THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Studying MOOC completion at scale using the MOOC replication framework

Citation for published version:
Andres, JML, Baker, R, Gasevic, D, Siemens, G, Crossley, S & Spann , C 2018, Studying MOOC completion at scale using the MOOC replication framework. in LAK '18 Proceedings of the 8th International Conference on Learning Analytics and Knowledge. ACM Press, pp. 71-78. DOI: 10.1145/3170358.3170369

Digital Object Identifier (DOI):
10.1145/3170358.3170369

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
LAK '18 Proceedings of the 8th International Conference on Learning Analytics and Knowledge

Using the MOOC Replication Framework to Examine Course Completion

ABSTRACT

Research on learner behaviors and course completion within Massive Open Online Courses (MOOCs) has been mostly confined to single courses, making the findings difficult to generalize across different data sets and to assess which contexts and types of courses these findings apply to. This paper reports on the development of the MOOC Replication Framework (MORF), a framework that facilitates the replication of previously published findings across multiple data sets and the seamless integration of new findings as new research is conducted or new hypotheses are generated. In the proof of concept presented here, we use MORF to attempt to replicate 15 previously published findings across 29 iterations of 17 MOOCs. The findings indicate that 12 of the 15 findings replicated significantly across the data sets. Results contradicting previously published findings were found in two cases. MORF enables larger-scale analysis of MOOC research questions than previously feasible, and enables researchers around the world to conduct analyses on huge multi-MOOC data sets without having to negotiate access to data.

CCS CONCEPTS

• Applied computing ~ Education

KEYWORDS

MOOCs, MORF, MOOC Replication Framework, completion, multi-MOOC analysis, replication, meta-analysis.

ACM Reference format:

[Redacted]. 2018. Using the MOOC Replication Framework to Examine Course Completion. In *Proceedings of the International Conference on Learning Analytics and Knowledge, Sydney, Australia, March 2018 (LAK'18)*, 8 pages. DOI: [TO BE PROVIDED]

1 INTRODUCTION

Massive Open Online Courses (MOOCs) have created new opportunities to study how learning occurs across contexts, with thousands of courses offered, millions of users registered, and billions of student-platform interactions [1]. Both the popularity of MOOCs among students [2] and their benefits to those who complete them [3] suggest that MOOCs present a new, easily scalable, and easily accessible opportunity for learning. A major criticism of MOOC platforms, however, is their frequently high attrition rates [4], with only 10% or fewer learners completing many popular MOOC courses [1, 5]. As such, a majority of research on MOOCs in the past three years has been geared towards understanding and increasing student completion. Researchers have investigated features of individual courses, universities, platforms, and students [2] as possible explanations of why students complete or fail to complete.

A majority of MOOC research has been limited to single courses, often taught by the researchers themselves, which is due in most part to the lack of access to other data, as well as challenges to researchers in working with data sets much larger than those they are used to. While understandable, the practice of conducting analyses on small samples often leads to inconsistent findings and questions about the generalizability and replicability of what is learned. In the context of MOOCs, for example, one study investigated the possibility of predicting course completion based on forum posting behavior in a 3D graphics course [6]. They found that starting threads more frequently than average was predictive of completion. Another study investigating this relationship in two courses on Algebra and Microeconomics found the opposite to be true; participants that started threads more frequently were less likely to complete [7]. Research in single courses has the risk of producing contradictory findings which are difficult to resolve. Running analyses on single-course data sets limits the generalizability of findings, and leads to inconsistency between published reports [8].

In another example of this problem, one study investigating the relationship between students' motivations in taking the course and course completion across three open online learning environments found that students who were taking a course for credit were more likely to complete [4]. An attempt to replicate this finding in a different MOOC found that this feature was not a statistically significant predictor of completion [9].

The current limited scope of much of the current research within MOOCs has led to several contradictory findings of this nature, duplicating the "crisis of replication" seen in the social psychology community [10]. The ability to determine which findings generalize across MOOCs, which findings don't, and in what contexts less universal findings are relevant, will lead to trustworthy and ultimately more actionable knowledge about learning and engagement in MOOCs.

