Methodological Challenges in the Analysis of MOOC Data for Exploring the Relationship between Discussion Forum Views and Learning Outcomes

Yoav Bergner
Educational Testing Service
Princeton, NJ 08541
ybergner@ets.org

Deirdre Kerr
Educational Testing Service
Princeton, NJ 08541
dkerr@ets.org

David E. Pritchard
M.I.T.
Cambridge, MA 02139
dpritch@mit.edu

ABSTRACT

Determining how learners use MOOCs effectively is critical to providing feedback to instructors, schools, and policy-makers on this highly scalable technology. However, drawing inferences about student learning outcomes in MOOCs has proven to be quite difficult due to large amounts of missing data (of various kinds) and to the diverse population of MOOC participants. Thus significant methodological challenges must be addressed before seemingly straightforward substantive questions can be answered. The present study considers modeling final exam performance outcomes on early-stage ability estimates, discussion forum viewing frequency, and overall assessment-oriented engagement (AOE, seen as a proxy measure of motivation). These variables require careful operationalization, analysis of which is the principle contribution of this work. This study demonstrates that the effect sizes of discussion forum viewing activities on final exam outcomes are quite sensitive to these choices.

Author Keywords

MOOCs; discussion forums; social learning.

INTRODUCTION

Massive open online courses (MOOCs), a recent modality of distance learning wherein course materials are made available online and are freely accessible by anyone with computer access, have been rapidly gaining popularity as new platforms and courses come online. As of August 2014, over 2000 MOOCs were being offered through more than 50 initiatives (www.mooc-list.com), and these numbers had more than doubled over the prior year. MOOCs are generally viewed as having great value because they provide expanded opportunities to learn and near-instantaneous feedback and support. Additionally, the large number of enrollees and clickstream interaction logs in any given MOOC provide a vast amount of fine-grained data that can help researchers understand how people learn and how best to support learning in an online environment.

This program of research began with the hope of capitalizing on these properties in order to examine the impact of MOOC discussion forum use on learning outcomes. Simply put, we wanted to study whether viewing discussion board threads while doing homework resulted in final exam gains attributable to this behavior, i.e. controlling for other factors. It seemed prudent to try to account for enrollees with different levels of prior ability and engagement/motivation, as MOOC students are known to have diverse populations. Thus, final exam performance would be our outcome variable; prior ability, engagement/motivation (or some proxy), and discussion forum usage would be covariate predictors. Along the way, however, we perceived that the challenges of operationalizing all of the variables gained more and more importance to the validity of our inferences.

Indeed, recent work by other authors concentrated on the sensitivity of analytical inferences to operationalization of predictor variables such as time-on-task estimation [18]. In reference to that work, this paper may also be seen as an attempt to "penetrate the black box" of a particular MOOC analysis. Thus, we raise the following auxiliary research questions: Does the method of quantifying discussion forum use significantly impact the analysis of its effect on performance? Given that motivation matters, does the decision of which filter to use to exclude unmotivated students change the results of the analysis? Issues of prior ability estimation are myriad; we discuss these briefly below but get into more details in a separate study [4].

In the remainder of this paper, we examine the impact of methodological decisions on the quality and type of inferences that can be drawn from examining MOOC forum use, focusing specifically on methods of quantifying discussion forum use and filtering unmotivated students.

The organization is as follows. By way of motivating our original substantive questions, we first review related literature on the impact of discussion forums in online learning. We then describe our data set. Next, we turn to the challenges of MOOC analyses, in general and specifically to the variables under consideration. We describe different methods for and results from operationalization choices with regards to discussion forum usage, motivation proxies, and prior ability estimates. Finally, we consider the impact of these variables on performance using multiple linear regression models for final exam score.

DISCUSSION FORUMS IN ONLINE LEARNING

The impact of discussion forums on learning in MOOCs and other online courses is still not well understood, although the literature on the subject dates back to the 1990s. While some early research on discussion forums cautioned about the shortcomings of computer-mediated dialogue as compared with face-to-face interactions [25], much of that research explored the benefits of the cognitive processes involved in the use of discussion forums, such as reshaping ideas and constructing meaning with the help of peers [3,21]. Later research (but still prior to the MOOC era) focused on measuring the level and quality of student activity in the forums, for example using data mining and text mining [8]. Cultivation of successful asynchronous discussion was linked to measures of discussion quality [2]. Artificial intelligence approaches for classifying effective synchronous collaborative learning [23] were also applied to asynchronous forums in a graduate level course [24].

