

# Attrition in Massive Open Online Courses

René F. Kizilcec

Department of Communication  
Stanford University  
kizilcec@stanford.edu

Sherif Halawa

Department of Electrical Engineering  
Stanford University  
halawa@stanford.edu

## ABSTRACT

Attrition in online learning is generally higher than in traditional settings. High attrition in large-scale open learning environments prompted a shift from a monolithic notion of success toward a more nuanced treatment that accounts for the diversity in learners' motivation. A systematic analysis of individual differences in attrition in 17 diverse massive open online courses ( $N > 60,000$ ) revealed substantial achievement gaps: women were less likely to persist than men, and learners in Africa, Asia, and South America persisted less and scored lower than learners in Europe, North America, and Oceania. More persistent learners were, however, only marginally more satisfied with their achievement. Successful online learners who left the course early reported higher levels of goal striving, feelings of social belonging, and growth mindset than unsuccessful ones ( $N = 756$ ). The shortage of time was the primary obstacle for most learners, and many were hindered by insufficient volitional control. A better understanding of why learners disengage from MOOCs allows for earlier recognition of impending attrition, which enables targeted interventions to help learners achieve their goals.

## Author Keywords

Online learning, MOOC, persistence, dropout, individual differences, volition, social belonging, mindset

## ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; K.3.1. Computers and Education: Computer Uses in Education

## INTRODUCTION

Educational environments have become increasingly diverse. Traditional schools and universities have a characteristically rigid structure, including instructor-defined—even nationally agreed—syllabi, fixed time schedules, entry requirements, and material costs to enter and exit. Novel institutional structures have been developed to overcome particular constraints. Community colleges, for instance, were created in an attempt to democratize education by offering instruction at a lower cost, with lower admission criteria, and with more flexible

schedules to accommodate those who cannot afford to be full-time students [9]. Distance learning programs deliver education to remote parts of the world and to people who simply could not attend in-person classes. Course materials, including assessments, were delivered through mail (correspondence education), radio, television, and eventually the Internet, thereby addressing geographical and time related constraints of traditional instruction [16].

The latest generation of online learning environments, characterized by massive open online courses (MOOCs), has pushed the boundary on the scale of education [26]. By design, MOOCs provide course materials to millions of people worldwide. This scale could be achieved by pre-recording lectures, designing assessments that can be graded automatically, and by leveraging the momentum of the number of people involved (e.g., to facilitate peer learning or peer assessment). Maybe by virtue of their large scale, their prominent instructors, or their adherence to contemporary interface designs, MOOCs rapidly became an online media phenomenon. People would sign up weeks in advance of the course launch date, many of whom would never even enter the course site. And among those who enroll and enter the site, a large proportion tends to only “sample” some content and leave again [13]. Many of the prototypical behaviors observed in MOOCs [13, 3] resemble those on online media platforms. This trend has also been reflected in the diversity of MOOC learners' motivations for enrolling [14].

Shortly after the first wave of courses had finished, extensive media coverage led to MOOCs becoming associated with high attrition rates [15, 19]. Early MOOC research cautioned against dichotomizing learners into successes and failures based on course completion [13, 21]. Instead, more nuanced categorizations based on learner behavior [13, 4], motivations [14], or intentions [28] have been suggested. Ultimately, perspectives on persistence and attrition in MOOCs depend on how MOOCs have been conceptualized. For instance, MOOCs can be thought to have bridged two worlds: one is governed by the user-centric norms of online media, where everyone is encouraged to be as active as they wish; the other adheres to the “grammar of schooling” [25], which presupposes instructor-defined goals that students strive to achieve [14]. This view of MOOC participation clearly adds complexity to interpretations of attrition. The present work extends the empirical base of research on attrition to inform the theoretical discourse and pave the path for novel practices.

This paper presents a systematic investigation of attrition in MOOCs, based on self-report and behavioral data collected

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

L@S'15, March 14–15, 2015, Vancouver, Canada.

Copyright © 2015 ACM ISBN/14/04...\$15.00.

DOI string from ACM form confirmation

from 66,076 online learners in 18 courses overall. A better understanding of why learners disengage from MOOCs allows for earlier recognition of impending attrition, which enables targeted interventions to help learners achieve their learning goals. We begin by briefly reviewing the large literature on attrition in educational environments with a focus on important developments in understanding its causes. Building on this foundation of prior work, Study 1 offers insights into reasons for disengaging from MOOCs and explores relationships with prior behavior and reported intentions. In Study 2, to develop a deeper understanding of MOOC attrition, we invited around 6,500 learners who were predicted to have disengaged from a MOOC to provide feedback. Their open response answers about what challenges they faced while taking the MOOC were iteratively coded. We investigated the role of volitional control for learners who reported having not enough time, and examined goal striving, social belonging, and growth mindset in relation to being successful in a MOOC.

## RELATED WORK

Research and theorizing on attrition in education has a rich history. This review is intended to serve as a foundation to build on with the current research. We therefore simultaneously develop research questions and hypotheses to investigate and test in the current work. The focus of this brief review is on how ways of thinking about attrition have evolved over the last decades.

### In-person Education

The majority of early work on attrition centered around theoretical models of students' decision to persist or dropout of a traditional higher education setting. The literature on attrition in higher education is largely concerned with students disengaging from a course of study rather than a single course. An early model suggested that students' persistence is largely driven by their prior behavior, attitudes, and norms [8]. The psychological processes involved in turning an intent to learn into the decision to persist were thought to be mediated by volition, the extent to which the student engages in goal-directed behavior in the face of distraction [6]. Motivation alone is necessary but not sufficient for persistence, because students may fail to sustain efforts in the absence of strong self-regulatory skills.

**H1** Successful learners exhibit higher levels of goal striving than unsuccessful learners.

The next generation of psychological models, which were highly influential in the literature, emphasized the critical role of students' "fit" in the institution. Tinto's [24] student integration model posited that college students' decision to persist is a function of prior experiences and individual students' characteristics, and experiences during college. While prior experiences and characteristics are fixed, schools can influence the college experience, including the degree of social and academic integration. Tinto operationalized academic and social integration by GPA scores and the frequency of positive interactions with peers and instructors, respectively. This resonates with recent work highlighting

the critical role of students' feelings of social belonging in achievement-oriented environments (e.g. [27]).