While there has been some initial interest in data sharing within MOOCs, prior efforts have not yet changed this state of affairs. Individual universities store data on dozens of MOOCs, but have mostly not yet made this data available to researchers in a fashion that enables large-scale analysis (although individual examples of multi-MOOC analyses exist [cf. 11, 12]). The edX Research Data Exchange (RDX) has made limited data from multiple universities accessible to researchers at other universities [13], but has also restricted the data available due to concerns about privacy, restricting key data necessary to replicating many published analyses. The moocDB data format and moocRP research platform were developed with a goal of supporting research in this area [14]. Their tool allows for the implementation of several analytic models, with the goal of facilitating the re-use and replication of an analysis in a new MOOC. However, the use of moocRP has not yet scaled beyond analyses of single MOOCs, making it uncertain how useful it will be for the types of broad, cross-contextual research that are needed to get MOOC research past its own replication crisis.

2 MORF: GOALS AND ARCHITECTURE

One of the common approaches to resolving the uncertainty caused by contradictory findings is to conduct meta-analyses [15], where the results of several previous findings are integrated together to produce a more general answer to a research question. The meta-analysis research community has developed powerful statistical techniques for synthesizing many studies together despite incomplete information. By definition, however, a meta-analysis must wait on the completion of analyses by multiple research groups.

An alternate approach is to collect large and diverse data sets to then test published findings in. Such an approach has historically been infeasible in learning contexts, where data sources were, up until relatively recently, disparate, incompatible, and small. Even though large amounts of data have become available for individual intelligent tutoring systems over the last decade [16], the differences in the design of different tutoring systems and the semantics of data fields (even when the data field has the same name across different systems, or the systems share a common data format as in the Pittsburgh Science of Learning Center DataShop [16]), has made statistical analyses across multiple platforms relatively rare. However, analysis across large ranges of courses becomes more feasible for MOOCs, where a small number of providers generate huge amounts of data on courses with very different content, but relatively similar high-level design.

To leverage this opportunity, we have developed MORF, the MOOC Replication Framework, a framework for investigating research questions in MOOCs within data from multiple MOOC data sets. Our goal is to determine which relationships (particularly, previously published findings) hold across different courses and iterations of those courses, and which findings are unique to specific kinds of courses and/or kinds of participants. In our first report on MORF [9], we discussed the MORF architecture and attempted to replicate 21 published findings in the context of a single MOOC. We found that nine of these 21 findings replicated successfully in the MOOC studied, and results contradicting the previously published findings were found in two cases. In this paper, we report the first large-scale use of MORF, attempting to replicate 15 published findings in 29 iterations of 17 MOOCs, listed in Table 1.

In its current version, MORF represents findings as production rules, a simple formalism previously used in work to develop human-understandable computational theory in psychology and education [17, 18]. This approach allows findings to be represented in a fashion that human researchers and practitioners can easily understand, but which can be parametrically adapted to different contexts, where slightly different variations of the same findings may hold.

The production rule system used in MORF was built using Jess, an expert system programming language [19]. All findings were converted into if-else production rules following the format, “If a student who is <attribute> does <operator>, then <outcome>.” Attributes are pieces of information about a student, such as whether a student reports a certain goal on a pre-course questionnaire. Operators are actions a student does within the MOOC. Outcomes can represent a number of indicators of student success or failure including watching a majority of videos [e.g., 11, 20] or publishing a scientific paper after participating in the MOOC [e.g., 21]. In the current study, we focus on the most commonly-studied research question, whether or not the student in question completed the MOOC. Not all production rules need to have both attributes and operators. For example, production rules that look at time spent in specific course pages may have only operators (e.g., spending more time in the forums than the average student) and outcomes (i.e., whether or not the participant completed the MOOC) [e.g., 22].

Each production rule returns two counts: 1) the confidence [23], or the number of participants who fit the rule, i.e., meet both the if and the then statements, and 2) the conviction [24], the production rule’s counterfactual, i.e., the number of participants who match the rule’s then statement but not the rule’s if statement. For example, in the production rule, “If a student posts more frequently to the discussion forum than the average student, then they are more likely to complete the MOOC,” the two counts returned are the number of participants that posted more than the average student and completed the MOOC, and the number of participants who posted less than the average, but still completed the MOOC. As a result, for each MOOC, a confidence and a conviction for each production rule can be generated.

Table 1. The full list of courses included in the current study, and the number of iterations each course was offered.