Correlations of discussion activity with external performance measures have been the subject of several studies ranging from high school [15] to college [17,19] to graduate school [24], with mixed results. Correlations of 0.51 were found for topical student discussion behaviors (coded by hand) with concept-test performance in a physics course using the learning online network with computer-assisted personalized approach (LON-CAPA) learning management system [17]. Operationalizing discussion behavior purely by counts, [15] found correlations of 0.27-0.44 between project performance and activity volume in the forums for secondary school computer science. [19] performed a multiple regression analysis of quiz scores in two college psychology courses, finding that only content-page-hits were significant, not counts of discussion posts or reads. [24] also found no significant correlations between number of posts and student success in a graduate level course, but success variability was very low and the number of students was only 18.

Prior to MOOCs, the largest number of students in any of these studies was 214 [17]. This is one profound difference in the MOOC era, where tens of thousands of students participate and often thousands complete an online course. More recent analyses of discussion forum use in large MOOCs include the following: one analysis found that superposters elicited more posting from their less prolific peers, but the study did not analyze the impact of posting behavior on performance [14]. A randomized controlled trial comparing students with access to chat and discussion forms to students with access to only discussion forums found no differences in retention or performance between groups [6]. Background characteristics of forum users and the communication networks they formed were analyzed in [12], which found that higher performing students participated more in discussion forums but did not interact exclusively with other higher performing students.

MOOC DATA SET

The data for this study come from the Spring 2012 Circuits and Electronics MOOC on the MITx platform. Descriptive measures of discussion forum usage, homework performance, and final exam scores were extracted from the MOOC clickstream logs using parsers written in Python [22]. Over 100,000 students registered for this course, though only half as many attempted to solve at least one problem in the course. Roughly 9000 attempted at least one problem on the final exam, and 7157 earned certificates.

Each access by a student to the discussion forum was recorded in the click-stream logs of the MOOC, as were the times when the student first opened each weekly homework assignment and the time of the last submit (the “homework window”). Thus it was straightforward to enumerate the number of threads viewed each week during the homework window. In this course, the most commonly referenced resource during homework solving was the discussion forum [22], which was structured as a Q&A board with up-voting and search capability (other course resources included lecture videos, an online textbook, and a wiki). Interestingly, most of this activity was “voyeuristic” not contributive: 67% of active students viewed (that is, clicked on—without scroll information and/or eye-tracking sensors, one cannot say for sure whether students read the threads they opened) at least one discussion thread between the first time they opened the homework and their last submission, whereas fewer than 10% posted a question, comment, or answer. Moreover 95% of all discussion activity in this course (by number of events) was viewing, not posting.

Because discussion forum content was generated by students, the forum was not as rich in the first few weeks of the course until participation reached a critical level, as shown in Figure 1.

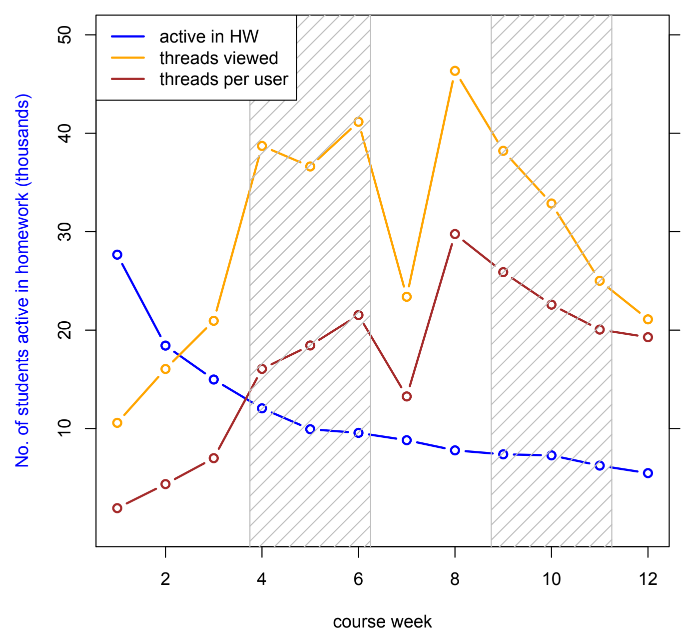

Figure 1: MOOC activity over time. Grey bars indicate early stage and late stage intervals on either side of the midterm.

