

Learning is Not a Spectator Sport: Doing is Better than Watching for Learning from a MOOC

Kenneth R. Koedinger
Carnegie Mellon University
5000 Forbes Avenue
koedinger@cmu.edu

Jihee Kim
Carnegie Mellon University
5000 Forbes Avenue
smile.jihee@gmail.com

Julianna Zhuxin Jia
Carnegie Mellon University
5000 Forbes Avenue
zhuxinj@andrew.cmu.edu

Elizabeth A. McLaughlin
Carnegie Mellon University
5000 Forbes Avenue
mimim@cs.cmu.edu

Norman L. Bier
Carnegie Mellon University
5000 Forbes Avenue
nbier@cmu.edu

ABSTRACT

The printing press long ago and the computer today have made widespread access to information possible. Learning theorists have suggested, however, that mere information is a poor way to learn. Instead, more effective learning comes through doing. While the most popularized element of today's MOOCs are the video lectures, many MOOCs also include interactive activities that can afford learning by doing. This paper explores the learning benefits of the use of informational assets (e.g., videos and text) in MOOCs, versus the learning by doing opportunities that interactive activities provide. We find that students doing more activities learn more than students watching more videos or reading more pages. We estimate the learning benefit from extra doing (1 SD increase) to be more than six times that of extra watching or reading. Our data, from a psychology MOOC, is correlational in character, however we employ causal inference mechanisms to lend support for the claim that the associations we find are causal.

Keywords: Learning by doing; MOOCs; learning prediction; course effectiveness; Open Education; OER

ACM Classification Keywords K.3.1

INTRODUCTION

The use of online learning resources to provide and support instruction is on the rise and spectacularly so [5]. Further, there is a growing recognition and interest in the opportunity to apply and contribute to the learning sciences by conducting education research through online learning environments [29, 16]. While there are notable

success stories of online courses that were shown to be more effective than traditional instruction [e.g., 20], the more typical situation is that, at best, online courses achieve the same outcomes at lower cost [e.g., 6, 9]. Perhaps more importantly, we do not know enough about what features of online courses are most important for student learning.

The prototypical feature of a Massive Open Online Course (MOOC) is lecture videos, but many MOOCs also include activities such as questions for students to answer or problems for them to solve, in some cases with immediate online feedback. What is more important for student learning? Is it the information students get from watching lecture videos, the practice and feedback they get from online questions or problems, or some combination. The idea of scaling the best lecturers for open access is a compelling feature of MOOCs. However, in contrast to the passive form of learning characterized by watching video lectures or reading text, a diversity of learning theorists have recommended more active learning by doing [e.g., 3, 11, 28]. Many argue for learning by doing as it focuses on authentic activities that are more representative of knowledge use in the real world. More fundamentally, learning by doing is important because most of human expertise involves tacit knowledge of the cues and conditions for deciding when, where, and what knowledge to bring to bear in complex situations [36]. In this view, there appears to be no verbal shortcut to acquiring expertise. It is gained by observing examples (or “models”), attempting to engage in expert activities with feedback and as-needed instruction (“scaffolding”), and having that support be adaptive to advancing proficiency (“fading”) [cf., 10].

Intelligent tutoring systems provide such support through adaptive feedback and hints during learning by doing [1, 2, 34]. When well designed, they yield significant learning gains [e.g., 24, 27, 35]. “Well-designed” means designed using theory of learning and instruction [cf., 14, 26], data-driven methods including cognitive task analysis

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
L@S 2015, March 14–18, 2015, Vancouver, BC, Canada.
ACM 978-1-4503-3411-2/15/03.
<http://dx.doi.org/10.1145/2724660.2724681>

[8, 17], and classroom design research iterations [4, 19]. Interactive activities in Carnegie Mellon University's *Open Learning Initiative* courses attempt to mimic some of the behavior of intelligent tutors and, importantly, follow these good design practices [21, 31].

In the past few years, much attention has been paid to the demonstrated potential of Massive Open Online Courses (MOOC) to scale access to educational resources [25].

While the exact definition and role of MOOCs continues to be debated, much of the popular dialogue surrounds the three major MOOC platforms – Coursera, Udacity and EdX – who describe MOOCs as online courses with open, unlimited enrollment [23]. Though specific activities vary from course to course, video-based lectures and student discussion forums typically form the core of the MOOC instructional experience. We call this the “lecture model”. The Open Learning Initiative, which has offered online learning environments since 2002, takes a different approach focusing on rich and interactive learn-by-doing activities, aligned with student-centered learning outcomes, and designed around science-based learner models. We call this the “learn-by-doing model.” Both models offer rich datasets, though with different focuses and capturing different kinds of learner interactions.

