

# Attrition in Online Learning: Understanding Persistence and Dropout in Massive Open Online Courses

Omitted for blind review

Institution

Address

Email

## ABSTRACT

Abstract goes here.

## Author Keywords

Online learning, persistence, attrition, dropout, disengagement, massive open online courses, MOOC, psychological factors

## 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 and by accommodating people with less flexible schedules [9]. Distance learning programs intended to deliver education in remote parts of the world and for 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 [19].

The latest generation of online learning environments, characterized by massive open online courses (MOOCs), has pushed the boundary on the scale of education [29]. 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 grading [17, 4]). 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 [15]. Many of the prototypical behaviors observed in MOOCs [15, 3] (other?) resemble those on online media platforms, such as YouTube or tumblr. This trend has also been reflected in the diversity of MOOC learners’ motivations for enrolling [16].

Shortly after the first wave of courses had finished, extensive media coverage led to MOOCs becoming associated with high attrition rates [18, 22, 10]. Early MOOC research cautioned against dichotomizing learners into successes and failures based on course completion [15, 24]. Instead, more nuanced categorizations based on learner behavior [15, 5] (other?), motivations [16], or intentions [31] have been suggested. Ultimately, perspectives on persistence and attrition in MOOCs depend on how MOOCs have been conceptualized. Kizilcec and Schneider [16] proposed that MOOCs have bridged two different worlds: one world is governed by the user-centric norms of online media, where everyone is encouraged to be as active as they wish; the other world adheres to the “grammar of schooling”, which presupposes instructor-defined goals that students strive to achieve [28]. Viewing MOOC participation as bridging these two worlds undoubtedly adds a layer of complexity to interpretations of attrition.

This paper presents a systematic investigation into attrition in MOOCs, based on self-report and behavioral data collected from [N] online learners in [M] courses. 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, we sought to develop a deeper understanding of attrition in MOOCs by conducting a case study. Targeting around 7,000 online learners who were predicted likely dropouts in a particular course were invited to provide feedback via a survey. TODO: complete paragraph based on what we actually do in Study 2.

## RELATED WORK

Research and theorizing on attrition in education has a rich history. This review is intended to serve as a foundation to

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...\$15.00.  
DOI string from ACM form confirmation

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 developed 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. 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, i.e., the extent to which the student engages in goal-directed behaviors in the face of distraction [7]. Hence, motivation alone is necessary but not sufficient for persistence. Students may fail to sustain efforts in the absence of strong self-regulatory skills.

**RQ1** How do prior learner characteristics (demographics, geographic location, intentions and motivations, and previous online courses) influence learners' likelihood of dropping out of a MOOC?

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

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 [27] 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 [30].

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

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 (i.e., instilling a growth mindset) was found to halve attrition rates and increase academic achievement [23]. The results highlight the powerful influence of mindset and caution against overinterpreting non-persistence as an indicator of goal achievement.

**RQ2** What proportion of learners who dropped out were satisfied with their achievements in the course?

**H3** Successful MOOC learners have a more fluid conception of intelligence (i.e., more of a growth than fixed mindset) than unsuccessful ones.

### **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 (e.g., demographics), academic and environmental variables (e.g., study habits, financial resources, work and family obligations), and academic and psychological outcomes (e.g., GPA, satisfaction). The significant change from Tinto's original model was the inclusion of environmental variables to account for the added complexity of nontraditional students' lives (see [13], for another adaptation of Tinto's model for distance education).

Rovai [25], 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, while the theoretical models become more developed, they became harder to apply to real-world settings, and hence, the empirical evidence to support them remained sparse. An exploratory study of reasons for attrition suggested the following eight constructs, based on over 1,000 online education students: academic and technical skills, learner motivation, time and support for studies, cost and access to the Internet, and technical problems [21]. 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 [20].

### **MOOCs**

A survey of over one 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 [11]. 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 [14]. 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 [32]. In

**Table 1. Summary statistics of learner demographics and behavior in 14 MOOCs**

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

<sup>1</sup> mean (SD) <sup>2</sup> bachelors' or higher degree <sup>3</sup> self-reported in the feedback survey

addition, a number of machine learning approaches yielded promising results for early dropout prediction (e.g. [26, 12]), but fell short of offering insights into why a given learner would leave the course.