Course Title	Number of Iterations
Artificial Intelligence Planning	2
Animal Behavior and Welfare	1
Astrobiology	2
AstroTech: The Science and Technology Behind Astronomical Discovery	2
Clinical Psychology	1
Code Yourself! An Introduction to Programming	1
E-Learning and Digital Cultures	3
EDIVET: Do you have what it takes to be a veterinarian?	2
Equine Nutrition	2
General Elections 2015	1
Introduction to Philosophy	4
Mental Health: A Global Priority	1
Fundamentals of Music Theory	1
Nudge-It (Introduction to the obesity problem)	1
Philosophy and the Sciences	2
Introduction to Sustainability (Explores topics like ecosystem degradation and resource limitation)	1
The Life and Work of Andy Warhol	2

A chi-square test of independence can then be calculated comparing each confidence to each conviction. The chi-square test can determine whether the two values are significantly different from each other, and in doing so, determine whether the production rule or its counterfactual significantly generalized to the data set. Odds ratio effect sizes per production rule are also calculated. In this study, we tested MORF on 29 data sets obtained from the University of Edinburgh's large MOOC program. In integrating across MOOCs, we choose the conservative and straightforward method of using Stouffer's [25] Z-score method to combine the results per finding across the multiple MOOC data sets, to obtain a single statistical significance result across all MOOCs. We also report mean and median odds ratios across data sets.

3 SCOPE OF ANALYSIS

In a first report on MORF's infrastructure, we attempted to replicate a set of 21 previously published findings in a single MOOC on Big Data in Education [9]. Six findings analyzed in this first report required questionnaire data that was not available for the broader set of MOOCs investigated in the current study. As such, the current study analyzes the remaining 15 of these findings on MOOC completion across 29 iterations of 17 different MOOCs offered through Coursera by the University of Edinburgh. There was a total of 514,656 registrants and 86,535,662 user events across these 29 MOOC data sets.

Within the context of these MOOCs, we investigate previously published findings from five papers demonstrating that discussion forum behaviors were associated with successful course completion. This category of findings was studied for two reasons. First, it has importance to the design of effective MOOCs. Understanding the role that discussion forum participation plays in course completion is important to designing discussion forums that create a positive social environment that enhances learner success [28]. Second, it represents a type of finding that has been difficult to investigate at scale with existing data sets, since there has been limited sharing of the type of discussion forum data necessary for this type of research, due to the difficulty of deidentifying this type of data. Prominent findings on MOOC completion involving time spent within the forums, as compared to other activities, were also considered.

These five past papers found that writing longer posts [5, 26], writing posts more often [8, 26], starting a thread, receiving replies on one's thread, and replying to others' threads [5, 8, 27], and just generally spending more time in the forums and on quizzes [22] were significantly associated with course completion. The original papers on these findings involved one edX MOOC on Electronics [22], and Coursera MOOCs on Surviving Disruptive Technology [27], Algebra [5, 7], Microeconomics [5, 7], and Big Data in Education [26]. The full list of findings investigated is given in Table 2.

One area of particular interest for many MOOC researchers is learners' failure to complete MOOC courses, due the problem's importance and potential actionability. Completion is important even beyond the context of a single MOOC. Though not all MOOC learners have the goal of completion [29], completion is one of the best predictors of eventual participation in the community of practice associated with the MOOC [21]. As such, understanding why learners fail to complete MOOCs may enable the design of interventions that increase the proportion of students who succeed in MOOCs. The studies included in this paper's set of analyses sought to understand which student behaviors were significantly related to course completion, as a step towards designing interventions.

Table 2. The previously published findings on MOOC completion included in the study, presented as production rules, as well as the articles the findings are drawn from.

#	If	Then	Source
1	Participant spends more time in forums than average	Likely to complete	[22]
2	Participant spends more time on assignments than average	Likely to complete	[22]
3	Participant's average length of posts is longer than the course average	Likely to complete	[5, 26]
4	Participant posts on the forums more frequently than average	Likely to complete	[7, 26]
5	Participant responds more frequently to other participants' posts than average	Likely to complete	[5]
6	Participant starts a thread	Likely to complete	[5]
7	Participant starts threads more frequently than average	Not likely to complete	[8]
8	Participant has respondents on threads they started	Likely to complete	[27]
9	Participant has respondents on threads they started greater than average	Likely to complete	[27]
10	Participant uses more concrete words than average	Likely to complete	[26]
11	Participant uses more bigrams than average	Likely to complete	[26]
12	Participant uses more trigrams than average	Likely to complete	[26]
13	Participants uses less meaningful words than average	Likely to complete	[26]
14	Participant uses more sophisticated words than average	Likely to complete	[26]
15	Participant uses a wider variety of words than average	Likely to complete	[26]

In the first of these five articles, De Boer and colleagues [22] explored the impact of resource use on achievement within edX's first MOOC, Circuits and Electronics, offered in Spring 2012. The class reportedly drew students from nearly every country in the world. The study correlated course completion to the amount of time spent on different online course resources, and found that time spent on the forums and time spent on assignments were predictive of higher overall final scores (required for course completion with a certificate), even when controlling for prior ability and country of origin. These results show that time allocation is an important predictor of student success in MOOCs.