**RQ1** How do prior characteristics (demographics, geographic location, intentions, and prior experience) influence learners' likelihood of disengaging from the course?

**H2** Successful learners exhibit a greater sense of social belonging than unsuccessful learners.

Tinto's work, which specifically targets traditional college students, prompted universities to be proactive in establishing environments that support student integration. Research on attrition in community college settings reiterates the importance of academic and social integration, but points out that non-persistence could indicate success depending on students' intent—students may leave after accomplishing their goal [2]. Nevertheless, a brief psychological intervention that taught community college students an incremental theory of intelligence (instilling a growth mindset [7]) was found to halve attrition rates and increase academic achievement [20]. This highlights the critical role of mindset and caution against overinterpreting non-persistence as an indicator of goal achievement.

**RQ2a** What proportion of learners who disengaged were satisfied with their achievements in the course?

**RQ2b** How does learners' level of satisfaction depend on their behavior in the course?

**H3** Successful learners have a more growth mindset than unsuccessful learners.

### Distance Education and e-Learning

Distance education, in contrast to traditional in-person education, attracts a different student demographic and typically provides fewer opportunities for social integration. However, students in distance learning programs tend to lead social lives outside of school, maybe working part-time and living with their partner. Building on Tinto's model, Bean and Metzner [1] proposed a conceptual model of persistence that would be more applicable to nontraditional students. Persistence is thought to be a function of background characteristics (demographics, etc.), academic and environmental variables (study habits, financial resources, work and family obligations, etc.), and academic and psychological outcomes (GPA, satisfaction, etc.). The significant change from Tinto's original model was the inclusion of environmental variables to account for the added complexity of nontraditional students' lives.

Rovai [22], combining Tinto's and Bean and Metzner's models with factors specific to online learning and pedagogical styles, proposed a composite persistence model specifically for students in online distance education programs. Among other novel factors, Rovai's model acknowledges the critical role of computer literacy in online learning. And yet, these increasingly complex theoretical models were difficult to apply to real-world settings, and hence, the empirical evidence to support them remained sparse. An exploratory study of reasons for attrition of over 1,000 online students found eight

**Table 1. Summary statistics of learner demographics and behavior in 17 MOOCs**

	Topic Area	Enrolled	Survey Responses			Female	Age <sup>1</sup>	Education <sup>2</sup>	Dropout <sup>3</sup>	Videos Watched		
			Initial	Feedback	Both					> 30%	> 50%	> 80%
C1	Business	61,096	9,943	2,367	1,282	42%	39 (30,49)	n.a.	10%	39%	20%	7%
C2	Business	41,081	6,544	883	630	40%	39 (30,49)	97%	9%	32%	13%	2%
C3	CS	27,104	6,104	2,854	2,116	35%	30 (22,41)	74%	2%	67%	40%	7%
C4	CS	56,154	3,406	1,123	352	11%	33 (27,42)	91%	21%	34%	16%	11%
C5	CS	57,636	5738	402	342	17%	33 (26,43)	91%	14%	66%	55%	28%
C6	Engineering	16,220	2,540	269	182	13%	33 (27,45)	92%	15%	62%	49%	17%
C7	Engineering	7,178	1,312	202	167	12%	27 (22,39)	85%	7%	51%	18%	2%
C8	Humanities	7,634	3,631	152	127	92%	33 (26,45)	93%	7%	45%	11%	0%
C9	Humanities	6,475	3,315	326	262	90%	32 (27,47)	91%	16%	69%	50%	17%
C10	Humanities	9,903	4,227	1,570	1,066	52%	39 (29,52)	89%	3%	84%	74%	21%
C11	Humanities	630	167	80	40	50%	30 (27,40)	97%	16%	74%	55%	16%
C12	Math & Stats	49,317	3,090	862	510	33%	41 (29,55)	88%	31%	27%	10%	1%
C13	Math & Stats	32,006	26,142	5,336	4,096	58%	28 (16,46)	n.a.	4%	69%	38%	8%
C14	Math & Stats	16,857	2,892	363	282	9%	31 (26,38)	99%	14%	61%	45%	25%
C15	Math & Stats	37,017	11,629	1,854	1,521	20%	35 (29,45)	98%	8%	78%	67%	34%
C16	Math & Stats	27,236	2,711	637	610	82%	43 (34,51)	n.a.	1%	64%	36%	5%
C17	Math & Stats	13,132	3,629	843	626	n.a.	n.a.	n.a.	5%	70%	54%	14%

<sup>1</sup> mean (upper, lower quartile) <sup>2</sup> bachelor's or higher degree <sup>3</sup> self-reported in the feedback survey

relevant factors: academic and technical skills, learner motivation, time and support for studies, cost and access to the Internet, and technical problems [18]. An analysis of student behavior on an online education platform showed that 31% of variation in achievement could be accounted for by a small set of participation measures [17].

## MOOCs

A survey of over a hundred learners who dropped out of a MOOC showed that the majority indicated having too little time due to work responsibilities, not enough social support inside and outside of the course, and insufficient academic and technical support from the course [10]. A qualitative analysis of public records, especially forum posts, from 42 MOOCs suggested the following reasons for attrition: lack of time, learner motivation, feelings of isolation, lack of interactivity, insufficient prior knowledge or skills, and hidden costs [12]. A survival analysis of close to 800 learners who had posted on a MOOC discussion forum suggested that the likelihood of dropout was lower for those who i. actively participated in the first week of the course; ii. served as an authority figure in the community on the forum, and iii. did not engage in a particular subcommunity on the forum [29]. In addition, a number of machine learning approaches yielded promising results for early dropout prediction (e.g. [23, 11]), but fell short of offering insights into why a given learner would leave the course.

**RQ3a** What reasons do learners report for disengaging from a course?

**RQ3b** How do reasons for disengaging vary across prior learner characteristics (cf. RQ1)?

## STUDY 1: ATTRITION & SATISFACTION

The present study extends the existing literature on persistence with a representative quantitative treatment of attrition in MOOCs. We address the research questions developed in the previous section with a combination of self-report and behavioral data from 17 diverse MOOCs. Going beyond

specific research questions, this observational analysis is intended to inform our understanding of attrition in MOOCs to better support learners in these environments. This can be achieved through general changes, for instance, to the course format; or by targeted interventions that support a particular group of learners, such as those who exhibit low levels of goal striving (cf. Study 2).