An opportunity emerged recently to compare the instructional features of these two different models in terms of how variations in student use of them impacts the learning outcomes they achieve. In 2013, Georgia Institute of Technology and Carnegie Mellon University (CMU) collaborated to incorporate elements of CMU's Open Learning Initiative (OLI) “Introduction to Psychology” learning environment into Tech's *Introduction to Psychology as a Science* MOOC. Taught via the Coursera platform, OLI materials were available as part of the larger course, in addition to lectures, quizzes and other Coursera-based activities. This paper explores the impact of the use of the OLI elements on learning, in comparison to use of MOOC elements (alone or in association with the OLI materials). As part of this exploration, we examine which OLI features are most associated with learning and if we can infer causal relationships between these features. In addition, we analyze the potential for predicting the course dropout using student performance in OLI activities and survey-based demographic information.

The key research questions we pursue are:

1. What factors determine whether or not students stay in the course or dropout?
2. Do students who use OLI features learn more than those using the MOOC only?
3. What variations in course feature use (watching videos, reading text, or doing activities) are most associated with learning? And can we infer causal relationships?

A key goal is to provide evidence relevant to alternative hypotheses about what makes a course effective. The “lecture model” suggests that students' primary source for learning is through listening to lectures. The “learn-by-doing model” suggests that students' primary source for learning comes from answering questions and solving problems with feedback. Of course, it may be that both sources are critical.

COURSE FEATURES AND DESIGN

In considering the features available in the course, we divide components into two broad categories: passive/declarative information and active/interactive activities. Students learn from passive/declarative information by reading, watching or studying; these features include video lectures, lecture slides and other expository materials (text). Active/interactive features by definition require students to be more active, and include quizzes, exams, discussion forum participation and interactive activities that provide targeted feedback and hints. Although engagement with the full range of learning materials assigned was encouraged, the final grade in the course was awarded based upon a combination of quiz scores, final exam score and two written assignments. In addition, the course contained two additional features intended to support research rather than learning: a pre/post test and student background survey. Neither of these elements was factored into students' grades.

Introduction to Psychology as a Science was designed as a 12-week introductory survey course, as is often taught during the first year of college. For each week of class, the course targeted a major topic area (e.g. Memory, Sense and Perception, Abnormal Behavior, Brain Structures and the Nervous System); these topics were broken into 3-4 sub-topics, each supported by a pre-recorded video lecture (10-15 minutes, with downloadable slides) and included assigned modules and learning outcomes in the OLI learning environment. A high-stakes quiz assessed students against these outcomes at the end of each week.

The **Coursera MOOC platform** provided general course structure (registration, syllabus, etc), video lectures and slides, discussion forums, writing assignments, quizzes (with questions drawn from OLI item banks, see Figure 1b for an example), and a final exam (with questions created by the instructor, see Figure 1c for an example). The background survey was also administered via this platform, which focused on demographic information (gender, age, education, occupation) as well as some questions to assess learner intent and opinion.

The **OLI Learning Environment** was embedded in the Coursera platform using Learning Tools Interoperability (LTI) for a seamless learner experience. The corresponding OLI modules included a variety of

expository content (text, examples, images, and video clips) and a large number of interactive activities. Broadly, these activities serve two purposes. “Learn By Doing” activities, intended to support student outcome achievement, provide feedback targeted to diagnose misconceptions and robust hints to support students. In Figure 1a, we show a screenshot of a Learn by Doing activity from the unit on Personality covered in week 9 of the course. “Did I Get This” activities provide a self-comprehension check for students. They are introduced at points when students are expected to have achieved mastery and do not provide hints, though they do offer feedback [31].

Complete the table below by dragging each of the major factors of personality based on the Five-Factor (Big Five) Model of Personality to their proper location, between the corresponding traits of both extremes. Note that each factor represents a dimension, or range, between two extremes.

Low Extreme Traits	Factor	High Extreme Traits
Calm, even-tempered, unemotional, hardy	Neuroticism	Worrying, temperamental, emotional, vulnerable
Reserved, loner, quiet	Openness to experience	Affectionate, joiner, talkative
Down-to-earth, conventional, uncreative, prefer routine		Imaginative, original, creative, prefer variety
Antagonistic, ruthless, suspicious		Sympathetic, softhearted, trusting
Lazy, aimless, quitting		Hardworking, ambitious, persevering

Extraversion Conscientiousness Agreeableness

Hint: Someone who is high in extraversion is outgoing and decisive, whereas someone low on this trait is retiring and withdrawn.

That's incorrect. You are looking here for the factor that is associated with the traits that describe the low and high ends of that dimension.

(a) OLI Activity

Which of the following is not one of the five groups in the five-factor model used to describe personality?

A. Agreeableness
B. Intelligence
C. Neuroticism
D. Extraversion

(b) Quiz question

Which of the following is an example of a projective personality test?

Thematic Apperception test
Minnesota Multiphasic Personality Inventory
Myers-Briggs Type Indicator
Sixteen Personality Factor Questionnaire

(c) Final Exam question

Figure 1. (a) Screen shot of a Learn By Doing OLI activity from the unit on Personality © OLI. (b) Corresponding quiz question © OLI. (c) Related final exam question © Dr. Anderson Smith, GA Institute of Technology.