Two studies by Yang and her colleagues [5, 7] explored dropout rates, confusion, and forum posting behaviors within two Coursera MOOCs, one on Algebra and the other on Microeconomics. Their first study developed a survival model that measured the influence of student behavior and social positioning within the discussion forum on student dropout rates on a week-to-week basis. The second study attempted to quantify the effect of behaviors indicative of confusion on participation through the development of another survival model. They found that the more a participant engaged in behaviors they believed indicative of confusion (i.e., starting threads more frequently than the average student), the lower their probability of retention in the course. The findings of these two studies on the relationship of posting behavior (i.e., starting threads, writing frequent and lengthy posts, and responding to others' posts) to course completion are crucial to the design of MOOCs because they suggest that social factors are associated with a student's propensity to drop out during their progression through a MOOC.

Crossley and colleagues [26] conducted a similar investigation on the relationship between discussion forum posting behaviors and MOOC completion in a MOOC on Big Data in Education. In their study, they also found that a range of linguistic features, computed through natural language processing, were associated with successful MOOC completion, including the use of concrete, meaningful, and sophisticated words, and the use of bigrams and trigrams. Concreteness is assessed based on how closely a word is connected to specific objects. "If one can describe a word by simply pointing to the object it signifies, such as the word apple, a word can be said to be concrete, while if a word can be explained only using other words, such as infinity or impossible, it can be considered more abstract [30, p. 762]." Meaningfulness is assessed based on how related a word is to other words. According to the definition in [30], words like "animal," for example, are likely to be more meaningful than field-specific terms like "equine". Lexical sophistication involves the "depth and breadth of lexical knowledge [30]." It is usually assessed using word frequency indices, which look at the frequency by which words from multiple large-scale corpora appear in a body of text [30]. More concrete or more sophisticated words were found to be associated with a greater probability of course completion, while more meaningful words were found to be associated with a lower probability of course completion. The findings of their study have important implications for how individual differences among students that go beyond observed behaviors (e.g., language skills and usage choices) can predict success.

As mentioned, the current study attempts to replicate 15 previously published findings relating to participant behaviors and MOOC completion. These findings are presented in Table 2 as if-then production rules; the previous articles the findings were drawn from are also included. The findings are divided into three categories: findings involving data drawn from clickstream logs concerning time spent on specific activities within the MOOC (Rules 1-2), findings involving data drawn from the discussion forum that look at the participants' posting behavior (Rules 3-9), and findings involving data from the forum posts that look at linguistic features of the participants' contributions (Rules 10-15). The Tool for the Automated Analysis of Lexical Sophistication 1.4, or TAALES [30], and the Tool for the Automatic Analysis of Cohesion 1.0, or TAACO [31], were used to generate the linguistic variables used in the analyses.

In TAALES, sophistication is derived from word occurrence across multiple large-scale corpora and are computed using five frequency indices: the Thorndike-Lorge index based on Lorge's 4.5 million-word corpus on magazine articles [32], the Brown index [33] based on the 1 million-word London-Lund Corpus of English Conservation [34], the Kucera-Francis index based on the Brown corpus, which consists of about 1 million words published in the US [35], the British National Corpus (BNC) index based on about 100 million word of written and spoken English in Great Britain [36], and the SUBTLEXus index based on a corpus of subtitles from about 8000 films and television series in the US [37]. TAALES returns a sophistication score per corpus. The more words from these five corpuses are used, the higher the respective sophistication score is. For more information on these corpora, see [30]. Bigram and trigram frequency are two other metrics of lexical sophistication [30], i.e., the more bigrams and trigrams used, the more sophisticated a body of text is. Bigrams are pairs of consecutive words, and trigrams are groups of three consecutive words. TAALES calculates frequencies of commonly used bigrams and trigrams from an 80-million-word sub-corpus of the BNC. The use of bigrams and trigrams are normed to the length of its text post.