## Methods

Seventeen MOOCs on topics in a variety of disciplines were selected for investigation (cf. Table 1). For each course, we merged data from the initial course survey (distributed in the first weeks of the course), the feedback surveys (distributed in the last weeks of the course), and behavioral data collected in the learning environment. On the initial course surveys, learners reported their gender, age, education, prior experience with the topic and MOOCs, enrollment intentions (on the OLEI scale [14]), and how many hours per week they intend to spend on the course. On the feedback survey, learners reported their satisfaction with the course (6-item index, range from 1 to 7,  $M = 5.29$ ,  $SD = 0.94$ ,  $\alpha = .82$ ), and whether they ‘stopped participating in the course before it ended’ or ‘remained active until the end’. Those who indicated that they stopped participating also reported ‘what influenced their decision to stop participating’ using a grid of 14 binary-choice items (Figure 2 lists reasons).<sup>1</sup> Geographic location was determined by IP address and aggregated by continental region.

Table 1 provides summary statistics for enrollment, survey response rates, demographics, self-reported dropout, and a subset of behavioral indicators for each course. Behavioral indicators were chosen to represent course milestones: watching over 30%, 50%, 80% of lecture videos, attempting over 30%,

<sup>1</sup> Instead of a select-all-that-apply question, learners were encouraged to consider each reason in turn by using binary-choice items. The response options were “Did influence my decision to stop” and “Did NOT influence my decision to stop” to mitigate acquiescence bias relative to a Yes/No or Agree/Disagree scale.

50%, 80% of assessments, and achieving a grade above the 60<sup>th</sup>, 80<sup>th</sup> percentile. A regression analysis predicting course milestones and performance was performed based on a subset of learners who completed the initial course survey in the 17 MOOCs in our dataset. Out of the 63,620 learners who started the survey, we selected those with up to three missing values in the dataset, yielding a sample of  $N = 44,382$ . Missing values were multiply imputed based on other survey responses and pooled estimates are reported, which account for the added uncertainty.<sup>2</sup>

## Results

We address the research questions developed in the previous section in three parts: first, a regression analysis of differential attrition and performance (*RQ1*); second, a comparison of satisfaction levels across subsets of learners (*RQ2ab*), and third, a descriptive analysis of reported reasons for disengaging from the course (*RQ3ab*).

### Individual Differences in Attrition

We fit a logistic mixed-effects (hierarchical) model to predict multiple binary persistence and performance metrics (milestones) to estimate individual differences in learner characteristics, including demographic and geographic features. Correlations between predictors were low ( $|r| < 0.16$ ), except for age and education ( $r = 0.27$ ), which were each significant predictors if the other was excluded. The coefficient on education, for instance, reflects the association between having a bachelor's degree and achieving a milestone, partialing out all other predictors, including age. Fourteen MOOCs in the dataset were modeled as a random effect, in order to generalize to the population of similar MOOCs.<sup>3</sup>

Figure 1 presents transformed regression coefficients with 95% confidence intervals from each model. The estimates represent changes in the likelihood of achieving a milestone relative to an arbitrary baseline population. The regression intercept corresponds to the following baseline group that derives from the coding of regressors: average aged male learners in North America without a bachelor's degree, without prior experience with the topic or other MOOCs, prepared to spend up to 2 hours/week on the course, and not intending to complete it. In this sample of survey respondents, the baseline probability for each milestone was high: videos watched (27% above 30%, 16% above 50%, 8% above 80%) assessments attempted (78% above 30%, 55% above 50%, 33% above 80%), and grade (42% above the 60<sup>th</sup>, 31% above the 80<sup>th</sup> percentile).

A number of significant and substantial individual differences emerged. Women were 10 to 15% less likely than men to persist with lectures and assessments. Women, who constituted 41% of learners in the sample, were also less likely to score a grade above the 60<sup>th</sup> percentile than men. Learners in Africa ( $n = 2,097$ ), Asia ( $n = 8,199$ ), and South America ( $n = 2,574$ ) were 20 to 50% less likely to persist in video watching and tended to achieve lower grades than learners

in North America ( $n = 18,353$ ). In contrast, learners in Europe ( $n = 11,992$ ) exhibited 20 to 50% higher persistence with lectures and assessments than learners in North America. While learners in Asia exhibited lower persistence with videos, they were more likely to attempt more assessments than learners from North America. Learners in Oceania ( $n = 1,166$ ) performed similarly to those in North America and Europe.

Besides gender and geography, the older the learner, the more likely they were to persist in lectures and assessments, but the lower their relative performance. More educated learners, those with a bachelor's or higher degree, were somewhat more likely to start assessments, compared to those with less education, but were less likely to score highly on them. Not surprisingly, prior experience, intending to learn about all topics in the course, and being prepared to spend more hours in the course were strong predictors of persistence. Learners who took prior MOOCs, however, were less likely to persist, yet achieved higher grades.

### Learner Satisfaction

Attrition can be a sign of success if a learner has achieved their personal goal. We investigated how satisfied learners were with their achievements for those who reported to have stopped participating (*RQ2a*) and those who participated to different degrees (*RQ2b*). Satisfaction ratings were compared with a mixed-effects regression to model differences across courses and generalize to a larger population of MOOCs. Learners who reported that they stopped participating were 9% less satisfied than those who reported not to have stopped ( $t_{15140} = 18, p < .001$ ). Moreover, we examined if persistent learners, those who watched more lectures, were also more satisfied. While the proportion of videos significantly predicted higher satisfaction ( $t_{15700} = 14, p < .001$ ), the effect was relatively weak. Those who watched over 75% of videos were only 5% more satisfied than those who watched under 25% ( $M_{<25\%} = 4.97, M_{>75\%} = 5.22$ ).