**RQ3a** What reasons do learners report for dropping out of a MOOC?

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

**RQ4** How does behavior in MOOCs vary by reasons for dropping out?

### STUDY 1: REASONS FOR DISENGAGING

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 seventeen 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 (see Table ). 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. Table provides summary statistics for enrollment, survey response rates, demographics, self-reported dropout, and behavioral indicators for each course. Self-reported dropout was measured in the feedback survey by asking learners if they stopped participating in the course. The three behavioral indicators are the proportions of learners who viewed at least 30, 50, and 80% of the lecture videos.

Information about age, gender, and education level were only available for learners who completed the initial course survey.

### Measures

Learners reported reasons for enrolling in the course (using the OLEI scale [16] on the course survey) and leaving the course (on the feedback survey) on a grid of binary-choice items. The feedback survey for asked learners whether or not they have stopped participating before the end of the course, and if they did, what reasons influenced their decision to stop. For each reason, the student was asked to select one of two options: "Did influence my decision to stop" and "Did NOT influence my decision to stop". Compared to a question where the student only checks a box if the reason applies, this design aims to distinguish between respondents who skip reporting reasons and those who express that the reason did not apply to them. We also computed the fraction of videos in the course each learner had watched to relate self-report measures to learner behavior.

We first report the proportion for each reason for dropping out. This was computed as the number of respondents who reported that the reason applied to them divided by the total number of those who responded to the question (reporting that it applies or that it does not apply). We also computed the median and interquartile range of the proportions across all courses for each reason. The resulting proportions are shown in Figure 1. Second, we performed three correlational analyses were performed to investigate associations between these measures.

We examined the strength of correlations between reasons for dropout. The correlation matrix for the reasons were computed separately for each course, and then linearly combined to yield the average correlation matrix shown in Table . Following the same procedure, we computed the correlation matrix between reasons to enroll and reasons for attrition. Furthermore, we computed correlations between reported reasons for dropout and certain learner behaviors. Finally, for each learner, we computed the fraction of course

videos visited (the learner started playing the video) and fraction watched (the learner stayed with the video for 3 minutes, or for the whole video duration if shorter than 3 minutes). We then computed the matrix of correlations between reasons for dropout and these variables for each course. These matrices were averaged and the result is shown in Table [TableToBeAdded].

## Results

(Note: this section is work in progress, we will go through each RQ and present results that address them here.)

Table .

How do prior learner characteristics (demographics, geographic location, intentions and motivations, and previous online courses) influence learners' likelihood of dropping out of a MOOC? (RQ1)

What proportion of learners who dropped out were satisfied with their achievements in the course? (RQ2)

What reasons do learners report for dropping out of a MOOC? (RQ3a)

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

How does behavior in MOOCs vary by reasons for dropping out? (RQ4)

The two most commonly reported reasons for dropout are related to one's ability to commit time to the course. The correlation between the two reasons is  $r = 0.32$ . A more detailed analysis we conditioned on responses to these two reasons revealed that 30% of respondents who reported at least one of these two time reasons reported no other reasons for dropout, while 24% reported one other reason and 46% reported more than one.

Difficulty of an assignment or exam was the third most reported reason for dropout. It is most positively correlated with reasons citing difficulties, including "Too advanced for my background" ( $r = 0.4$ ), "Materials not explained properly" ( $r = 0.37$ ), and "Technical difficulties" ( $r = 0.33$ ).

The proportion of respondents who selected "Technical difficulties" exceeded 20% in 10 out of the 17 courses. This reason, however, exhibits positive correlations stronger than 0.2 with 7 other reasons. [What might we like to study more about respondents who selected this reason?]

In the median, 19% of respondents reported they learned what they wanted in the course, and this reason was moderately positively correlated with reasons citing no or incomplete interest in the course, namely "I was only exploring the course" ( $r = 0.36$ ), "The course is not challenging enough for me" ( $r = 0.31$ ), and "The course is not what I was looking for" ( $r = 0.28$ ).