METHODS

It is important to point out that using data from natural student use of MOOCs adds uncertainty in making inferences about causal relationships as compared to using data from experimental designs. This uncertainty is further increased by the large attrition or dropout that is typical in MOOCs. The sample of students involved in any particular analysis is determined by student participation and effects that might be attributed to other

factors (e.g., course features) might instead be so-called “selection effects”, that is, effects of sampling differences based on the choices or selections that students make. Nevertheless, there is a real opportunity to use the large and naturally-occurring data that comes from MOOCs to provide initial, if not confirming, evidence of factors of potential importance for course participation and learning outcomes.

Table 1 shows different subsets of students as indicated by different forms of participation in the course. We refer to it in describing how samples were selected to address our research questions.

Our first research question is: What factors determine whether or not students stay in the course or dropout? 27720 students registered in the Coursera MOOC Psychology course while 1154 students completed it (see Table 1). We are interested in what indicators or features may predict dropouts throughout the course, and we use quiz and final exam participation as estimates of student dropout. For example, if a student has a score for quiz 4 but none of the remaining quizzes or the final, we consider that student to have dropped out after quiz 4. We are interested in factors that predict future dropouts. In addition to whether students used the OLI material or not, we also included quiz participation and quiz score in a logistic regression model to predict final exam participation.

Our second research question is: Do students who use OLI learn more than students who only use the MOOC materials? MOOC+OLI students (N=9075) are those who registered to use the OLI materials. MOOC-only students (N=18,645) did not (see Table 1). To address the question, we did a quasi-experimental comparison of learning outcomes between the MOOC+OLI students who took the final (N=939) with the MOOC-only students who took the final (N=215).

Our third research question is: What variations in course feature use (watching videos, reading text, or doing activities) are most associated with learning? And can we infer causal relationships? In the results section, we describe an exploratory data analysis to identify relationships between usage of these features (gathered from the log data [15]) and our two measures of learning, quizzes scores and final exam score. To frame that analysis, we present some global data on feature usage. Of all MOOC registrants, 14,264 (51.4% of total) started to watch at least one lecture video. Of the 9075 students (32.7% of total) registered for OLI material study, 84.5% (7683 students) accessed at least one page of OLI readings and visited or revisited an average of 69 pages with a maximum of 1942 pages (variable *pageview*). On average, 33 unique pages were viewed with a maximum of 192 unique pages. Of the 9075 OLI registered students, 62.3% (5658 students) started at least one interactive

	Students	Percent total (percent subgroup)	Average Score or Feature Usage/Max
All students	27720	100%	
Pre-test	12218	44.89%	6.9/20
Quizzes	2373 ¹	8.7%	7.1 /10
Final	1154	4.2%	25.6/ 40
MOOC only	18645	67.3% (100%)	
Pre-test	4872	17.6% (39.9%)	5.8/20
Quizzes	496	1.8% (4.1%)	6.3/10
Final	215	1% (1%)	22.8/40
OLI registered	9075	32.7% (100%)	
Pre-test	7346	26.5% (80.9%)	8.6/20
Quizzes	1876	6.8% (20.7%)	7.5/10
Final	939	3.4%	26.3/40
<i>Playing video</i>	902	(96.1%)	164.4/4460
<i>Reading pages</i>	939	(100%)	296.8/1942
<i>Doing activities</i>	939	(100%)	387.2/695
<i>Non-activity pages</i>	939	(100%)	182/1759

Table 1. Student participation in assessments and in course features; average score or usage.

¹6912 students took quiz 1 and 1136 students took quiz 11. 2374 is the average across all 11 quizzes.

activity with an average of 127 and a maximum of 695 activities started (variable *activities started*). Table 1 shows these same activity use statistics for the subset of OLI registered students who took the final exam (939 students).

Because activities sit on pages and students must go to those pages to do the activities, we created a new measure to represent pages accessed beyond those needed to get to

a page. A scatterplot between *pageview* and *activities_started* indicated a lower bound on pages seen for a given number of activities and the bound is reasonably estimated at the maximum of 695 activities, where no one did this many activities in fewer than 206 pageviews. We used this ratio of about 3.4 activities per page to subtract out the pages due to activity access and computed a new variable, *non activities pageview*, that is arguably a more pure measure of the variation in the reading students did.

After the exploratory data analysis, the results present a search for potential causal relationships, including estimates of the strength of these relationships, by using a causal inference system called Tetrad [32]. Tetrad is a program that creates, estimates, tests, predicts with, and searches for causal and statistical models. Tetrad has a distinct suite of model discovery search algorithms that include, for example, an ability to search when there may be unobserved confounders of measured variables, to search for models of latent structure, to search for linear feedback models, and to calculate predictions of the effects of interventions or experiments based on a model¹. We used Tetrad to infer the causal relationships between pretest score, course feature use variables (watching videos, reading pages, doing activities), quiz scores, and final exam score. To aid causal model search, we specified some extra time-based constraints, particularly that leader measures could not use earlier ones (i.e., the final scores cannot cause quiz scores, quiz scores cannot cause feature use, and feature use cannot cause pre-test score). We used Tetrad's PCD algorithm for causal model search and we normalized the data (using a correlation matrix input to PCD) so that resulting model coefficients could be better compared and interpreted.