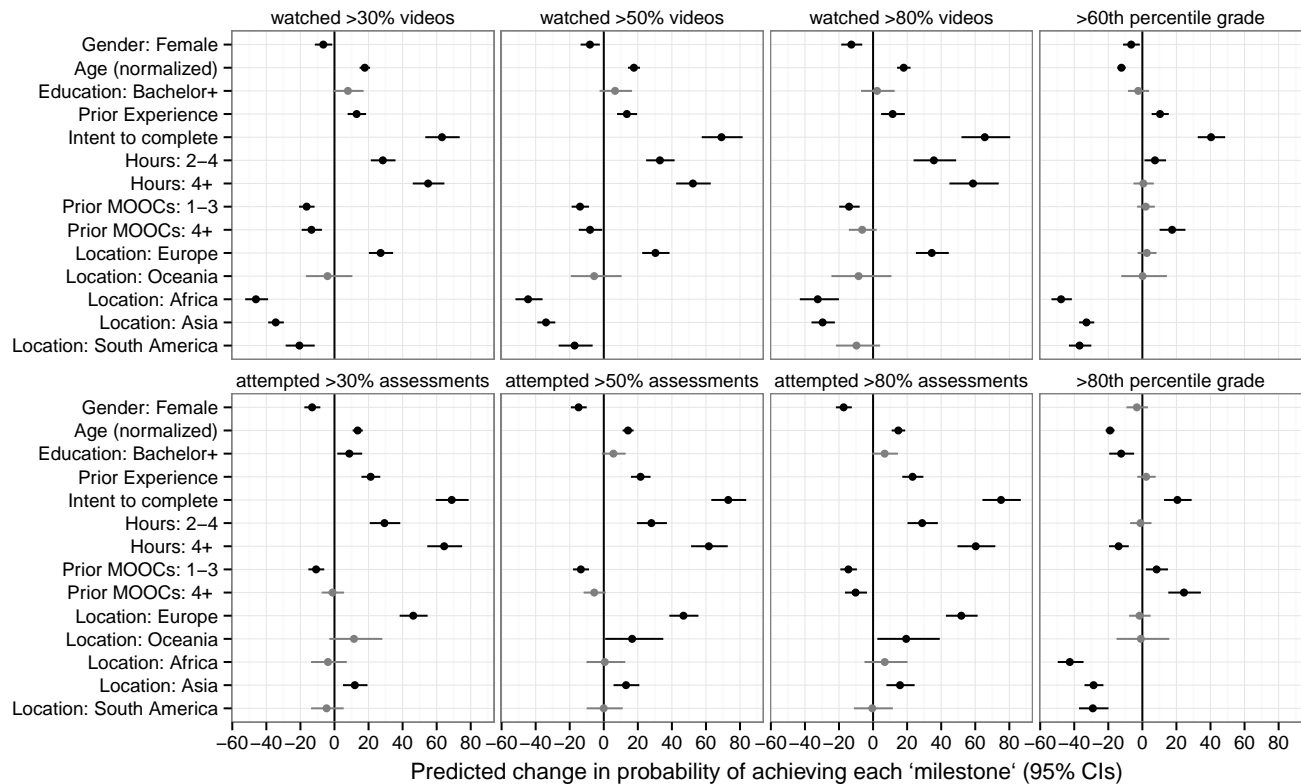
### Reasons for Attrition

Learners' reasons for disengaging from a MOOC were examined using survey responses (*RQ3a*). Across 17 MOOCs, learners selected reasons for disengaging from a pre-defined list. This list may not be collectively exhaustive, which is addressed in Study 2. The distribution of chosen reasons is illustrated in Figure 2. Around half of the respondents indicated two time-related reasons as influencing their decision to stop participating. Learners faced issues keeping up with deadlines and found that the course demanded too much time. Other reasons for attrition were less prevalent, but still reported by around 10 to 25% of respondents. Notably, just under 20% of respondents in a typical course (median proportion) stopped participating because they learned all they intended to learn.

To investigate the underlying factor structure of the reported reasons for disengaging, we conducted an exploratory factor analysis. The optimal number of factors was determined by the parallel analysis criterion and corresponding scree plot. The following five factors emerged: first, *general time issues*

<sup>2</sup>Multiple imputation and subsequent pooled analyses were performed with the *R* mice package.

<sup>3</sup>Data for 3 of the 17 MOOCs was unavailable due to anonymization.



**Figure 1.** Estimates from logistic mixed-effects model predicting persistence and performance milestones. An estimate of -10 indicates a 10% lower chance of achieving a milestone.

(course requires too much time; can't keep up with deadlines); second, *late start* (starting course late); third, *difficulty* (too advanced, exam too difficult); fourth, *format & content* (didn't enjoy online format, confusing videos, materials not explained well), and fifth, *goals & expectations* (only exploring, learned all I wanted, not what I'm looking for, not challenging, credential not valuable). The technical difficulties item failed to load onto any factor.

We expected reasons for disengaging from the course to vary by learner characteristics (*RQ3b*), such as learners' motivations for enrolling in the course. A correlation matrix of learners' enrollment intentions and their subsequent reasons for disengaging revealed a small number of notable associations. Taking the course with friends or colleagues was correlated with disengaging because the course was not challenging enough, the material was too advanced for the learner's background, and the credential was not valuable enough to the learner ( $r = .24, .20, .16$ , respectively). Not surprisingly, learners who enrolled to experience an online course tended to stop participating because they were only exploring the course ( $r = .17$ ). Learners looking to improve their English tended to experience technical difficulties ( $r = .17$ ), and those who enrolled for personal growth and enrichment tended to struggle to keep up with deadlines ( $r = .16$ ).

## Discussion

A number of striking findings emerged from the preceeding analyses. We found strong evidence for gender inequalities in attrition and, to some extent, performance. Women were less likely to persist, even when adding a number of other demographic, geographic, and preference indicators to the model, which highlights the robustness of the relationship. The MOOCs in our dataset were on diverse topics, including some on topics in the humanities and business with high rates of participation from women. Besides gender differences, attrition was substantially higher for learners from Africa, Asia, and South America, compared to North America, Oceania, and Europe, in terms of lectures watched. We also identified a large achievement gap in terms of grades between learners from these continents. The current generation of MOOCs is seemingly not immune to inequalities that have pervaded traditional educational settings. It remains unclear, however, to what extent these gaps result from differences in Internet access and language barriers, or potentially, from experiences of psychological threat to learners' identity or sense of social belonging [27].

