Studying MOOC Completion at Scale Using the MOOC Replication Framework

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

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

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

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

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. 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].

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, 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
To address this limitation, we have developed MORF, the **MO**OC **R**eplication **F**ramework, 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 [11], we discussed the MORF architecture and attempted to replicate 21 published findings in the context of a single MOOC.

MORF represents findings as production rules, a simple formalism previously used in work to develop human-understandable computational theory in psychology and education [14]. 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 [15]. 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 [16], or the number of participants who fit the rule, i.e., meets both the if and the then statements, and 2) the conviction [17], 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 and risk ratio effect sizes per production rule are also calculated. Stouffer's [18] Z-score method can be used in order to combine the results per finding across multiple MOOC data sets, to obtain a single statistical significance.

Currently, 40 MOOC data sets and 21 production rules related to pre-course survey responses, time spent in course pages, forum posting behaviors, forum post linguistic features, and completion are incorporated in the framework.

3. FUTURE WORK

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.

Second, 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. Understanding how the features of a 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. Doing so 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.

4. 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] 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).

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

[11] 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*.

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

[16] 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).

[17] 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.

[18] 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.