Another set of interesting positive correlations exists between "Course is not what I was looking for" and other reasons citing the materials' level of difficulty/challenge, the value of

the credentials to the respondent, and the extent to which she enjoyed the online course format, suggesting .... [TBC]

The reason that exhibited the strongest positive correlations with other features was "Videos were confusing". it correlates most strongly with "Materials not explained properly" ( $r = 0.66$ ) and "Did not enjoy online course format" ( $r = 0.54$ )

Unlike correlations among reasons for disengagement, correlations of reasons for disengagement with enrollment intentions were mostly weak ( $-0.14 < r < 0.24$ ). Among the strongest positive correlations were the cases listed in Table ??

## Discussion

(Note: this section is also work in progress.)

Even though study 1 accentuates "time" as a reported reason for dropout, interpreting what time means is not straight forward. Expectancy-value theory models the learner as a rational agent who allocates time to different activities to maximize subjective gain. In other words, a learner might disengage from the course if she gets occupied by other higher priority tasks. Because learners can join or leave MOOCs freely. Joining a MOOC does not necessarily imply that a learner is capable of allocating the time for it according to the rational framework that expectancy value theory predicts. Moreover, previous research on volitional control suggests that learners allocate time to the next highest priority task, but could rather get distracted by lower priority activities that promise nearer gratification. So what does lack of time for the course mean? In study 2, we attempt to investigate various possible dropout reasons that learners might report as "lack of time".

## STUDY 2: UNDERSTANDING DISENGAGEMENT

To gain a better understanding of the attrition patterns identified in Study 1, we designed a smaller, more focused follow-up study. In Study 2, we reached out to learners who were likely to drop out of a course and asked them to provide feedback. To this end, we developed a predictive model to identify at-risk learners and asked them to complete a survey. One goal of this study was to further investigate research question RQ3a about reasons for attrition based on learners' own descriptions of the challenges they faced. Specifically, in Study 1, most learners reported that they disengaged due to a shortage of time. A variety of factors can influence the amount of time learners have at their disposal. Higher priority obligations, such as professional, academic, and personal commitments, are certainly 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 [6]. For instance, 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 presence of these two factors for learners who report having not enough time:

**H4a** Learners who report having 'not enough time' can be grouped into those with higher priority obligations (pro-

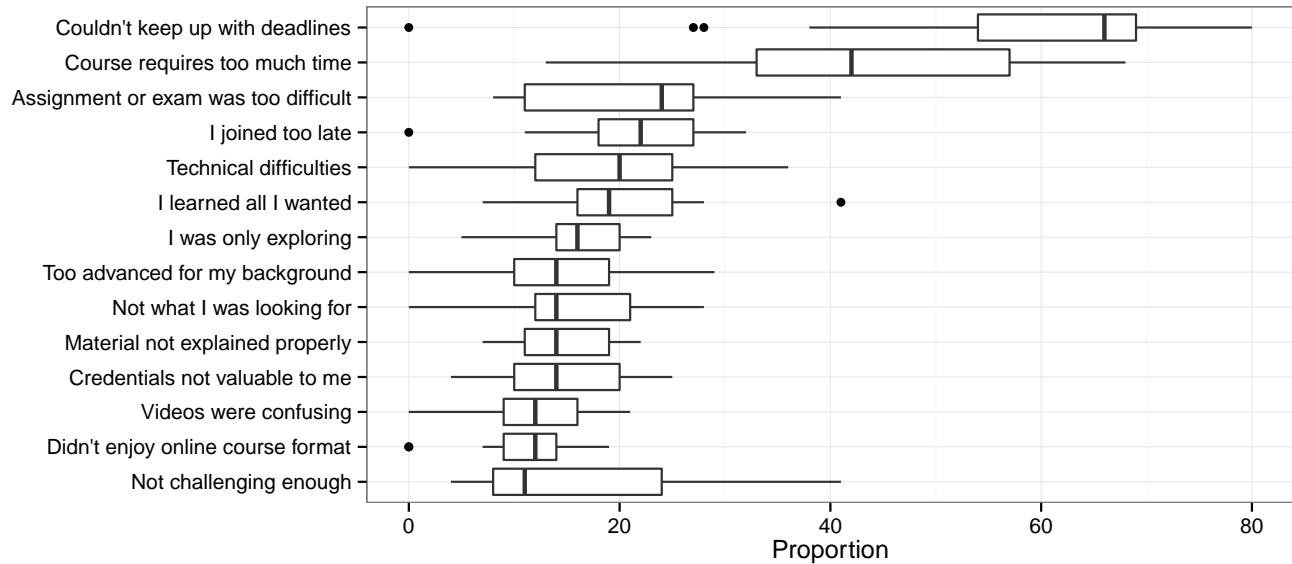


Figure 1. Proportion of learners who indicated each reason for attrition across 17 MOOCs

Table 2. Correlations among reasons for disengagement based on a linear combination of the courses' correlation matrices

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
R1: Assn or exam was too difficult	1.0	0.17	0.28	0.3	0.16	0.14	0.37	0.24	0.26	0.2	0.09	0.33	0.4	0.28
R2: Couldn't keep up with deadlines		1.0	0.06	-0.03	0.01	-0.05	0.05	-0.07	-0.06	0.32	0.17	0.04	0.08	0.04
R3: Credentials not valuable to me			1.0	0.34	0.34	0.25	0.34	0.37	0.39	0.17	0.17	0.27	0.28	0.35
R4: Didn't enjoy online course format				1.0	0.24	0.15	0.53	0.36	0.44	0.02	0.18	0.34	0.33	0.54
R5: I was only exploring					1.0	0.36	0.2	0.26	0.31	0.11	0.18	0.21	0.16	0.18
R6: I learned all I wanted						1.0	0.12	0.31	0.28	0.0	0.12	0.16	0.16	0.09
R7: Material not explained properly							1.0	0.36	0.45	0.15	0.13	0.31	0.35	0.66
R8: Not challenging enough								1.0	0.45	0.0	0.12	0.28	0.21	0.32
R9: Not what I was looking for									1.0	0.07	0.12	0.3	0.33	0.44
R10: Course requires too much time										1.0	0.06	0.08	0.19	0.12
R11: I joined too late											1.0	0.22	0.11	0.14
R12: Technical difficulties												1.0	0.27	0.3
R13: Too advanced for my background													1.0	0.33
R14: Videos were confusing														1

fessional, 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.

Another goal of Study 2 was to collect data to test hypotheses *H1*, *H2*, and *H3* on the role of goal striving, social belonging, and growth mindset in the success of MOOC learners. To this end, we included a self-report measures for each construct in the survey that was sent to at-risk learners.

## 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 [N] attempted more than one assignment.

### Predicting Disengagement

The objective was to build an 'early warning' prediction model that could flag learners who are likely to drop out in the future. We defined the following three criteria by which to evaluate the performance of the model.

**Prediction accuracy:** Prediction recall (the fraction of correctly predicted dropouts) should be high to deliver the survey to as many future dropouts as possible. The false positive rate (the fraction of non-dropouts incorrectly predicted as dropouts) should be low to avoid sending a survey to learners who do not experience challenges in the courses.

**Prediction lag:** The period between a learner's last site activity and the time they are flagged as being likely to drop out should be short. A shorter prediction lag improves the likelihood of successfully intervening to help the learner get back on track.

**Transferability:** Whether or not a learner drops out of the course can not be determined before the end of the course. We thus had to train the prediction model using data from previous courses. The performance of the prediction model

**Table 3. Correlations between reasons for disengagement and enrollment intentions based on a linear combination of the courses’ correlation matrices**