One production rule studied in this paper is a re-parameterized version of an original finding that was carried over into the current study from the first use of MORF in a single MOOC [9]. Rule 8 was the original finding, i.e., participants having respondents on their threads in the discussion forum. Within [9], we created a variant of this rule, Rule 9, participants having more respondents on their threads than average, due to the relatively low numbers of threads with zero respondents in some MOOCs.

4 USING MORF

The production rule analysis of MORF makes use of two different kinds of data: 1) clickstream events used to analyze the rules relating to the amount of time spent in the forums and on the assignments, and 2) relational database forum data used to analyze the rules relating to forum behavior and linguistic features. MORF utilizes Amazon Web Services (AWS) for data storage of the clickstream events, which are stored in Amazon S3 buckets, and database access for the forum-related data via Amazon's Relational Database Service (RDS).

When MORF is run, it connects securely and remotely to AWS to access all necessary data. The user simply needs to state which courses and course iterations they intend to run the production rule analysis on, and once the analysis is complete, the user is presented with the list of MOOCs stored in MORF's AWS storage, whether or not each production rule replicated significantly within each course iteration, and the significance level and effect size for each analysis, as well as the overall analysis.

Utilizing such an architecture protects privacy and data ownership by enabling users to run analyses without seeing any of the data. Users are also able to either contribute their own data sets to MORF, or conduct their own analyses against MORF's data set, which is currently comprised of 131 iterations of 61 MOOCs.

5 RESULTS

The results of the 15 analyses across MOOCs can be found in Tables 3 and 4. In Table 3, each row represents the result of testing each previously published finding across the full set of MOOCs. The table reports each finding, again presented as an if-then production rule, the respective Z-scores and p-values for the analysis across MOOCs, as well as the number of MOOCs in which the finding significantly replicated, the number of MOOCs that had the counterfactual replicate, and the number of MOOCs where the finding failed to replicate in in either direction. Counterfactuals that are statistically significant overall, across MOOCs, are marked by shaded bands. Findings that failed to replicate in either direction are italicized. Table 4 reports the mean and median odds ratio effect sizes of each production rule across the 29 data sets.

As shown in Table 3, two of the 15 previous findings had their counterfactuals come out statistically significant, i.e., they had the opposite result from the result previously reported. Whereas Yang and colleagues [7] found that students who start threads on the forums more frequently than the average student are less likely to complete, we found that in 27 cases out of 29 (with 0 positive replications and 2 null effects) that students who start threads less frequently are less likely to complete. Also, whereas Crossley and colleagues [26] found that students who used a wider variety of words in their forum posts than the average student were more likely to complete, we found in 13 cases out of 29 (with 2 positive replications, and 14 null effects) that students who used a narrower variety of words were more likely to complete. Finally, one finding, which originally stated that students who used more concrete words in their forum posts than the average student were more likely to complete, failed to replicate overall in either direction (with 3 positive replications, and 5 negative replications). The remaining 12 of the 15 previous findings replicated significantly across the 29 data sets.

6 IMPLICATIONS

Twelve of the fifteen production rules investigated significantly replicated across the data sets. The previously published findings related to time spent in the forums and on assignments – stating that more time spent on these activities is associated with completion – replicated significantly across all 29 data sets. These findings indicate that spending more time with the course content, either through engaging in or observing the discussions in the forums or through engaging with the course assignments, is associated with completion.

This is likely for multiple reasons. More motivated participants are likely to spend more time within the MOOC and are also more likely to complete. Spending more time with the material may also increase the chance of successful performance and completion. In an environment such as MOOCs, where students have the freedom to disengage at any point in the course, knowing that time spent in the discussion forums is associated with remaining engaged till completion indicates that attention should be spent on designing engaging and positive discussion forum experiences that encourage participation.

Beyond this, most rules on posting behaviors replicated significantly across the 29 data sets as well. These rules found that writing longer posts, writing posts more frequently, responding more frequently to other students' posts, and having others respond more frequently to one's own posts are all significant predictors of completion. Interactions among and between students and course staff, and certainly, the behavior of posting and responding frequently on the forums implies, at the very least, an interest to learn. This greater effort spent in participation in many cases is probably also associated with learning from one's peers, an important aspect of MOOCs.

Table 3. The meta-analysis results per finding. Shaded bands indicate that our replication found the reverse of the published finding. Italics represent null results.