As seen in this figure, the number of students actively doing homework in our data set (active in homework, blue line) decays over time, while activity in the forums increases before leveling off (threads per user, brown line). The midterm exam occurred between weeks 7 and 8, which explains the dip and then surge in discussion forum activity, as it was not permitted to post questions or answers about the midterm. The greyed regions of Figure 3 represent two three-week intervals, which we label "early stage"—weeks 4-6, after the discussion forum had fully taken off but before the midterm—and "late stage"—weeks 9-11, after the midterm but before the final exam. To smooth out week-to-week variation, we summed over views within each three-week long interval, as discussed below.

CHALLENGES IN OPERATIONALIZING PREDICTORS

MOOCs differ from standard courses in a number of ways that make analyzing enrollee behavior difficult. These include higher than usual variability in prior educational attainment [20] and assessment motivation [26], large amounts of missing data, and affordances of multiple attempts on both formative and summative assessments [4]. Due to these issues, several researchers have noted that traditional measures of participation and achievement may need to be reconsidered in the context of MOOCs [5,7,13,16]. In this section, we introduce three sets of challenges, one for each predictor variable:

1. How can *prior ability* be estimated so that performance models can control for prior ability?
2. How should *discussion forum usage* be quantified? Is it a static quantity, or does it change over time?
3. Can we identify students who appear to be *disengaged/unmotivated*? What effect would excluding those students have on the effect size of forum usage?

Prior Ability

Enrollees in MOOCs range from high school students to professionals with earned doctorates [20]. Because overall performance is likely to depend on prior ability, this factor should be accounted for in any analysis of "treatment effects" from discussion forum usage. However, prior ability is typically unavailable information. Not all MOOCs survey incoming students, and those that do often survey sparsely. Enrollees in the Spring 2012 Circuits and Electronics MOOC were not given a pretest. Therefore, prior ability had to be inferred from the course data. In this study, we chose to estimate prior ability levels from performance on homework assignments in the first three weeks of the course, when enrollees had just begun to learn the content and before discussion forum use had taken off. The main idea was that early stage ability estimates were not likely to be affected by discussion forum usage, whereas final exam performance might be.

Because homework assignments allowed an unlimited number of attempts, the variability of the eventually correct (EC) score (the official score of record) was quite low. However, scoring items based on whether they were solved correctly on the first attempt (CFA) resulted in a far more normal distribution (see Figure 2). A host of options for scoring homework in the presence of missing data and multiple attempts was described in [4]. While approaches based on polytomous item response models were most predictive of final exam scores, a reasonable improvement of the EC score was obtained for observed scores based on CFA. For simplicity, we use the mean CFA score, which is the proportion of homework problems attempted by each enrollee in the first three weeks of the course that were solved correctly on the first attempt. Skipped items are ignored, rather than scored as incorrect. For detailed considerations of homework scoring in MOOCs, we refer the reader to [4].

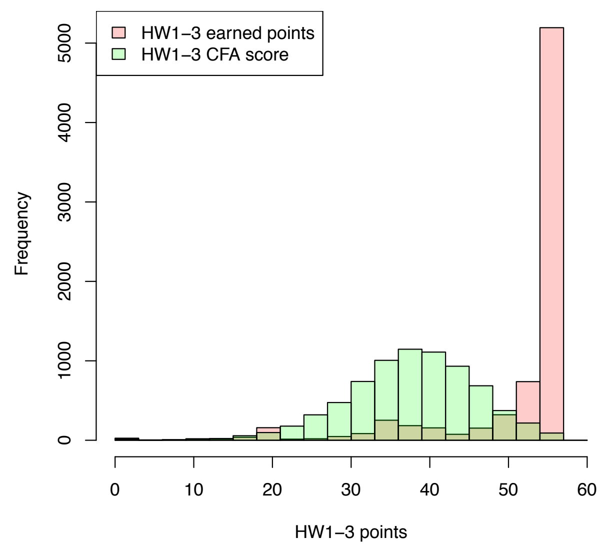

Figure 2: EC and CFA score distributions

It should be noted that the issues of homework scoring also arise in the final exam, which is our outcome measure. We do not consider alternate scoring options, e.g. CFA scoring or item response theory, for the final exam. Only three attempts were allowed versus unlimited attempts on homework, and we did not want to punish students for strategically using their available attempts. However, there remain issues of examinee motivation, as discussed below.