RESULTS AND DISCUSSION

Predicting Course DropOut and Completion

From the total number of registered students (27,720), only 4% take the final exam. Of the 9075 students who registered for the OLI materials, only 10% took the final exam. Such high attrition rates in MOOC courses are not uncommon [12] as many students register for reasons other than completing the course (e.g., determine their interest in the material, check their knowledge on the topic, etc.). But, we were interested in examining when students dropout (i.e., is there a crucial point in the course for students to not continue to participate) and what factors might contribute to dropping out.

¹ <http://www.phil.cmu.edu/tetrad/>

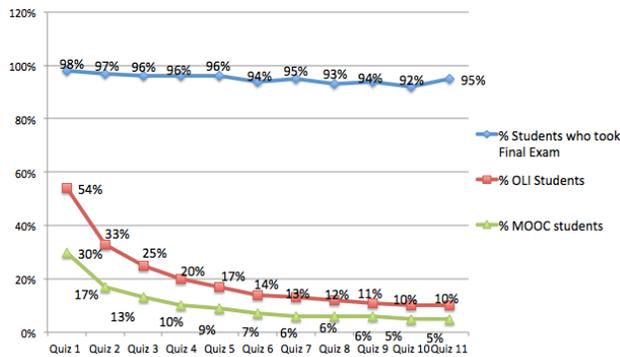


Figure 2. The participation rate in quizzes among three groups of students, registered in MOOC (N = 27720), also registered in OLI (N = 9075), and also taking the final exam (N = 939).

Figure 2 shows different participation rates among three student groups: 1) the 27720 students who registered for the course, 2) the 9075 students who registered to use the OLI materials, and 3) the 939 of those MOOC+OLI students who took the final. Quiz participation decreases over time both for all MOOC students and for the MOOC+OLI subset. The rate of quiz participation is consistently higher for MOOC+OLI students than for MOOC students in general, about twice as high. Seriousness to take the extra step to register in OLI appears *not* to be enough to explain greater MOOC+OLI participation rates. If we further restrict the sample to those who completed the first quiz (another indicator of early seriousness), we still find the MOOC+OLI students are more likely to take quiz 11 (18.5%) than the MOOC-only students (14%). The biggest drop in participation comes between quiz 1 and 2 with 43% of students dropping overall (39% of MOOC+OLI and 50% of MOOC only). Perhaps not surprisingly, the quiz participation of the MOOC+OLI students who attended the final exam (N = 939) is quite high with 98% taking quiz 1 and 95% taking quiz 11.

We also explored whether quizzes’ participation and/or quizzes’ scores can predict final exam participation. We used a logistic regression model with final exam participation as the outcome variable. The predictor variables were Pretest participation and Pretest score and 22 others for participation in each of the 11 quizzes and for scores on each of the 11 quizzes. Table 2 shows the coefficients, standard errors and P values for all predictors that are highly significant, at the $p < 0.01$ level. Participations in quizzes later in the course (i.e., quizzes 7 and 11) are, perhaps not surprisingly, good predictors of final exam participation. Additionally, there is an indication that how well students are doing in the course may also be predictive. A student’s score on quiz 1, above and beyond having merely taken it (and taken other quizzes), is associated with higher final exam participation. Given this is the first quiz, this result may suggest that students that are either underprepared for the

course or are not engaging sufficiently with the first unit materials to learn from them are unlikely to continue through to the end of the course and take the final exam.

Factors	Estimate	Std. Error	Pr(> z)
(Intercept)	-6.34020	0.19564	< 2e-16 ***
Quiz 1 Score	0.24434	0.07698	0.0015 **
Took Quiz 7 or Not	2.95023	0.71933	4.11e-05 ***
Took Quiz 11 or Not	4.95794	0.69094	7.20e-13 ***

Table 2. Significant factors in a logistic regression model predicting final exam participation.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 AIC: 1469

Comparing Learning Outcomes of MOOC-only and MOOC+OLI Students

The first measure of comparison was students’ final exam scores. MOOC-only students (N=215) had an average final exam score of 56.9% and the MOOC+OLI students (N=939) averaged 65.7%. This difference is highly significant based on a t-test ($p < .001$), however, applying a t-test directly is not appropriate because the group score distributions are not normally distributed. They are skewed toward higher scores and include a number of low outliers. We employ a simple transformation (a cubic) of the measure ($final_score^3$) to produce normal distributions. Applying a t-test to the transformed data once again yields a highly significant difference ($p < 0.001$).

