

Studying MOOC Completion at Scale Using the MOOC Replication Framework

Juan Miguel L. Andres
Ryan S. Baker
University of Pennsylvania
3700 Walnut Street
Philadelphia, PA 19104
+1 (877) 736-6473
andresju@gse.upenn.edu,
rybaker@upenn.edu

George Siemens
Catherine A. Spann
University of Texas Arlington
701 S Nedderman Drive
Arlington, TX 76019
+1 (817) 272-2011
gsiemens@gmail.com,
caspann17@gmail.com

Dragan Gašević
University of Edinburgh
Old College, South Bridge
Edinburgh EH89YL, UK
+44 (131) 650-1000
dragan.gasevic@ed.ac.uk

Scott Crossley
Georgia State University
38 Peachtree Center Ave.
Atlanta, GA 30303
+1 (404) 413-5000
sacrossley@gmail.com

ABSTRACT

Research on learner behaviors and course completion within Massive Open Online Courses (MOOCs) have 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, and that two findings replicated significantly in the opposite direction. 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.

Keywords

MOOC, MORF, replication, meta-analysis.

1. INTRODUCTION

Massive Open Online Courses (MOOCs) have created new opportunities to study learning at scale, with millions of users registered, thousands of courses offered, 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 3 years has been geared towards 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 this research, however, 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. In order to increase access to data and make analysis easier, researchers at UC Berkley developed an open-source repository and analytics tool for MOOC data [6]. Their tool allows for the implementation of several analytic models, facilitating the re-use and replication of an analysis in a new MOOC.

Running analyses on single data sets, however, still limits the generalizability of findings, and leads to inconsistency between published reports [7]. 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 [8]. They found that starting threads more frequently than average was predictive of completion. Another study investigating the relationship between forum posting behaviors, confusion, and completion in two courses on Algebra and Microeconomics found the opposite to be true; participants that started threads more frequently were *less* likely to complete [9].

In another example, 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 the 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 [10].

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 [11]. The ability to determine which findings generalize across MOOCs, and what contexts findings stabilize, will lead to knowledge that can more effectively drive the design of MOOCs and enhance practical outcomes for learners.

2. MORF: GOALS AND ARCHITECTURE

One of the common approaches to addressing contradictory findings is to conduct meta-analyses [12], where the results of several analyses are integrated together to produce a more general answer to a research question. 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, 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 [13],

with a common data format, 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) 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 of MORF [10], we discussed the MORF architecture and attempted to replicate 21 published findings in the context of a single MOOC. 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.

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	1
Philosophy and the Sciences	2
Introduction to Sustainability	1
The Life and Work of Andy Warhol	2

MORF represents findings as production rules, a simple formalism previously used in work to develop human-understandable computational theory in psychology and education [14, 15]. 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 was built using Jess, an expert system programming language [16]. All findings were programmed 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 are, in the case of the current study, whether or not the student in question completed the MOOC (but could represent other outcomes, such as watching more than half of the videos). 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).

Each production rule returns two counts: 1) the confidence [17], or the number of participants who fit the rule, i.e., meets both the if and the then statements, and 2) the conviction [18], 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.

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. Stouffer’s [19] Z-score method was used in order to combine the results per finding across the multiple MOOC data sets, to obtain a single statistical significance across all MOOCs.

3. SCOPE OF ANALYSIS

In the 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 [10]. The current study analyzes 15 of these findings on MOOC completion across 29 iterations of 17 different MOOCs offered through Coursera by the University of Edinburgh. Six findings required questionnaire data that was not available for the current set of MOOCs. There was a total of 514,656 registrants and 86,535,662 transactions 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. 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. These papers found that writing longer posts [5, 20], writing posts more often [7, 20], starting a thread, receiving replies on one’s thread, and replying to others’ threads [5, 7, 21], 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 [21], Algebra [5, 9], Microeconomics [5, 9], and Big Data in Education [20]. 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. Though not all MOOC learners have the goal of completion, it may be possible to design 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. Among 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, 9] 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 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 affect student dropout during their progression through a MOOC.

Crossley and colleagues [20] 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, however, 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 [23, p. 762]." Meaningfulness is assessed based on how related a word is to other words. 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 [23]." 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 [23]. 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.

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][20]
4	Participant posts on the forums more frequently than average	Likely to complete	[9][20]
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	[9]
8	Participant has respondents on threads they started	Likely to complete	[21]
9	Participant has respondents on threads they started greater than average	Likely to complete	[21]
10	Participant uses more concrete words than average	Likely to complete	[20]
11	Participant uses more bigrams than average	Likely to complete	[20]
12	Participant uses more trigrams than average	Likely to complete	[20]
13	Participants uses less meaningful words than average	Likely to complete	[20]
14	Participant uses more sophisticated words than average	Likely to complete	[20]
15	Participant uses a wider variety of words than average	Likely to complete	[20]

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, or TAALES [23], and the Tool for the Automatic Analysis of Cohesion, or TAACO [24], were used to generate the linguistic variables used in the analyses. In this study, we used TAALES v1.4 and TAACO v1.0.