If	Then	Z	p	+	-	null
More time in forums	Likely to complete	26.93	< 0.001	29	0	0
More time on assignments	Likely to complete	26.93	< 0.001	29	0	0
Longer posts than average	Likely to complete	11.76	< 0.001	15	1	13
Posts more frequently than average	Likely to complete	26.04	< 0.001	27	0	2
Responds more frequently than average	Likely to complete	23.84	< 0.001	25	0	4
Starts a thread	Likely to complete	12.34	< 0.001	15	0	14
Starts threads more frequently than average	Not likely to complete	26.39	< 0.001	0	27	2
Has respondents	Likely to complete	22.29	< 0.001	26	0	3
Has respondents greater than average	Likely to complete	22.72	< 0.001	24	0	5
Uses more concrete words	Likely to complete	1.51	0.131	3	5	21
Uses more bigrams	Likely to complete	12.68	< 0.001	15	1	13
Uses more trigrams	Likely to complete	12.84	< 0.001	16	1	12
Uses less meaningful words	Likely to complete	10.18	< 0.001	16	0	13
Uses more sophisticated words	Likely to complete	17.54	< 0.001	20	0	9
Uses wider variety of words	Likely to complete	-4.11	< 0.001	2	13	14

One rule in this area, however, replicated significantly in the opposite direction. The finding originally stated that students who start threads more frequently are less likely to complete [9]. Its counterfactual, however, which states that students who start threads less frequently than the average student are less likely to complete, replicated significantly across 27 of the 29 MOOCs. Yang and colleagues interpreted starting a thread as indicating confusion, and indeed, this may motivate some students to start threads. It is likely, however, that students start threads for many reasons beyond confusion, including to share ideas [38], make personal contact with other students [38, 39], and even to insult their instructor [40]. It may be valuable in future work to more thoroughly study the content of discussion threads in order to see if different posts have different associations to student outcomes.

Table 4. The mean and median odds ratio effect sizes per finding across the 29 data sets. Shaded bands indicate that our replication found the reverse of the published finding. Italics represent null results.

If	Then	Odds Ratio mean	Odds Ratio median
More time in forums	Likely to complete	27.235	12.060
More time on assignments	Likely to complete	251.979	121.349
Longer posts than average	Likely to complete	1.362	1.238
Posts more frequently than average	Likely to complete	4.667	3.406
Responds more frequently than average	Likely to complete	2.959	2.569
Starts a thread	Likely to complete	1.874	1.676
Starts threads more frequently than average	Not likely to complete	4.601	3.571
Has respondents	Likely to complete	2.321	1.997
Has respondents greater than average	Likely to complete	2.544	2.250
Uses more concrete words	Likely to complete	1.036	1.076
Uses more bigrams	Likely to complete	1.376	1.292
Uses more trigrams	Likely to complete	1.390	1.281
Uses less meaningful words	Likely to complete	0.799	0.782
Uses more sophisticated words	Likely to complete	1.623	1.472
Uses wider variety of words	Likely to complete	0.987	0.875

In terms of the linguistic features of the participants' forum posts, the analysis found that students more likely to complete the MOOCs produced more sophisticated language and used more bigrams and trigrams, but used less meaningful words, replicating the findings of Crossley and his colleagues [26]. However, Crossley et al.'s previous findings on concreteness failed to replicate (but did not replicate in reverse either).

One of the findings that did replicate was the negative relationship between using meaningful words and course completion. Within TAALES, meaningful words are words with greater numbers of associations to other words, regardless of domain [22]. In other words, the finding seen here (replicating [26]) may be because words interpreted as linguistically meaningful by TAALES may be less relevant to course content than other words. Using fewer meaningful words could thus mean that participants were using field-specific terms in their discussion posts. Conversing using field-specific terms could imply better understanding of the content being taught in the course. By contrast, lexical sophistication involves the "depth and breadth of lexical knowledge [30]." Word sophistication, bigram use, and trigram use are all measures of lexical sophistication within TAALES. The findings positively linking lexical sophistication to course completion, thus, imply that more sophisticated posts are associated with remaining engaged in the course. More sophisticated language may also be associated with positive understanding of the course content.

One production rule turned out to be significant in the reverse direction from what was reported in its original article. The finding was part of a set of linguistic features that were correlated with course completion [26]. The rule originally states that participants who post on the forums using a wider variety of words than the average student were more likely to complete. This analysis, however, found that using a narrower variety of words was significantly related to course completion. One possibility is that students who use a considerable variety of words are not focusing on words of specific importance for their current course, but are instead rambling on a range of other (often unrelated) topics [cf. 40, 41].