As mentioned above, the difference in these self-selected groups may be a consequence of features of the students rather than of the OLI instruction. Students who registered to use the OLI materials may simply be better students. One way to test for (but not completely eliminate) the possibility of such a selection effect is to build a statistical model using all the information we have about student characteristics. We can then test whether a difference based on OLI use still remains after accounting for these other characteristics. In particular, we created a linear regression with final exam score as the outcome variable and, in addition to instructional group (MOOC+OLI vs. MOOC-only), we included six other student characteristic variables: pretest score, Quiz 1 score, occupation, age, education and gender. Because not all students answered the survey, our sample is now reduced to 551 students, 251 in MOOC+OLI and 301 in MOOC-only. Table 3 shows only the significant coefficients in the model and we see that Quiz 1 score and education make significant independent contributions to the final exam score and, importantly, the increase due to

OLI use remains. None of the other variables (pretest score, occupation, age, or gender) make a significant independent contribution to the final exam score.

	Estimate	Std Error	P-value
(Intercept)	16.90	3.32	5.21e-07***
OLI use	1.43	0.55	0.009 **
Quiz 1 score	1.06	0.11	<2e-16 ***
Education = PhD	3.96	1.84	0.032 *

Table 3. Significant factors in a linear regression predicting final exam score.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05

The model parameters indicate that students with a PhD (N=41) get 3.96 more questions (out of 40) correct on average than students in other educational groups, that every extra point on the pre-test yields 1.06 more questions correct, and that use of OLI (at least registering) produces 1.43 more questions correct. Just because students register to use OLI, does not guarantee that they do. And, similarly, just because students are in the MOOC does not mean they take advantage of the features it provides, such as watching the lecture videos. The next section investigates the log data to explore the impact of estimated student use of such features.

Variations in Course Feature Use Predict Differences in Learning Outcomes

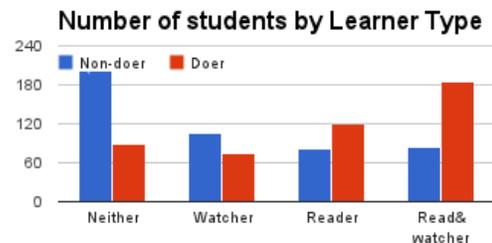
Exploratory Data Analysis: Doing, not Watching, Better Predicts Learning. As a method of exploratory data analysis, we simply performed a median split on each of our metrics of instructional feature use -- videos played, pages accessed (beyond activities), and activities started. A student is a “watcher” if they played more than a median number of videos, a “reader” if they accessed more than a median number of pages, and a “doer” if they started more than a median number of activities. Figure 3a shows that the most frequent combinations are the extremes, either being on the low half for all three, a non-watcher-non-reader-non-doer shown in the leftmost bar (N=201) or being on the high half for all three a watcher-reader-doer shown in the leftmost bar (N=184). The next most frequent combinations are reader-doers (N=120), who are using the OLI features, and watchers (N=105), who are using the distinctive MOOC feature, video lectures.

Which type of student appears to learn the most? Figure 3b shows the results for the total quiz score and indicates that the doers do well on the quizzes (score of about 94 points) even without being on the high half of reading or watching (the red bars are equally high). (Note, however,

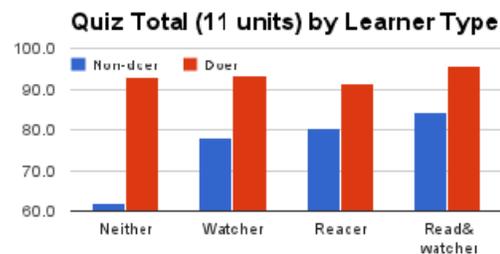
that the doers tend to read and watch more than the non-doers in the matched reading and watching groups.)

Those in the lower half of doing (the non-doers) do not perform as well on the quizzes, but those who are on *either* the high half of reading *or* watching do better (80 points) than those low on both (62 points). In other words, doing the activities may be sufficient to do well on the quizzes, but if you do not do the activities, you will be better off at least reading or watching.

a)



b)



c)

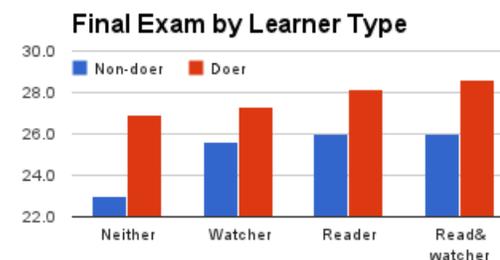


Figure 3. (a) Most students are either on the low half of watching, reading, and doing (the first blue Neither column for Non-doers) or high on all three (the last red Read&watcher column for Doers), but many (>70) are present in the other 6 combinations. (b) Doers consistently do better in the total quiz score and next best is either watching or reading. (c) Final exams appear to show greater sensitivity to reading and watching further boosting the benefits of doing, but doing is still the clear dominant factor.