Levels of self-reported satisfaction were generally high and did not vary much by level of persistence, although learners who reportedly disengaged from the course tended to be less satisfied than those who reportedly persisted until the end. Learners' reasons for disengaging from the course clustered into five key factors: general time issues; problems due to starting the course late; course difficulty; issues with the

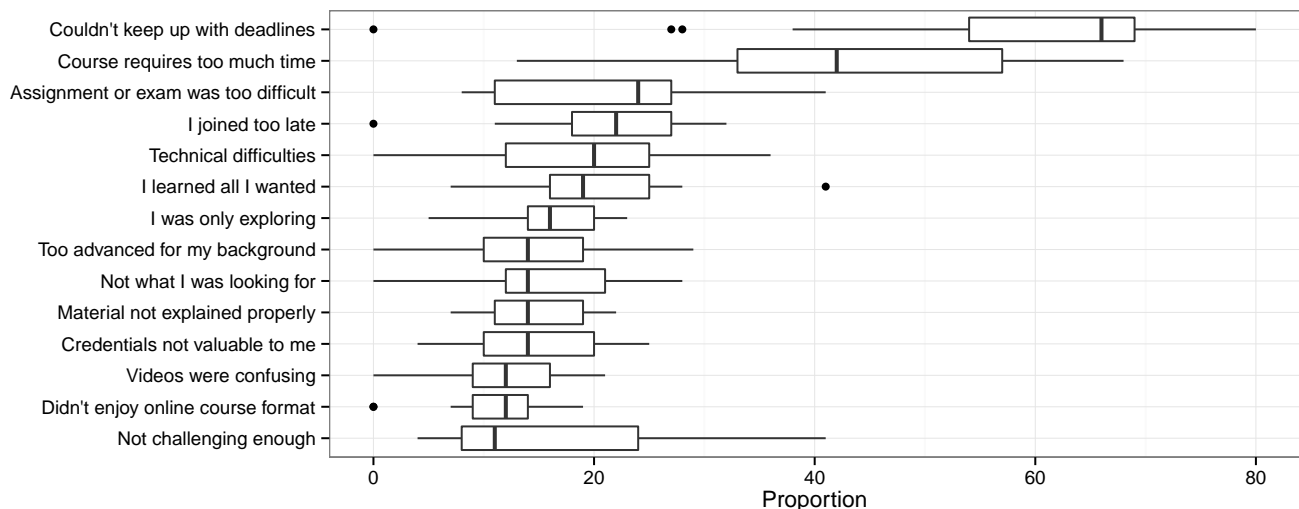


Figure 2. Proportion of learners who indicated each reason for attrition across 17 MOOCs

course format or content, and a mismatch or fulfillment of expectations. General time issues were overall the most pronounced. Attrition in community colleges has been interpreted as a sign of success in prior work. This view resonates in the MOOC setting, where around 20% of learners who disengaged reportedly had achieved their goal, and where progress in the course and learner satisfaction were only weakly related.

Learners may disengage from the course when they become too occupied by higher-priority tasks. Enrolling in a MOOC neither guarantees that a learner can or intends to allocate enough time to go through the course. According to expectancy-value theory, the learner is a rational agent who allocates time to different activities to maximize subjective gain. A low capacity for volitional control [5] could lead learners to misallocate time to lower-priority activities that promise nearer gratification, instead of allocating it to the next highest priority task. This implies that learners' general time issues may be a product of high-priority obligations as well as a result of a low capacity for volitional control.

## STUDY 2: UNDERSTANDING ATTRITION

To gain a better understanding of the attrition patterns identified in Study 1, we designed a smaller, more focused follow-up study. This study investigates reasons for attrition based on psychological measures and learners' own descriptions of the challenges they faced. We invited learners who were likely to drop out of a course to provide feedback. These learners were identified with a predictive model and invited to complete a survey. The survey included self-report measures of psychological constructs to test hypotheses *H1*, *H2*, and *H3* on the role of goal striving, social belonging, and growth mindset in the success of MOOC learners. Moreover, most learners indicated in Study 1 that they disengaged due to a shortage of time. A variety of factors could influence the amount of time learners have at their disposal. Higher priority obligations, such as professional, academic, and personal commitments

are presumably large contributors. Another critical factor, especially in the context of MOOCs as largely unguided learning environments, is learners' volitional control to engage in self-regulated learning [5]. Low volitional control could reduce learners' ability to allocate time for the course in the presence of potential distractors of subjectively lower importance than taking the course (e.g., watching TV). We constructed the following hypotheses about the role of volition in attrition due to a shortage of time:

**H4a** Learners who report having 'not enough time' can be grouped into those with higher priority obligations (professional, academic, and personal) and those with low volitional control.

**H4b** Learners who do not report specific higher-priority obligations tend to exhibit lower volitional control than those who specify particular obligations.

## Methods

The particular MOOC under observation was an undergraduate-level course on an advanced topic in computer science. It was offered in 2014 through Coursera. There were 20,048 enrolled learners; 10,510 watched at least one video, and 5,912 attempted at least one assignment. The following subsections describe how disengaging learners were identified, which measures were included in the feedback survey, and how open responses were iteratively coded.

### Predicting Disengagement

We constructed an 'early warning' prediction model to flag learners who had disengaged from the course or were about to disengage. The following three criteria were chosen as benchmarks to evaluate the practical and theoretical value of our prediction model.

**Prediction accuracy:** Prediction recall (fraction of correctly predicted disengaged learners) should be high to deliver the survey to as many disengaged learners as possible. The false

positive rate (fraction of active learners who are incorrectly predicted as disengaged) should be low to avoid sending surveys to learners who do not experience challenges in the course.

*Prediction lag:* The period between a learner's last site activity and the time they are flagged as disengaged should be short. A shorter prediction lag improves the likelihood of successfully intervening to help learners resume the course.

*Transferability:* Whether or not a learner disengages from the course cannot be determined before the end of the course. We thus trained the prediction model using data from previous courses. The performance of the prediction model should be good in courses other than those on which it was trained.