Discussion Forum Usage

The average number of threads viewed per week was shown in Figure 1. We now explore the distribution over MOOC users of the early stage and late stage intervals (grey regions in Figure 1; the purpose of summing was to smooth out week-to-week variation.) We are interested in knowing both the distribution of counts within each interval—e.g. is it simple or bimodal?—as well as across the intervals—i.e. do learners exhibit consistent discussion usage over time, or does it change? These are important considerations for modeling the effect of discussion views. Consider students who purposefully increase their reference to forums after the midterm and reap performance gains as a result. Modeling their usage as constant over time would distort the positive effect.

As shown in Figure 3, the early/late view count variables are of mixed type: many students do not view any threads, but among those who view at least one, the counts are roughly log-normally distributed. We have added 0.37 to all counts, such that after log-transformation, the students with zero counts appear in the disjoint bin at -1. As seen in the figure, there are roughly 1600 students in this bin for both early stage and late stage counts.

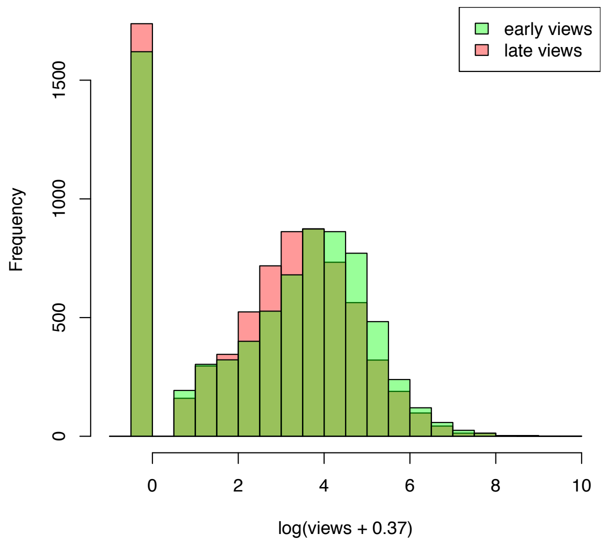

Figure 3: Distribution of view counts (log-transformed)

Figure 3 does not reveal whether there are students who significantly increase or decrease their discussion viewing between these time periods. Moreover, determining what amount of change is significant is a subtle point.

To address this question, we plot early view counts (scaled) against the difference between early and late counts (also both scaled) in Figure 4. Scatterplot and point density are both shown. There is a floor effect, which appears as a diagonal lower bound in the figure, representing students who went from a finite number of threads viewed in the early period to zero in the late period. Another salient feature is that for medium to large values of early counts, the change (from early to late counts) seems to be a random effect around zero (no change). This random description does not however fit all of the data. There does appear to be a clump of students on the upper left, whose viewing counts increase from very low levels to moderate levels. And there are some whose viewing decreases beyond the noise threshold. We chose to identify these students as outliers from the random distribution.

We determined empirical means and variances after removing low values and then drew a random sample of 7000 data points from a bivariate normal distribution with center $\mu = (4.17, -0.27)$ and with covariance matrix $\Sigma = (1.15, 0, 0, 0.84)$. Elliptical contours are drawn at the 95% and 99% confidence level in the figure. We have also included reference lines at the vertical mean value plus and minus log(2). The purpose of this second boundary is to define a criterion for those students whose early view counts were extreme outliers but whose change was still modest. Since the vertical axis is a difference of logarithms (or the log of the ratio), points outside this inner region represent doubling (or halving) in the counts.

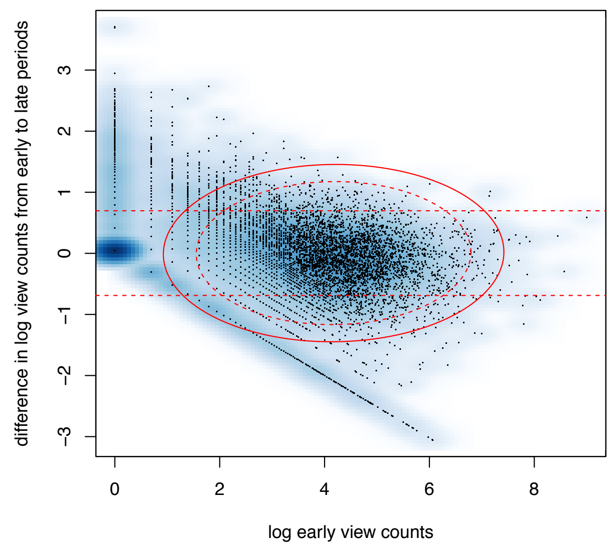

Figure 4: Change in discussion view counts against early counts. Ellipses denote 95% and 99% confidence intervals around a bivariate normal uncorrelated distribution. Dashed lines at +/- log(2) denote doubling thresholds.