Figure 3c indicates that, as for quiz score, a higher final exam score is more typical of those on the higher half of doing (about 28 points) and next best is either reading or watching (about 26 points) with low on all being the worst (23 points). In contrast with the quizzes, there is at least a hint in the final exam score showing that simply doing (26.7) is further enhanced by watching (27.4), reading (28.2), and both (28.7).

	Quiz		Final	
	Poor	Excellent	Poor	Excellent
High watcher only	46%	5%	38%	15%
High doer only	13%	42%	13%	16%
High both	2%	52%	0%	31%

Table 4. Percent of students of different extreme types (highest quartile) that do poorly or excellently on the quizzes and final.

Table 4 summarizes an analysis of the extremes that starts by splitting each of the watching, doing, quiz, and final variables into four equal groups of students and focuses on the lowest and highest quarters of these groups. The first row in Table 4 illustrates how watching lots of the videos but doing few activities leaves one quite likely to do poorly on quizzes (46% of such students) or on the final (38%) and quite unlikely to do an excellent job on quizzes (5%) and final (15%). The second row shows how simply doing a lot even without much watching of video lectures avoids poor performance on quizzes and the final (only 13% for both) and enhances excellent performance on quizzes (42%). However, a lot of doing alone does not facilitate excellent performance on the final (16%). As seen in the last column of the third row, reaching excellent levels of transfer to performance on the final is one place where it appears that extensive video watching is beneficial, but only for those who are also extensively engaged in the learning by doing activities (31% as opposed to 15% and 16% for those who are only high watchers or high doers alone).

Why might high video watching along with high doing aid excellent performance on the final, whereas lots of doing without much video watching appears sufficient for excellent performance on the quizzes? One possible reason is that the final was created by the MOOC professor and may have had items on it that are not covered in the OLI material, but only in the lecture videos. Learning by doing activities may generally better support learning, but if certain items on the final involved concepts or skills that were not required in those activities but that were presented in the videos, then students that watched lots of video lectures are more likely to get those items correct. Alternatively, we find that replacing “high watcher” with “high reader” in Table 4 produces similar results, that high doing is enough to increase chances of excellent quiz performance but it takes combining high reading along with high doing to increase chances of excellent quiz performance. In other words, it may not be the video lecture per se, but the declarative content present in video *or* reading, that appears important, on top of lots of learning by doing, for excellent performance on the final exam.

Is Course Feature Use Causally Related with Student Outcomes? As introduced in the Methods, we used Tetrad, a tool for causal inference, to evaluate whether associations between key variables, pre-test, course features (doing, reading, and watching), quiz total, and final are potentially causal. Figure 4 shows the causal model that a Tetrad search identified. The model is a good one in that a chi square (df=7) test shows its predictions are not statistically different from the data ($p = 0.39$)².

The model indicates direct causal impacts from all course features, doing activities, reading materials, and watching videos, to a higher total quiz score. The most influential impact comes from doing activities, with a normalized coefficient of 0.44 (a 1 standard deviation increase in doing activities produces 0.44 sd increase in quiz score). The strength of this relationship is more than six times the impact of watching video or reading pages (both with coefficients of about .065) and more than three times the combined impact of watching and reading.

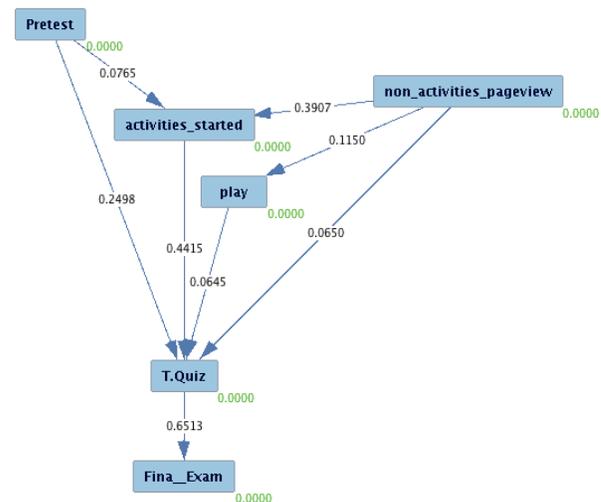


Figure 4. Tetrad inference of causal relationships between pretest score, course features (doing activities, watching videos, reading pages), total quizzes score, and final exam.

Looking at other features of the model, we see that a higher pretest score directly causes a higher quiz score (coefficient = 0.25). Thus, effects of course feature use are adjusted for and above and beyond this influence. Higher pretest also causes greater activity use, though weakly (0.08). Within the three course features, more reading causes both more doing and more watching, with

² In Tetrad, the null hypothesis of the chi square test is that the population covariance matrix is equal to the estimated covariance matrix determined as a function of the free model parameters. The test essentially asks whether the predictions for associations between variables that are not linked in the causal graph are different from the actual associations -- a model is good when the prediction is not significantly different from the data.

a larger coefficient for doing ($0.39 > 0.12$). The higher influence of reading on doing may reflect the proximity of these features within OLI pages, whereas the videos are located in the MOOC proper. Higher quiz scores cause higher final scores and this relationship is quite strong (0.65). The final exam was developed by the MOOC instructor whereas the unit quizzes more directly correspond with the content of the associated OLI modules. The strength of the connection suggests that student-learning improvements as measured by the quizzes (and highly influenced by doing more activities) do well transfer to the final exam.