The training set was composed of learners' interaction data (predictors) and disengagement states (outcome) from 20 MOOCs. The large number of courses should improve model transferability by favoring features that are strongly predictive across different courses. Five hundred learners from each MOOC were included in the training set (*train*), and 500 other learners from the same MOOCs were selected for the first test set (*test*<sub>1</sub>). A second test set of similar size was composed from 20 other MOOCs (*test*<sub>2</sub>). Features extracted from the interaction data included different aspects of learners' video, assignment, and forum activity; their pace of viewing videos relative to the course pace, as well as various aspects of engagement time, such as the last time the learner interacted with the course (cf. [11], for additional information on the prediction model).

To control the length of the prediction lag, the model was only provided with learners' activity data until they were absent for 14 consecutive days. In the training phase, if the learner disengaged but the model did not detect it within 14 days of their last activity, the case was counted as a false negative. This construction results in a model with features that all fall within the defined prediction lag period. The model was fit by logistic regression with two-fold learner-level and course-level cross-validation. Prediction performance measured by the area under the ROC curve (AUC) very high (0.931 for *train*, 0.929 for *test*<sub>1</sub>, 0.922 for *test*<sub>2</sub>). Two of the predictors in the fitted model are highly influential. The first is the number of days since the learner has last been active. The likelihood that the learner drops out increases substantially if the learner disengages for 14 days or more. However, the likelihood of such a learner re-engaging increases with the fraction of released videos the learner has viewed prior to the absence; this is the second most influential feature in the model. The fact that almost identical results were obtained for *test*<sub>1</sub> and *test*<sub>2</sub> suggest good transferability of the model. A learner was flagged if their dropout probability exceeded 50%. Out of the 20,048 enrolled learners in the course, only those 10,510 were considered who watched at least one lecture video, and 6,074 of them were flagged as likely to drop out. The survey was sent out to 6,053 learners (excluding 21 learners with invalid email addresses or duplicate accounts).

## Feedback Survey

Every learner who was predicted to disengage from the course was sent an email kindly requesting their help: "You are enrolled in [course name], but you've been less active recently. Could you help us understand why?" A low response rate was expected for this subpopulation that was defined by low engagement. And yet, 756 out of 6,053 learners started the survey (12.5% response rate), and 459 completed it (61% completion rate). Missing values were multiply imputed by predictive mean matching using responses to all survey questions. All estimates were pooled across ten imputations, with standard errors adjusted for added variance. The majority of respondents were male (85%), 35.6 years old on average ( $SD = 12.5$ ), and 80% had achieved a bachelor's or more advanced degree.

Learners were asked to report how satisfied they were with their progress in the course, and whether they were using the course materials more, less, or exactly as much as they would have liked. These items served as measures of personal success. They were then asked to openly report "what challenges, inside or outside of the course, [they] experienced while [they were] taking this course, if any?" The instructions encouraged them to list all challenges they could think of. This question was asked prior to any survey questions that could suggest particular reasons for disengaging.

Two research assistants independently developed codebooks for the resulting 448 non-empty open responses. Their codebooks were consolidated and applied on Amazon Mechanical Turk (AMT) in three iterations. The codebook was updated in each iteration to reliably fit the open response data. Specific updates to the initial codebook were informed by code frequency, code correlations, and inter-coder agreement. In the first iteration, 250 randomly selected responses were each coded by four 'classification experts,' which is an AMT-specific qualification. In the second iteration, the remaining responses were coded in the same way. In the final iteration, all responses were coded by two classification experts (ties were resolved by two researchers).<sup>4</sup> The iterative coding yielded 399 relevant coded responses, which were analyzed in combination with the remaining survey data.

Learners self-reported social belonging (17 items on their sense of social and academic fit [27],  $M = 4.66$ ,  $SD = 0.71$ ,  $\alpha = 0.86$ ); mindset (4 items on the stability of intelligence and talent<sup>5</sup>;  $M = 4.44$ ,  $SD = 0.92$ ,  $\alpha = 0.77$ ), goal striving (4 items on motivation, perceived importance, commitment, and confidence;  $M = 3.12$ ,  $SD = 0.95$ ,  $\alpha = 0.82$ ), and capacity for volitional control (one item about lower-priority distractions hindering the learner's course progress;  $M = 2.73$ ,  $SD = 1.29$ ).

## Results

We further investigated research question *RQ3a* about reasons for disengaging from the course, and tested hypotheses *H4ab* about volitional control. In a second set of analyses, we

<sup>4</sup>For additional details on the coding procedure and codebooks, see [URL].

<sup>5</sup>Adapted from <http://mindsetonline.com/testyourmindset/>.

tested our hypotheses about the role of psychological factors in the self-ascribed success of MOOC learners.

#### Psychological Factors

Hypotheses *H1*, *H2*, and *H3* about goal striving, perceived social belonging, and mindset, respectively, were tested by comparing self-described successful with unsuccessful learners. Success was measured by relative progress and satisfaction with progress: 68% of respondents reported using the course materials less than they would have wanted, and 57% reported not being satisfied with their progress in the course (18% were very or extremely dissatisfied). The two measures of success were correlated,  $r = .35, t_{754} = 10, p < .001$ ; satisfied respondents were, however, equally likely to be successful and unsuccessful in terms of their progress. Table 2 provides summary statistics for each psychological measure by these two success metrics. Successful learners exhibited higher levels of goal striving according to both progress and satisfaction measures. Social belonging was significantly higher for learners who were successful in terms of satisfaction, but those who were successful in terms of progress. The pattern was reversed for growth mindset, such that learners who were successful in terms of progress (but not satisfaction) exhibited a stronger growth mindset. This pattern could reflect a difference in the underlying process by which these psychological factors affect learner behavior and perception. The observed differences are expected to be larger in the full population of learners, compared to the small, homogenous sample in this study, consisting of learners unlikely to persist.