As a result of this exploratory analysis, we divided our initial population into an *overall group* (N = 6505), whose discussion viewing during homework could be seen as unchanging over time and thus aggregated into a variable V_O, and a *change group* (N = 989), whose viewing change V_C should be modeled instead. V_O is the sum of the early and late stage counts, and V_C is the difference. Each would subsequently be treated as a continuous variable in an overall model or a change model, respectively.

Assessment-oriented Engagement and Total Time as Proxy Measures of Motivation

Inferences about ability from standard measures of performance may not always be valid in a MOOC due to differences in enrollees' motivations for taking the course. The expectancy-value model [9] puts the validity problem as follows: achievement motivation is influenced by both the individual's expectancies for success and the subjective value attached to success on the task. If the value of success is low, the examinee's achievement motivation will be low. Motivation thus acts as a source of construct-irrelevant variance and impacts the validity of score-based inferences [10]. In a meta-analysis of twelve empirical studies, [26] found that motivated students scored on average 0.59

standard deviations higher than their unmotivated counterparts. Such a result highlights the need to evaluate examinee motivation and possibly filter data from unmotivated test-takers to strengthen the assumption that a score obtained from an assessment accurately reflects the underlying abilities/traits of interest [1].

Consider the final exam score, which typically counts heavily toward qualification for a certificate (in the course under study, the final counted for 40% of the cumulative grade). However, the MOOC certificate is largely symbolic when it confers no degree credit. Thus, enrollees whose motivations for taking the course do not include certification may well view the final exam as low-stakes. The consequentiality of certificates may, in fact, change as more MOOCs seek accrediting status and even charge fees accordingly.

In the following, we consider three solutions to this problem, which is essentially the problem of whom to include. The first is to use a heuristic cutoff with respect to proportion of items attempted in the initial and final ability assessments. In the second solution, we attempt to filter out unmotivated students using a simple measure that should be relatively insensitive to the initial and final assessments, namely total time spent online in the course. The third and most intricate solution will be to use a latent class cluster analysis to model the course population as a mixture of classes based on cumulative evidence of assessment-oriented engagement (AOE). Thus both AOE and time-on-task are effective proxy measures for motivation, but we continue to use the original term in order to make contact with validity literature.

Motivation heuristic filter on attempts

Screening out students who attempted less than 60% of the HW1-3 items (which constitute our proxy measure of "prior ability") or less than 60% of the final exam leaves 6210 students. This proportion is chosen to match the passing grade threshold of the course; in order to achieve this minimum, a student must at the very least attempt the same fraction of assessment items. This cutoff ignores the proportion of attempts on items in between Week 3 and the final exam, which will enter into the latent class analysis.

Although this is a filter based on attempts and not scores, it raises selection bias issues. While low-performing students who at least attempted many items would remain, this filter does, by definition, remove low scoring students. Thus our proxy for motivation is wrapped up in the outcome variable of our analysis. The rationale for solution two is partly a response to the bias of solution one.

Motivation heuristic filter on time

What if there were students who invested significant amounts of time and effort in this course but were simply unable to answer many questions and were disinclined to guess? Alternately, what if there were students who carelessly attempted many items, but whose investment in the course was more accurately reflected in low overall time commitment. Rather than filter on proportion of assessment items, we considered overall time spent in the course as a proxy for motivation. All activity, including video views, was included in this time aggregate, which is roughly log-normally distributed (slightly skewed to the left) with a median value around 100 hours. At a minimum time cutoff of 30 hrs (~1.5 standard deviations below), 679 students would be excluded, leaving 6815.

Motivation via latent class analysis of AOE

In the third approach, rather than determine whom to include or exclude, we seek to identify self-similar groups of students based on a pattern throughout the course. We could then model the effect of discussion viewing separately for all groups. Our idea is related to the approach in [16], where week-by-week trajectories were clustered. The results of that analysis were largely interpreted in terms of proportion of assessment attempted, so we went directly to that measure as a basis for clustering. We used five measures based on proportion of assessment items attempted: homework in weeks 1-3, homework in weeks 4-6, midterm exam, homework in weeks 9-11, and final exam. Each student's record of item attempts was thus mapped to a vector of five proportions, and these vectors were clustered using the Gaussian mixture model-based clustering algorithm in the MClust package [11] in R.