RELATED WORK, LIMITATIONS AND FUTURE WORK

Our results on the limited value of video watching for learning are particularly interesting given other results suggesting that video watching behavior is predictive of dropout [30]. Those who are motivated to watch lecture videos may well be interested in the course material and may stick with it, however, that is no guarantee that those experiences produce any substantial robust or lasting learning. Consistent with our results, much work on intelligent tutoring systems is premised on the power of learning by doing. A pre-MOOC analysis of course elements found that “electronic homework as administered by CyberTutor is the only course element that contributes significantly to improvement on the final exam” [22]. A recent analysis of two MOOC data sets was less conclusive “the wide distribution of demographics and initial skill in MOOCs challenges us to isolate the habits of learning and resource use that correlate with learning for different students” [7]. One potential important difference between some learning by doing activities in those MOOCs (e.g., use of a circuit simulator) and the ones provided by OLI is the availability of fast feedback and hints in the context of doing. Such interactive (not merely active) experiences may be particularly effective for learning.

An immediate opportunity for future work is to evaluate whether our results generalize to data from a second offering of the same MOOC that was run in Spring 2014 using the same materials and teaching team. In addition, during Spring of 2013, these OLI materials were used in a variety of two-year institutions as part of the OLI Community College evaluation project [13]. An analysis of this data offers an opportunity to consider the generalizability of these results beyond the MOOC setting.

Our analysis to date has not taken advantage of all the available data. In future work, we would like to also explore how student involvement in peer-graded writing assignments and in discussion forums is associated with learning outcomes and dropout. We can also improve on our estimates of the amount of watching, reading, and doing students engage in by not only seeing how often a

video, page, or activity is started, but also estimate the amount of time they spend on each (though variation in availability of resource/activity use start and end times makes doing so harder than it may seem).

If more time were spent doing activities than reading, our results may simply be a consequence of this extra time. Preliminary analysis of the log data shows that students, on average, do more activities than read pages (387 vs. 297, respectively), but spend less overall time doing activities than reading (21.6 hrs vs. 25.0 hrs, respectively). In other words, it appears students actually spend substantially less time per activity (3.4 min) than reading a page (5.0 min). Given the improvement we see in our learning outcomes and the exploratory analysis results, this time analysis further supports the notion that doing interactive activities has a greater impact on learning than passive reading. It will be interesting to get the results of a similar analysis to compare lecture video watching time.

Although our analysis considers elements of watching, reading and doing across the entirety of the course, a more fine-grained analysis is likely desirable and it could take advantage of the fact that learning materials in OLI have been carefully mapped to specific learning objectives and skills [cf., 15].

Going beyond this specific psychology MOOC, we look to doing and seeing an expansion of this kind of analysis to other MOOCs, online or blended courses that include both passive declarative elements and interactive problem-solving elements.

CONCLUSION

While many MOOCs do include questions and some online and offline homework assignments, some have argued that a key limitation of many online courses is that they lack sufficiently rich, well-supported activities with adaptive scaffolding for learning by doing [cf., 33, 18]. Our results support the view that video lectures may add limited value for student learning and that providing more interactive activities will better enhance student learning outcomes.

ACKNOWLEDGEMENTS

We thank Dr. Anderson Smith and other members of the Georgia Institute of Technology and Coursera teams that were critical to developing and delivering the Psychology MOOC. We used the 'Psychology MOOC GT' dataset, accessed via DataShop [15] (pslcdatashop.org) at <https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=863>. Support for developing and delivering this course was provided by the Bill and Melinda Gates Foundation. Support for data storage and analytics was provided by LearnLab (NSF grant SBE-0836012).