#### Reasons for Dropout

The majority of survey respondents (550 of 756; 73%) reported being only somewhat or less satisfied with their progress in the course. While many of them (66%) were substantially held back by other commitments that took up time they had planned to spend on the course, only one in four (25%) indicated being hindered by distractions of lower priority than the course (i.e. exhibit lower volitional control). These two time-related measures, which only differ in the assigned level of priority, were not significantly correlated ( $r = .05, t_{548} = 1.3, p = .21$ ). Reasons for learner attrition were measured using iteratively coded open responses. Table 3 provides each coded reason and percentage of open responses that indicated it. A number of significant correlations between these challenges stood out: having not enough time was negatively associated with disliking the format or teaching style ( $r = -.52$ ), needing additional course materials ( $r = -.26$ ), finding the course uninteresting or not valuable ( $r = -.45$ ), and facing technical difficulties ( $r = -.29$ ). Hence, if a learner had not enough time, this tended to be the primary challenge.

To test hypotheses *H4ab*, we examined the 336 open responses that indicated having too little time. It revealed that no particular commitment was specified in about half of them (53%). This distinction was a salient feature in the initial codebooks, but could not be coded with sufficient reliability across coders. Coders would have required additional training to avoid reading specific reasons into un-specific responses. We thus operationalized the distinction

by string matching to capture professional, academic, and personal reasons for having not enough time.<sup>6</sup> Relative to those who specified a reason, learners who indicated too little time without specifying a reason also reported a significantly lower level of volitional control, i.e. their progress was more severely hindered by lower-priority distractions ( $t_{331} = 2.9, p = .003, d = .32$ ). However, the two groups did not significantly differ in how much they were held back by commitments that took up time they had planned to spend on the course ( $t_{333} = -.75, p = .45$ ). These findings lend strong support to hypotheses *H4ab*: that is, while the large group of learners who reported having not enough time generally faced higher-priority obligations, a substantial proportion suffered from low volitional control—and reporting no specific reasons for having not enough time was indicative thereof.

**Table 3. Iteratively-Coded Reasons for Attrition ( $n = 399$ )**

What challenges have you experienced while taking this course?	%
Not enough time for the course	84
The course format or teaching style is not a good fit	18
The level of the course is too advanced	10
The course is not interesting or valuable enough	7
Additional course materials needed to learn the topic	4
Technical difficulties	2
Language barrier	1

#### Discussion

In this follow-up study, we examined the role of psychological factors in attrition and further investigated reasons for attrition. Personal success in a MOOC was associated with three established psychological factors: goal striving, social belonging, and growth mindset. Prior work in more traditional achievement-oriented environments found reduced levels of these psychological factors to result in lower persistence and performance. The current observational findings highlight the critical role of these factors in an online learning environment. Learners with higher levels of goal striving, social belonging, and growth mindset reported greater personal success (*H1*, *H2*, *H3*). An online learner with, for instance, a strong sense of social belonging may derive greater satisfaction from the course and persist for longer than one with a weak sense of belonging. These psychological factors tend to vary across people based on prior experiences and they can change as a result of contextual cues. Psychological factors could thereby, at least partially, account for individual differences in persistence and performance (Study 1).

The methodology used in this study can be extended to deliver timely and targeted interventions. The predictive model was successful in identifying a large number of learners who were facing challenges with taking the course. The majority of them were unsuccessful in their own terms. All three psychological factors under investigation were found to distinguish self-ascribed successful from unsuccessful learners. Recent work on psychological interventions found lasting positive effects of brief interventions involving reading and writing tasks (e.g., [27]). Cultivating a growth mindset was found to

<sup>6</sup>Pattern: `work|job|school|university|college|family|kids|health`



**Table 2. Descriptive and inferential statistics for psychological measures by learner success**

	<i>N</i>	Goal Striving ( <i>H1</i> )	Social Belonging ( <i>H2</i> )	Growth Mindset ( <i>H3</i> )
<i>Relative Progress</i>				
Successful	241	3.46 ( <i>SD</i> = 0.91)	4.70 ( <i>SD</i> = 0.74)	4.58 ( <i>SD</i> = 0.99)
Unsuccessful	515	2.96 ( <i>SD</i> = 0.93)	4.63 ( <i>SD</i> = 0.70)	4.38 ( <i>SD</i> = 0.87)
		$t_{201} = 6.3, p < .001, d = 0.55$	$t_{60} = 1.0, p > .25, d = .09$	$t_{152} = 2.4, p = .017, d = 0.21$
<i>Satisfaction</i>				
Successful	325	3.43 ( <i>SD</i> = 0.83)	4.76 ( <i>SD</i> = 0.70)	4.47 ( <i>SD</i> = 0.93)
Unsuccessful	431	2.88 ( <i>SD</i> = 0.97)	4.58 ( <i>SD</i> = 0.71)	4.42 ( <i>SD</i> = 0.91)
		$t_{166} = 7.2, p < .001, d = 0.59$	$t_{158} = 3.1, p = .002, d = .26$	$t_{175} = 0.6, p > .25, d = .05$

help community college students persist in prior work, and our findings provide correlational evidence that subjectively successful learners hold a more fluid notion of intelligence than their less successful peers. Leveraging predictive models of persistence, brief interventions could be delivered on-line and targeted to learners who face particular challenges, such as low goal striving, a fixed mindset, or perceived threat to their social belonging.

Through iterative coding, we arrived at a list of challenges faced by a subset of learners in a particular MOOC (Table 3). The list maps onto the factors identified in Study 1 (general time issues, late start, difficulty, format & content, expectations, and technical difficulties), with the additional challenge of a language barrier, which was not included as a response option in Study 1. The majority of learners indicated that they had not enough time: some struggled due to higher priority obligations that took up too much time, while others were hindered by low levels of volitional control (*H4a*). A learner needs to be highly self-regulated to be successful in a MOOC. Very few learners explicitly indicated issues with volitional control in their open responses. Learners may not possess the metacognitive skills to consciously notice this issue or they preferred not to admit to this undesirable trait. Learners who reported a shortage of time without providing specific reasons were more likely to have insufficient volitional control (*H4b*). This could be a useful proxy for identifying issues with volitional control in future work. Overall, the results suggest that learners' capacity for volitional control can be a critical ingredient for learners' success and warrants further investigation [5].