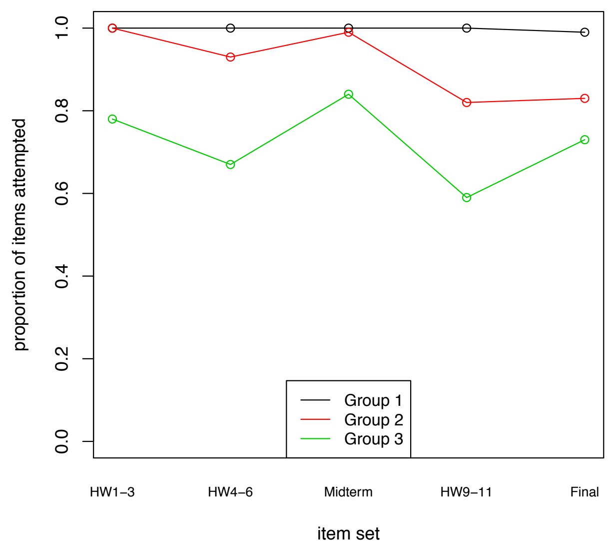

Figure 5: Mean values of proportion of items attempted for three latent class cluster groups.

The model-based approach used here differs from the clustering method in [16], but the results are consistent. The best fit was at three clusters. Mean values for proportion of items attempted are plotted in Figure 5. Groups 1-3 roughly correspond to what [16] called completing, disengaging, and sampling. Probably because we removed in advance students who did not attempt at least one final exam problem, we do not have an auditor group, typified by students who watch videos but do not attempt any assessment items.

SUBSTANTIVE ANALYSES

Having operationalized our predictors, we now turn to modeling the effect of discussion viewing on final exam performance. Using multiple linear regression, we examine the standardized regression coefficient for the discussion viewing term as a probe of effect size. Based on the exploratory analyses described above, discussion viewing was treated differently for those students whose usage levels were consistent overall versus those who changed their viewing amount between the early and late stages. We computed two different variables V_O and V_C for these two populations respectively. Variability in motivation was handled both through heuristic attempt-based and time-based filters as well as via latent class analysis.

Model and results using motivation filters

Consider the following linear model for predicting the final exam Y using prior ability θ and *overall* discussion view counts V_O,

$$Y = \beta_0 + \beta_1\theta + \beta_2 V_O$$

The change model is identical except for the substitution of view change for overall views. Importantly, the populations included for each model are different, as described above.

Table 1 reports standardized regression coefficients β_2 for these two models. The first column is the result when including all students who attempted at least one final exam problem and one homework item in weeks 1-3 (HW1-3 performance was the basis for estimating prior ability θ). The middle column shows results when excluding students who spent fewer than 30 hours online. The last column shows results excluding those who did not attempt at least 60% of *both* the final exam and the weeks 1-3 homework.

Table 1: Standardized regression coefficients for discussion viewing factor in two models under different data thresholds (white cells p < .001; grey cells not significant)

	No filter	Time > 30h	Attempt > 60%
Overall β_2	0.24	0.18	-0.01
Change β_2	0.19	0.19	0.16

The effect of discussion viewing in the overall model (first row of Table 1) appears to be significant when no filter is applied. But this unfiltered population contains hundreds of students who attempted very few assessment items, so these coefficients are not necessarily trustworthy. Indeed, the effect of overall viewing starts to decline as the population is refined in the next two columns. Screening out students who spent comparatively little time in the course reduces the effect but not by much. On the other hand, after screening out students who did not *attempt* at least 60% of those assessment items that formed the basis of the prior and outcome performance measures, the effect of discussion viewing disappears entirely.

At the least, it must be said that the effect size of discussion viewing in the overall model is sensitive to selection of students. We note that these models altogether explain only about 10% of the variance in the final exam. The midterm exam, for reference, is more predictive ($R^2 = 0.22$).

The effect of discussion views in the change model (second row), in contrast, appears to be more robust under selection for motivated students. At first glance, it is not clear whether increases in viewing are translating into higher scores or decreases in viewing are translating into lower scores. The latter could be consistent with attrition, for example. However, if attrition were the dominant explanation, then the third column coefficient would also be small, since course droppers would have been screened out. Thus the change model coefficients suggest that increasing discussion views are associated with higher final scores. We believe that interpretation of this effect is improved with reference to the latent class models, described next.

