015), 341–358.

[8] W. Matcha, R. Natthaphatwirata, N. A. Uzir, et al. 2024. Dropout is not always a failure! Exploration on the prior knowledge and learning behaviors of MOOC learners. *Journal of Computers in Education* (2024). doi:10.1007/s40692-024-00340-z

[9] D. F. O. Onah, J. Sinclair, and R. Boyatt. 2014. Dropout Rates of Massive Open Online Courses: Behavioural Patterns. In *Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN14)*. 5825–5834.

[10] D. Shah. 2021. A decade of MOOCs: A review of MOOC stats and trends in 2021. https://www.classcentral.com/report/MOOCs-stats-and-trends-2021. Accessed April 10, 2025.

[11] C. Taylor, K. Veeramachaneni, and U.-M. O'Reilly. 2014. Likely to stop? Predicting stopout in Massive Open Online Courses. *arXiv preprint* arXiv:1408.3382 (2014). doi:10.48550/arXiv.1408.3382

[12] Jacob Whitehill, Joseph Williams, Glenn Lopez, Cody Coleman, and Justin Reich. 2015. Beyond prediction: First steps toward automatic intervention in MOOC student stopout. *Available at SSRN 2611750* (2015).

[13] W. Xing. 2019. Exploring the influences of MOOC design features on student performance and persistence. *Distance Education* 40, 1 (2019), 98–113.

[14] C. Ye and G. Biswas. 2014. Early Prediction of Student Dropout and Performance in MOOCs using Higher Granularity Temporal Information. *Journal of Learning Analytics* 1, 3 (2014), 169–172. doi:10.18608/jla.2014.13.14