REFERENCES

1. Aleven, V., Beal, C. R., & Graesser, A. C. (2013). Introduction to the special issue on advanced learning technologies. *Journal of Educational Psychology*, 105(4), 929–931.
2. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4, 167–207.
3. Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, 86, 124–140.
4. Barab, S.A. & Squire, K.D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Science*, 13 (1) 1-14.
5. Bichsel, J. (2013). The State of E-Learning in Higher Education: An Eye toward Growth and Increased Access (Research Report), Louisville, CO: EDUCAUSE Center for Analysis and Research. Available online: <http://www.educause.edu/ecar>.
6. Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. I. (2013). Interactive learning online at public universities: Evidence from a six-campus randomized trial. *Journal of Policy Analysis and Management*, 33(1), 94–111.
7. Champaign, J., Fredericks, C., Colvin, K., Seaton, D., Liu, A. & Pritchard, D. Correlating skill and improvement in 2 MOOCs with a student's time on task. In *Proc. Learning@Scale 2014*, Retrieved from: <http://dx.doi.org/10.1145/2556325.2566250>
8. Clark, R.E., Feldon, D., van Merriënboer, J., Yates, K., & Early, S. (2007). Cognitive task analysis. In J.M. Spector, M.D. Merrill, J.J.G. van Merriënboer, & M.P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 577–593). Mahwah, NJ: Lawrence Erlbaum Associates.
9. Collins, E.D. (2013) “SJSU Plus augmented online learning environment: Pilot project report.” The Research and Planning Group for California Community Colleges. Available online: http://www.sjsu.edu/chemistry/People/Faculty/Collins_Research_Page/AOLE%20Report%20Final%20Version_Jan%201_2014.pdf
10. Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick. *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.
11. Dewey, J. (1916), (2007 edition). *Democracy and Education*, Teddington: Echo Library.
12. Hill, P. (2013). Emerging Student Patterns in MOOCs: A (Revised) Graphical View. *e-literate*. Available online: <http://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view/>
13. Kaufman, J.; Ryan, R.; Thille, C. and Bier, N. (2013) Open Learning Initiative Courses in Community Colleges: Evidence on Use and Effectiveness. Mellon University, Pittsburgh, PA. Available online: http://www.hewlett.org/sites/default/files/CCOLI_Report_Final_1.pdf
14. Koedinger, K.R., & Aleven V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review*, 19(3), 239–264.
15. Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. (2010). A Data Repository for the EDM community: The PSLC DataShop. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) *Handbook of Educational Data Mining*. Boca Raton, FL: CRC Press.
16. Koedinger, K.R., Booth, J. L., & Klahr, D. (2013). Instructional Complexity and the Science to Constrain It. *Science*, 342(6161), 935-937.
17. Koedinger, K. R., Corbett, A. C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36 (5), 757-798.
18. Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2014). Ubiquity symposium: MOOCs and technology to advance learning and learning research: Data-driven learner modeling to understand and improve online learning. *Ubiquity*, Volume 2014, Number May (2014), Pages 1-13. ACM New York, NY, USA. DOI:10.1145/2591682
19. Koedinger, K. R., & Sueker, E. L. F. (2014). Monitored design of an effective learning environment for algebraic problem solving. Technical report CMU-HCII-14-102.
20. Lovett, M., Meyer, O., Thille, C. (2008). The Open Learning Initiative: Measuring the effectiveness of the OLI learning course in accelerating student learning. *Journal of Interactive Media in Education*. <http://jime.open.ac.uk/2008/14/>.
21. Meyer, O., & Lovett, M. C. (2002). Implementing a computerized tutor in a statistical reasoning Course: Getting the big picture. In B. Phillips (Ed.) *Proc. of the Sixth International Conference on Teaching Statistics*.
22. Morote, E. & Pritchard, D. E. (2002). What Course Elements Correlate with Improvement on Tests in Introductory Newtonian Mechanics? National

- Association for Research in Science Teaching – NARST- 2002 Conference.
23. Ng, A. and Widom J. (2014). Origins of the Modern MOOC.
<http://www.cs.stanford.edu/people/ang/papers/mooc14-OriginsOfModernMOOC.pdf>
 24. Pane, J.F., Griffin, B., McCaffrey, D.F. & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at Scale. *Educational Evaluation and Policy Analysis*, 36 (2), 127 - 144.
 25. Pappano, L. (2012, November 2). *The year of the MOOC*. The New York Times.
 26. Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning (NCER 2007-2004)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
 27. Ritter S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14 (2):249-255.
 28. Schank, R. C., Berman, T. R. & Macperson, K. A. (1999). Learning by doing. In C. M. Reigeluth (Ed.), *Instructional Design Theories and Models: A New Paradigm of Instructional Theory (Vol. II)* (pp.161-181). Mahwah, NJ: Lawrence Erlbaum Associates.
 29. Singer, S.R. & Bonvillian, W.B. (2013). Two Revolutions in Learning. *Science* 22, Vol. 339 no. 6126, p.1359.
 30. 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. *Proc. of the 2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*
 31. Strader, R. & Thille, C. (2012). The Open Learning Initiative: Enacting Instruction Online. In Oblinger, D.G. (Ed.) *Game Changers: Education and Information Technologies (201-213)*. Educause.
 32. Tetrad IV. <http://www.phil.cmu.edu/tetrad>
 33. Thille, C. (2014). Ubiquity symposium: MOOCs and technology to advance learning and learning research: opening statement. *Ubiquity*, Volume 2014, Number April (2014), Pages 1-7. ACM New York, NY, USA. DOI: 10.1145/2601337
 34. VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227-265.
 35. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
 36. Zhu X., Lee Y., Simon H.A., & Zhu, D. (1996). Cue recognition and cue elaboration in learning from examples. In *Proc. of the National Academy of Sciences* 93, (pp. 1346±1351).