Overall, these findings suggest that there is considerable commonality in which behaviors are associated with success in MOOCs, across MOOCs on a heterogeneous range of topics, creating the possibility that interventions that encourage specific behaviors from the set studied here may have positive incomes on student success, even in entirely new courses.

7 CONCLUSION AND FUTURE WORK

In this paper, we investigate the degree to which previously published findings on MOOC course completion replicate across multiple new and different data sets. This was achieved through the development of the MOOC Replication Framework, or MORF, a framework that was used to attempt the replication of 15 previously published findings on MOOC completion on 29 MOOC data sets, drawn from 17 distinct courses on a range of topics. These 15 findings, represented as productions rules, were drawn from 5 studies that sought to understand the high attrition rate in MOOCs. Of these 15 findings, 12 successfully replicated across the 29 data sets, while 2 were statistically significant in the opposite direction. Through the development of MORF and the resulting analyses conducted, this study presents a larger-scale analysis of MOOC research questions than previously feasible.

Our next steps include extending our work published here in several ways. First, we plan to expand the current set of variables being modeled in MORF, both in terms of predictor (independent) variables and outcome (dependent) variables. This will enable us to replicate a broader range of published findings. Our first efforts do not yet include findings involving data from performance on assignments or behavior during video-watching, two essential activities in MOOCs which have been extensively researched in the last three years. To accomplish this goal, we intend to conduct a more comprehensive literature review. The findings in published papers can then be turned into production rules for replication on the current data set.

Second, we plan to move our framework beyond simply capturing findings that can be expressed and production rules, and also analyze findings that can only be expressed as more complex predictive models, in partnership with researchers at the [University of Redacted]. While we view production rules as a highly interpretable and reasonably flexible framework, more complex prediction models are already in use to determine which students are at risk of failing to complete a course [11, 12, 28]. Being able to test these more complex models for replication as well will broaden the applicability of the MORF framework.

Third, we plan to expand to an even greater range of data. Initially, we plan to apply the production rules to data from other MOOC courses. This should be a straightforward process as MORF is able to ingest raw edX and Coursera data seamlessly. At the time of this writing, we are nearing completion of the ingestion of edX and Coursera data from two other universities. Eventually, we hope to add data from other platforms as well.

Fourth, we intend to add to MORF a characterization of the features of the MOOCs themselves, towards studying whether some findings fail to replicate in specific MOOCs due to the differences in design, domain, or audience between MOOCs. Although 13 findings replicated overall, not all findings replicated in all MOOCs. Understanding how the features of the MOOC itself can explain differences in which results replicate may help us to explain some of the contradictory findings previously reported in single-MOOC research. With the large pool of courses MORF currently has access to, we intend to go beyond simple replication to study how factors like course design, target and actual population, domain, and instructor pedagogy influence the applicability of these findings. In turn, this will help us to understand which findings apply in which contexts, towards understanding how the different design of different MOOCs drive differences in the factors associated with student success.

Fifth, and perhaps most importantly, we are currently working with colleagues at the [University of Redacted] to create an infrastructure which will enable us to share access to MORF – while not sharing the data sets themselves – to a broader audience. This will enable a broader range of researchers to access and utilize large-scale MOOC data to conduct generalizable research on learning in this context. By broadening the base of access to large-scale learning data, we can incorporate a wider variety of ideas and a greater amount of energy and researcher time, with the hope of eventual speeding progress in this emerging scientific area.

ACKNOWLEDGMENTS

[Redacted]

REFERENCES

[1] Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. The International Review of Research in Open and Distributed Learning, 15(1).

[2] Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses.

[3] Zhenghao, C., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. (2015). Who's Benefiting from MOOCs, and Why. Harvard Business Review

[4] Clow, D. (2013). MOOCs and the funnel of participation. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 185-189). ACM

[5] Yang, D., Sinha, T., Adamson, D., & Rose, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In Proceedings of the 2013 NIPS Data-driven education Workshop (Vol. 11, p. 14)

[6] Andersson, U., Arvemo, T., & Gellerstedt, M. (2016). How well can completion of online courses be predicted using binary logistic regression?. In IRIS39-The 39th Information Systems Research Conference in Scandinavia, Ljungskile, Sweden, 7-10 August 2016.

[7] Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015). Exploring the effect of confusion in discussion forums of massive open online courses. In Proceedings of the Second (2015) ACM Conference on Learning@ Scale (pp. 121-130). ACM.