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 [25], the Brown index [26] based on the 1 million-word London-Lund Corpus of English Conservation [27], the Kucera-Francis index based on the Brown corpus, which consists of about 1 million words published in the US [28], the British National Corpus (BNC) index based on about 100 million word of written and spoken English in Great Britain [29], and the SUBTLEXus index based on a corpus of subtitles from about 8000 films and television series in the US [30]. 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 [23]. Bigram and trigram frequency are two other metrics of lexical sophistication [23], i.e., the more bigrams and trigrams used, the more sophisticated a body of text is. TAALES calculates these frequencies from an 80-million-word sub-corpus of the BNC.

One production rule is a parameterized version of the original finding and was carried over into the current study from the first use of MORF in a single MOOC [10]. Rule 8 was the original finding, i.e., participants having respondents on their threads in the discussion forum. Within [10], we created Rule 9 as a variation of the rule, i.e., participants having more respondents on their threads than average, due to the relatively low numbers of threads with zero respondents in some MOOCs.

4. RESULTS

The results of the 15 meta-analyses can be found in Table 3, where each row represents the meta-analysis result per previously published finding. The table reports each finding, again presented as an if-then production rule, the respective Z-scores and p-values for each meta-analysis, as well as the number of MOOCs in which the finding significantly replicated in, the number of MOOCs that had the counterfactual replicate, and the number of MOOCs the finding failed to replicate in in either direction. Statistically significant counterfactuals 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, 12 of the 15 previous findings replicated significantly across the 29 data sets. 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. Finally, one finding failed to replicate significantly in either direction.

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
<i>Uses more concrete words</i>	<i>Likely to complete</i>	<i>1.51</i>	<i>0.131</i>	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

5. IMPLICATIONS

Thirteen of the 15 production rules significantly replicated across the data sets. The previously published findings related to time spent in the forums and on assignments, in particular, 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 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.

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
<i>Uses more concrete words</i>	<i>Likely to complete</i>	<i>1.036</i>	<i>1.076</i>
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

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.

One rule, however, replicated significantly in the opposite direction. The finding originally stated that students who start threads more frequently are less likely to complete, and was drawn from a study where annotated confusion scores were used to predict a number of forum and confusion features, including the number of forum threads started [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. This may be because students start threads for reasons other than confusion, for instance due to being interested in the subject matter.

In terms of the linguistic features of the participants' forum posts, the analysis found (replicating [20]) that students more likely to complete the MOOCs produced more sophisticated language, used more bigrams and trigrams, but used less meaningful words. Previous findings on concreteness failed to replicate (but did not replicate in reverse either).

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 (and in [20]) 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 [23]." Frequency indices, bigram use, and trigram use are all measures of lexical sophistication within TAALES. The findings, 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 [20]. 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. 31, 32].

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.

6. 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 production system 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 productions rules were drawn from 5 studies that sought to address the high attrition rate

in MOOCs. Of these 15 findings, 13 were successfully replicated across the 29 data sets, while 1 was 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 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. Eventually, we hope to add data from other platforms as well.

Third, 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.

7. ACKNOWLEDGMENTS

Thanks to Srećko Joksimović of the University of Edinburgh for facilitating the access to the University's MOOC data, and to the Bill and Melinda Gates Foundation, the University of Pennsylvania, the University of Texas at Arlington, and the University of Edinburgh, for their support for this research.

8. 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] 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.
- [7] Lukasz, 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).
- [8] 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*.
- [9] 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.
- [10] Andres, J.M.L., Baker, R.S., Siemens, G., Gašević, D., & Spann, C.A. (in press). Replicating 21 Findings on Student Success in Online Learning. *Technology, Instruction, Cognition, & Learning*.
- [11] Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty replication in the education sciences. *Educational Researcher*, 0013189X14545513.
- [12] Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage publications.
- [13] 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.
- [14] 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.
- [15] Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial intelligence*, 33(1), 1-64.
- [16] Friedman-Hill, E. (2002). Jess, the expert system shell for the java platform. *USA: Distributed Computing Systems*.
- [17] 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).
- [18] 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.
- [19] 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.

- [20] 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*.
- [21] 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).
- [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] Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, 49 (4), 757-786.
- [24] 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*.
- [25] Thorndike, E. L., & Lorge, I. (1944). The teacher's word book of 30,000 words. New York, NY: Teachers College, Columbia University.
- [26] 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
- [27] Svartvik, J., & Quirk, R. (1980). A corpus of English conversation. Lund, Sweden: Gleerup.
- [28] Kucera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.
- [29] BNC Consortium. (2007), British National Corpus, version 3 (BNC XML ed.). Retrieved from www.natcorp.ox.ac.uk
- [30] 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 Methods*, 41, 977–990. doi:10.3758/BRM.41.4.977
- [31] 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.
- [32] 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*.