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Career change	0.03	0.02	-0.11	-0.03	-0.02	-0.06	0	-0.01	0	0	0.03	0.04	0	-0.03
Experience online course	0.06	0.01	0.01	0.1	0.17	0.09	0.02	0.01	-0.01	-0.02	0.05	0.12	0.12	0.08
Fun and challenge	0.05	0.07	-0.03	-0.01	0.02	-0.05	-0.01	0	-0.08	0.06	0.02	0.01	-0.01	0
General interest	0.01	-0.04	-0.04	0.02	-0.01	-0.07	-0.05	0.01	0.04	0.03	0.06	-0.03	-0.01	0.02
Improve my English	0.08	0	0.05	0.1	0.09	0.07	0.02	0.03	0.04	0	0.1	0.17	0.08	0.05
Presitigious uni/prof	0.03	0.15	-0.08	-0.07	-0.04	-0.06	-0.07	-0.1	-0.14	0.03	0.02	0	-0.03	-0.05
Relevant to job	-0.04	-0.01	0.02	0.05	0.04	0	0.01	-0.02	0	0	-0.03	0.09	0.02	0.01
Meet new people	0.05	-0.02	0.05	0.08	0.06	-0.04	0.03	0.13	-0.01	0.01	0.14	0.15	0.06	0.03
Other	0	-0.07	0.08	0.05	0	-0.03	0.03	0	0.01	-0.08	0	0.09	0.15	0
Personal growth	0.02	0.16	0	-0.03	0.02	-0.02	0.07	-0.01	0	0.03	0.1	0	-0.03	-0.02
Relevant to research	-0.02	0.01	0.1	0.06	0.05	0	0.02	0.03	-0.01	-0.02	0.13	0.1	-0.01	0.07
Relevant to school	-0.06	0	0.08	0.01	0.08	0.02	-0.01	0.02	-0.05	-0.04	0.08	0.02	-0.02	-0.01
Statement of accomp.	0.08	0.1	0.02	0.03	0.01	-0.02	0.08	0	0.02	0	0.06	0.09	0.14	0.02
Take with others	0	-0.02	0.16	0.06	0.12	0.05	0.06	0.24	0.11	0.04	0.09	0.15	0.2	0.08

should be adequate in courses other than those on which it was trained.

The training set was composed of learners’ interaction data (predictors) and dropout 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 (see [12], for additional details on the prediction model). The model was fit by logistic regression. Table 4 provides information on the prediction performance, measured by the area under the ROC curve (AUC).

**Table 4. Performance of the dropout prediction model**

	<i>train</i>	<i>test</i> <sub>1</sub>	<i>test</i> <sub>2</sub>
AUC	0.931	0.929	0.922

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 ( $x_1$ ). 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 ( $x_2$ ) prior to the absence; this is the second most influential feature in the model. Equation 1 is an approximation of the model with only these two predictors. A learner was flagged if their dropout probability exceeded a half:

$$Pr(\text{dropout}) = \frac{1}{1 + e^{-0.431x_1 + 6.388x_2 + 1.626}} > 0.5 \quad (1)$$

Out of the 20,048 enrolled learners in the course, we only considered those 10,510 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.<sup>1</sup> 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 bachelors’ 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 deliberately asked prior to any survey questions that could suggest particular reasons for disengagement.

Two research assistants independently developed codebooks for the resulting 448 non-empty open responses. Their codebooks were consolidated and applied on Mechanical Turk 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’. 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

<sup>1</sup>The *R* mice package was used to perform multiple imputation and subsequent pooled analyses.

resolved by two researchers).<sup>2</sup> The iterative coding yielded 399 relevant coded responses, which were analyzed in combination with the remaining survey data.

Learners also rated their sense of social and academic fit [30] (17 items,  $M = 4.66$ ,  $SD = 0.71$ ,  $\alpha = 0.86$ ); mindset (4 items on the nature of intelligence and talent<sup>3</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

### 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 ( $n = 399$ ). Table 5 provides each coded challenge 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 ( $r = -.29$ ).

**Table 5. Iteratively-Coded Reasons for Attrition**

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

To test hypotheses  $H4a$  and  $H4b$ , 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>4</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  $H4a$  and  $H4b$ : that is, while the large group of learners who report 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.

### 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 6 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 not progress. The pattern was reversed for growth mindset, with evidence for the hypothesized effect for progress but not satisfaction. Hence, the results fully support hypothesis  $H1$ , and partially support  $H2$  and  $H3$ .