[8] Łukasz, K., Sharma, K., Shirvani Boroujeni, M., & Dillenbourg, P. (2016). On generalizability of MOOC models. In Proceedings of the 9th International Conference on Educational Data Mining (No. EPFL-CONF-223613, pp. 406-411).

[9] Redacted

[10] Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty replication in the education sciences. Educational Researcher, 0013189X14545513.

[11] Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014, March). Understanding in-video dropouts and interaction peaks in online lecture videos. In Proceedings of the first ACM conference on Learning@ scale conference (pp. 31-40). ACM.

[12] Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout.

[13] edX. (2017). edX Research Guide. Retrieved from the edX website: http://edx.readthedocs.io/projects/devdata/en/latest/

[14] Pardos, Z. A., & Kao, K. (2015, March). moocRP: An open-source analytics platform. In Proceedings of the Second (2015) ACM conference on learning@ scale (pp. 103-110). ACM.

[15] Schmidt, F. L., & Hunter, J. E. (2014). Methods of meta-analysis: Correcting error and bias in research findings. Sage publications.

[16] Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. Handbook of educational data mining, 43.

[17] Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. Human-Computer Interaction, 12(4), 439-462.

[18] Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. Artificial intelligence, 33(1), 1-64.

[19] Friedman-Hill, E. (2002). Jess, the expert system shell for the java platform.USA: Distributed Computing Systems.

[20] Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). Your click decides your fate: Inferring information processing and attrition behavior from MOOC video clickstream interactions. arXiv preprint arXiv:1407.7131.

[21] Wang, Y. Baker, R. (2015) Content or Platform: Why do students complete MOOCs? MERLOT Journal of Online Learning and Teaching, 11 (1), 17-30.

[22] DeBoer, J., Ho, A., Stump, G.S., Pritchard, D.E., Seaton, D. and Breslow, L. (2013). Bringing student backgrounds online: MOOC user demographics, site usage, and online learning. engineer, 2, pp.0-81.

[23] Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Associations between Sets of Items in Massive Databases. In Proceedings of the ACM-SIGMOD Int'l Conference on Management of Data (pp. 207-216).

[24] Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997, June). Dynamic itemset counting and implication rules for market basket data. In ACM SIGMOD Record (Vol. 26, No. 2, pp. 255-264). ACM.

[25] Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. & Williams, R.M. Jr. (1949). The American Soldier, Vol. 1: Adjustment during Army Life. Princeton University Press, Princeton.

[26] Crossley, S., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to Completion: Success in an Educational Data Mining Massive Open Online Class. International Educational Data Mining Society.

[27] Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H. & Getoor, L. (2013, December). Modeling learner engagement in MOOCs using probabilistic soft logic. In NIPS Workshop on Data Driven Education (Vol. 2, pp. 1-7).

[28] Wen, M., Yang, D., & Rose, C. (2014, July). Sentiment Analysis in MOOC Discussion Forums: What does it tell us?. In Educational data mining 2014.

[29] Wilkowski, J., Deutsch, A., & Russell, D. M. (2014, March). Student skill and goal achievement in the mapping with google MOOC. In Proceedings of the first ACM conference on Learning@ scale conference (pp. 3-10). ACM.

[30] Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. TESOL Quarterly, 49 (4), 757-786.

[31] Crossley, S. A., Kyle, K., and McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. Behavior Research Methods.

[32] Thorndike, E. L., & Lorge, I. (1944). The teacher's word book of 30,000 words. New York, NY: Teachers College, Columbia University.

[33] Brown, G. D. A. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. Behavior Research Methods, Instrumentation & Computers, 16, 502–532. doi:10.3758/BF03200836

[34] Svartvik, J., & Quirk, R. (1980). A corpus of English conversation. Lund, Sweden: Gleerup.

[35] Kucera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

[36] BNC Consortium. (2007), British National Corpus, version 3 (BNC XML ed.). Retrieved from www.natcorp.ox.ac.uk

[37] Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Meth- ods, 41, 977–990. doi:10.3758/BRM.41.4.977

[38] Sharif, A., & Magrill, B. (2015). Discussion forums in MOOCs. International Journal of Learning, Teaching and Educational Research, 12(1).

[39] Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. Journal of Online Learning and Teaching, 9(2), 149.

[40] Comer, D., Baker, R., Wang, Y. (2015) Negativity in Massive Online Open Courses: Impacts on Learning and Teaching. InSight: A Journal of Scholarly Teaching, 10.

[41] Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating How Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains. International Educational Data Mining Society.