Model and results for latent class analysis

Table 2: Standardized regression coefficients for the overall viewing model with latent class cluster groups (white cells, p < .005; grey cells are not statistically significant)

$Y = \beta_0 + \beta_1\theta + \beta_2 V_O + \beta_3 G + \beta_4 \theta G + \beta_5 V_O G$							
	0.75	0.14	-0.09	0	0	0	G=1
				-0.76	0.05	0.05	G=2
				-0.96	0.09	0.53	G=3

In Table 2 we show the model equation and estimated parameters for overall viewing effect with latent class cluster assignments. There were significant interactions between the cluster groups G and the continuous prior ability and discussion variables for the overall model; therefore we include five coefficients. Group 1, the reference group, attempted almost all assessment items (see Figure 5). Because Group 2 and 3 attempted fewer items, the main effect for those groups ($\beta_3; p < .001$) is a lower expected final exam score. Indeed, Group 1 may be thought of as a more restrictive subsample from the third column of Table 1. The interpretation of this small negative β_2 is not necessarily that discussion views hurt, of course. Among Group 1 students, more viewing may indicate challenges with homework that transfer into challenges on the final.

Given that students in Group 3 omitted significant numbers of assessment items, why would such students reap more rewards from viewing discussion threads (β_5)? A possible explanation is that discussion viewing is a proxy for activity within Group 3. Indeed, there were positive correlations between overall views and final exam items *attempted* (0.38) as well as late-stage homework *attempted* (0.53). Students who viewed more also did more assessment items relative to other students in this group.

Finally, Table 3 shows the change model with latent classes. Comparing to the second row of Table 1, we see now that for Group 1, increasing views are no longer associated with higher final exam scores. Recall that this group comprises the most active population with respect to assessment items. Again, a plausible explanation is that increasing discussion views are simply an indication of increasing participation in Groups 2 and 3, for example due to late joiners to the course. The correlation between viewing change and final exam items attempted is low in both cases (roughly 0.06), but the correlation with late homework attempted is moderate (0.27 and 0.33 for Groups 2 and 3, respectively). For the sporadic users of assessment in these groups, the positive association of increasing discussion views over time is there, but it may be linked to increasing engagement with the homework.

Table 3: Change model including latent class cluster groups (white cells, p < .05; grey cells are not statistically significant)

$Y = \beta_0 + \beta_1\theta + \beta_2 V_C + \beta_3 G + \beta_4 \theta G + \beta_5 V_C G$							
	0.76	0.18	-0.05	0	0	0	G=1
				-0.80	-0.06	0.22	G=2
				-1.14	-0.15	0.21	G=3

CONCLUSIONS AND FUTURE WORK

We started out with a simple goal of studying the learning outcome benefit from viewing discussion threads while doing homework in a MOOC. Along the way, it became clear that operationalizing almost all of the variables in this equation presented challenges. We have considered solutions to several issues that are endemic to MOOCs: estimating prior ability; determining whether to use an overall or a change model of discussion viewing; and screening out unmotivated students for the purpose of increasing the validity of inferences.

In the end, neither overall discussion viewing (for those whose viewing was fairly steady) nor change in discussion view volume appeared to be significant for students who attempted most of the assessment items, i.e. Group 1. The gain that appears from a naïve application of a linear model to the larger student sample (Table 1, column 1) seems to be due to confounding discussing thread viewing with participation, among sporadic participants. More work would need to be done to decouple use of the discussion forum from assessment-oriented engagement, for example by treating the latter as a continuous measure rather than as an indicator on which to filter the population. Moreover, counting discussion thread views is a limited window into usage of the forums. We did not analyze posting or commenting in this analysis, nor did we discriminate between threads using textual analysis.

We did not say much about why the effect size of discussion viewing seemed insensitive to filtering students by overall time spent online. We suspect this is because there were hundreds of students who scored very highly on the final exam in this course but spent almost no time learning; in other words, these students already knew the content, but took the tests for fun or for the certificate.

As suggested above, we suspect that late joiners—whose increasing viewing over time appeared to associate with score gains—were a foil in this analysis. It would be interesting to dig deeper into how to model students whose trajectories of participation are increasing or decreasing over time. Also, although we used the final exam because it was an obvious choice, it may be possible to model the effect of discussion viewing on homework performance directly. There are subtleties to this, because multiple attempts increase the likelihood of correct responses. From a learning science perspective, looking at how students search the forums to get homework assistance may also be a fruitful direction.