## Discussion

Study 2 focused on reasons for attrition and psychological factors of learners who were likely to drop out of the course. Through iterative coding, we developed a set of challenges that were faced by a subset of learners in a particular MOOC (cf. Table 5). The large majority of learners who indicate that they had not have enough time was found to be splorable into two groups: those with higher priority obligations that take up too much time, and those who are challenged by low levels of volitional control ( $H4a$ ). MOOCs are an environment that requires learners to be highly self-regulated to be successful. The present findings emphasize the critical role of their capacity for volitional control, a factor that warrants further investigation [6]. Very few learners explicitly indicated issues with volitional control in their responses, potentially because it is not socially desirable to admit. The absence of specific reasons for their shortage of time was found indicative of issues with volitional control ( $H4b$ ). This could be a useful proxy for volitional issues in future research.

We investigated the extent to which personal success in MOOCs is associated with three established psychological factors: goal striving, social belonging, and growth mindset.

<sup>2</sup>For additional details on the coding procedure and codebooks, see [URL].

<sup>3</sup>Items were adapted from the mindset questionnaire available at <http://mindsetonline.com/testyourmindset/>.

<sup>4</sup>Pattern: *work|job|school|university|college|family|kids|health*

**Table 6. 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)
		$t_{166} = 7.2, p < .001, d = 0.59$	$t_{158} = 3.1, p = .002, d = .26$	$t_{175} = 0.6, p > .25, d = .05$

Prior evidence found the absence of each factor to cause lower persistence and performance in more traditional achievement-oriented environments. The present findings highlight the critical role of these factors in MOOCs based on observational evidence. Learners with higher levels of goal striving, social belonging, and growth mindset also reported higher levels of personal success (*H1, H2, H3*). The subpopulation of learners in this study was a small and intentionally skewed subset. The observed differences are expected to be larger in the full population of learners, compared to this homogeneous sample of learners who are unlikely to persist. Goal striving was most strongly associated with both measures of personal success. Success in terms of relative progress was associated with a stronger growth mindset, but not significantly with social belonging. In contrast, success in terms of satisfaction with progress was associated with stronger social belonging, but not significantly with growth mindset. Future work should examine if this pattern reflects a difference in the underlying process by which these psychological factors affect learner behavior and perception.

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., [30]). Leveraging predictive models of persistence, brief interventions could be delivered online and targeted to learners who face particular challenges, such as low goal striving, a fixed mindset, or perceived threat to their identity or social belonging.

## GENERAL DISCUSSION

## CONCLUSION

## ACKNOWLEDGMENTS

Omitted for blind review.

## 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. Cambre, J., Kulkarni, C., Bernstein, M. S., and Klemmer, S. R. Talkabout: small-group discussions in massive global classes. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 161–162.
5. Clow, D. Moocs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM (2013), 185–189.
6. 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.
7. Corno, L., and Kanfer, R. The role of volition in learning and performance. *Review of research in education* (1993), 301–341.
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. Guthrie, D. Moocs are toast or at least should be. *Forbes* (2013, July).
11. 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.
12. Halawa, S., Greene, D., and Mitchell, J. Dropout prediction in moocs using learner activity features. *eLearning Papers* (2014), 1–10.
13. Kember, D. A longitudinal-process model of drop-out from distance education. *The Journal of Higher Education* (1989), 278–301.
14. 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.



15. 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.
16. 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.
17. Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., and Klemmer, S. R. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 33.
18. Lewin, T. After setbacks, online courses are rethought. *New York Times* (2013, December).
19. Moore, M. G., and Kearsley, G. *Distance education: A systems view*. Wadsworth Publishing Company, 1966.
20. 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.
21. Muilenburg, L. Y., and Berge, Z. L. Student barriers to online learning: A factor analytic study. *Distance Education* 26, 1 (2005), 29–48.
22. Parr, C. Mooc completion rates ‘below 7%’. *Times Higher Education* (2013, May).
23. 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).
24. Rivard, R. Measuring the mooc dropout rate. *Inside Higher Ed* 8 (2013).
25. Rovai, A. P. In search of higher persistence rates in distance education online programs. *The Internet and Higher Education* 6, 1 (2003), 1–16.
26. 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).
27. Tinto, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* (1975), 89–125.
28. 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.
29. Waldrop, M. Online learning: Campus 2.0. *Nature* 495, 7440 (2013), 160–163.
30. 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.
31. Wilkowski, 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.
32. 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).