## GENERAL DISCUSSION

This study of over 60,000 online learners, combining behavioral and survey data, extends the empirical base of research on attrition in online learning environments. The three major findings are, *first*, the presence of substantial gaps in persistence and performance between learners with different demographic features and from different geographic locations; *second*, the relation between personal success in online learning (by learners' own judgment) and higher levels of goal-striving, social belonging, and growth mindset, and *third*, the challenges faced by online learners, with the shortage of time as the major obstacle, which was frequently related to learners' low capacity for volitional control. Moreover, we developed a survey item for measuring reasons for disengaging from a course (cf. Table 3), which should be tested and potentially extended based on more empirical evidence from additional courses.

In addressing research questions and testing hypotheses which emerged from a review of the literature, we sought to advance theory on attrition and provide empirical evidence. In line with early models of attrition that emphasized the importance of goal-directed behavior, we found goal striving to be higher in subjectively successful than unsuccessful learners. Moreover, volitional control was identified as a major contributor in reported time-related challenges with taking a MOOC. Tinto's [24] classic model of attrition emphasizes the role of academic and social integration, and incorporates individual learner characteristics as contributors to learners' decision to persist in a course of study. Our findings provide strong support for the critical role of individual characteristics, such as gender, age, intentions, experiences, and geographic location. Furthermore, we found evidence for the importance of social integration in terms of perceived social belonging [27]. In contrast to Tinto's concept of social and academic integration, which characterizes a general state in an institution, social belonging is a subjective experience based on perceptions and interpretations of environmental cues. Yet, this subjective experience is typically shared by individuals with a common identity; for example, female learners in male-dominated fields or black minority learners in college. Models of attrition should be updated to acknowledge the potential impact of psychological threats, which can cause achievement gaps that would otherwise be overlooked or misinterpreted.

There are many practical implications of this research and the one we would like to emphasize is the need for and feasibility of targeted psychological interventions in online learning environments. While the current work does not provide explanations for the presence of substantial achievement gaps, we can speculate that similar psychological processes as in traditional educational environments are involved. Brief and targeted online interventions (mindset, social belonging, goal striving, volitional control, etc.) can potentially narrow existing achievement gaps in online learning. More research on prediction models that identify struggling learners and targeted interventions to support them can make a big difference in online learning.

## ACKNOWLEDGMENTS

We thank Candace Thille, Emily Schneider, and Andreas Paepcke for their helpful feedback, and Elise Ogle and Ruth Bram for helping with the development of the codebook. The authors contributed equally to this work.

## REFERENCES

1. Bean, J. P., and Metzner, B. S. A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research* 55, 4 (1985), 485–540.
2. Bers, T. H., and Smith, K. E. Persistence of community college students: The influence of student intent and academic and social integration. *Research in Higher Education* 32, 5 (1991), 539–556.
3. Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., and Seaton, D. Studying learning in the worldwide classroom: Research into edxs first mooc. *Research & Practice in Assessment* 8 (2013), 13–25.
4. Clow, D. Moocs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM (2013), 185–189.
5. Corno, L. Volitional aspects of self-regulated learning. In *Self-regulated learning and academic achievement: Theoretical perspectives*, D. H. S. Barry J. Zimmerman, Ed. Routledge, 2001, ch. 6, 191–225.
6. Corno, L., and Kanfer, R. The role of volition in learning and performance. *Review of research in education* (1993), 301–341.
7. Dweck, C. *Mindset: The new psychology of success*. Random House LLC, 2006.
8. Fishbein, M., and Ajzen, I. *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley Publishing Company, 1975.
9. Goldrick-Rab, S. Challenges and opportunities for improving community college student success. *Review of Educational Research* 80, 3 (2010), 437–469.
10. Gütl, C., Rizzardini, R. H., Chang, V., and Morales, M. Attrition in mooc: Lessons learned from drop-out students. In *Learning Technology for Education in Cloud. MOOC and Big Data*. Springer, 2014, 37–48.
11. Halawa, S., Greene, D., and Mitchell, J. Dropout prediction in moocs using learner activity features. *eLearning Papers* (2014), 1–10.
12. Khalil, H., and Ebner, M. Moocs completion rates and possible methods to improve retention-a literature review. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, vol. 2014 (2014), 1305–1313.
13. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, ACM (2013), 170–179.
14. Kizilcec, R. F., and Schneider, E. Motivation as a lens to understand online learners: Towards data-driven design with the olei scale. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 2 (in press), 24.
15. Lewin, T. After setbacks, online courses are rethought. *New York Times* (2013, December).
16. Moore, M. G., and Kearsley, G. *Distance education: A systems view*. Wadsworth Publishing Company, 1966.
17. Morris, L. V., Finnegan, C., and Wu, S.-S. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education* 8, 3 (2005), 221–231.
18. Muilenburg, L. Y., and Berge, Z. L. Student barriers to online learning: A factor analytic study. *Distance Education* 26, 1 (2005), 29–48.
19. Parr, C. Mooc completion rates ‘below 7%’. *Times Higher Education* (2013, May).
20. Paunesku, D., Yeager, D., Romero, C., and Walton, G. A brief growth mindset intervention improves academic outcomes of community college students enrolled in developmental mathematics courses. *Unpublished manuscript, Stanford University, Stanford, CA* (2012).
21. Rivard, R. Measuring the mooc dropout rate. *Inside Higher Ed* 8 (2013).
22. Rovai, A. P. In search of higher persistence rates in distance education online programs. *The Internet and Higher Education* 6, 1 (2003), 1–16.
23. Taylor, C., Veeramachaneni, K., and O’Reilly, U.-M. Likely to stop? predicting stopout in massive open online courses. *arXiv preprint arXiv:1408.3382* (2014).
24. Tinto, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* (1975), 89–125.
25. Tyack, D., and Tobin, W. The “grammar” of schooling: Why has it been so hard to change? *American Educational Research Journal* 31, 3 (1994), 453–479.
26. Waldrop, M. Online learning: Campus 2.0. *Nature* 495, 7440 (2013), 160–163.
27. Walton, G. M., and Cohen, G. L. A question of belonging: race, social fit, and achievement. *Journal of personality and social psychology* 92, 1 (2007), 82.
28. Wilkowsky, J., Deutsch, A., and Russell, D. M. Student skill and goal achievement in the mapping with google mooc. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 3–10.
29. Yang, D., Sinha, T., Adamson, D., and Rosé, C. P. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop* (2013).