ACKNOWLEDGMENTS

We are grateful to edX for providing the raw data for this analysis, to Daniel Seaton for critical contributions to the processing of these data, and to helpful suggestions from reviewers. DEP would like to acknowledge support from a Google faculty award and from MIT.

REFERENCES

1. AERA, APA, NCME. *Standards for educational and psychological testing.* American Educational Research Association, Washington, D.C., 1999.
2. Andresen, M. Asynchronous Discussion Forums: Success Factors, Outcomes, Assessments, and Limitations. *Educational Technology & Society 12*, (2009), 249–257.
3. Bates, A.W.T. *Technology, E-Learning and Distance Education.* Routledge, 1995.
4. Bergner, Y., Colvin, K., and Pritchard, D.E. Estimation of Ability from Homework Items When There Are Missing and/or Multiple Attempts. *Proceedings of LAK 2015*, (2015).
5. Clow, D. MOOCs and the funnel of participation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge Discovery*, (2013), 185–189.
6. Coetzee, D. and Hearst, M.A. Chatrooms in MOOCs : All Talk and No Action. (2014), 127–136.
7. DeBoer, J., Ho, A.D., Stump, G.S., and Breslow, L. Changing “Course”: Reconceptualizing Educational

Variables for Massive Open Online Courses. *Educational Researcher March*, (2014), 74–84.

8. Dringus, L.P. and Ellis, T. Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education 45*, 1 (2005), 141–160.
9. Eccles, J.S. and Wigfield, A. Motivational Beliefs, Values, and Goals. *Annual Review of Psychology 53*, (2002), 109–132.
10. Eklof, H. Development and Validation of Scores From an Instrument Measuring Student Test-Taking Motivation. *Educational and Psychological Measurement 66*, 4 (2006), 643–656.
11. Fraley, C. and Raftery, A.E. Model-based Clustering, Discriminant Analysis and Density Estimation : *Journal of the American Statistical Association 97*, (2002), 611–631.
12. Gillani, N. and Eynon, R. Communication patterns in massively open online courses. *The Internet and Higher Education 23*, (2014), 18–26.
13. Ho, A.D., Reich, J., Nesterko, S.O., et al. HarvardX and MITx: The First Year of Open Online Courses, Fall 2012-Summer 2013. *SSRN Electronic Journal*, (2014).
14. Huang, J., Dasgupta, A., Ghosh, A., Manning, J., and Sanders, M. Superposter behavior in MOOC forums. *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14*, (2014), 117–126.
15. Kay, R.H. Developing a comprehensive metric for assessing discussion board effectiveness. *British Journal of Educational Technology 37*, 5 (2006), 761–783.
16. Kizilcec, R., Piech, C., and Schneider, E. Deconstructing Disengagement : Analyzing Learner Subpopulations in Massive Open Online Courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge Discovery*, (2013).
17. Kortemeyer, G. Correlations between student discussion behavior, attitudes, and learning. *Physical Review Special Topics - Physics Education Research 3*, 1 (2007), 010101.
18. Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R.S., and Hatala, M. Penetrating the black box of time-on-task estimation. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, ACM Press (2015), 184–193.
19. Ramos, C. and Yudko, E. "Hits" (not "Discussion Posts") predict student success in online courses: A double cross-validation study. *Computers & Education 50*, 4 (2008), 1174–1182.
20. Rayyan, S., Seaton, D.T., Belcher, J., Pritchard, D.E., and Chuang, I. Participation And performance In 8.02x Electricity And Magnetism: The First Physics MOOC From MITx. (2013), 4.
21. Rowntree, D. Teaching and learning online: a correspondence education for the 21st century? *British Journal of Educational Technology 26*, 3 (1995), 205–215.
22. Seaton, D.T., Bergner, Y., Chuang, I., Mitros, P., and Pritchard, D.E. Who does what in a massive open online course? *Communications of the ACM 57*, 4 (2014), 58–65.
23. Soller, A. Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in ... 12*, 1 (2001).
24. Song, L. and McNary, S. Understanding students' online interaction: Analysis of discussion board postings. *Journal of Interactive Online Learning 10*, 1 (2011), 1–14.
25. Thomas, M.J.W. Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning 18*, 3 (2002), 351–366.
26. Wise, S.L. and DeMars, C.E. Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. *Educational Assessment 10*, 1 (2005), 1–17.