

Brent J. Evans  
Rachel B. Baker  
JE Thomas S. Dee

---

**Online Appendix for:**

## **Persistence Patterns in Massive Open Online Courses (MOOCs)**

### **Selection of Words for Lecture Title Analysis**

We were led to this title word analysis graphically, by noting lecture outliers—some videos were watched multiple times by individual students, while others were watched by very few students. Preliminary analyses indicated that these outliers were potentially explained by the lecture titles—some indicated that they contained information on assignments and exercises, others were labeled optional. This led us to the conclusion that students were sensitive to word choice in lecture titles. We created the list of words to check by scraping all lecture titles from our data set and parsing these strings on spaces. We counted how many instances of each word we had and considered all words that appeared more than 20 times across all lectures. We eliminated words that were not meaningful (i.e. prepositions, conjunctions, or pronouns—“and,” “the,” “of”), were content-specific (e.g. “algorithm,” “encryption”), or did not convey information about the content of the video (e.g. “week,” “lecture,” “video,” “minute”). We also eliminated words that only appeared in lectures from one course (they would be dropped from the analyses due to our course fixed effects). We included the remaining words and then added to that list words similar in meaning (e.g. “basic,” “conclusion”) which may not have appeared 20 times. Our choice of words was thus empirically driven but also somewhat subjective. We also tried grouping these sets of words using factor analysis, but the data are more clearly presented listing words separately.

### **Registration & Participation**

Table A1 provides descriptive statistics across the 44 MOOCs we analyzed in this study. The table includes information about each course: its date of release, the number of students who registered for the course, total number of lecture videos for the course, and seven measures of student engagement and persistence: the percent of registrants who

watched at least one, 20%, and 80% of course videos; the percent of registrants who earned a certificate; and the percent of registrants earning a certificate conditional on watching at least 1, 20%, and 80% of videos. Analyzing this diverse array of outcomes illuminates critically important distinctions between the types of students engaged in MOOCs.

On average, over 48,000 students registered for each class, but there was substantial heterogeneity across courses. The largest course by registration numbers was Game Theory with over 128,000 students, and the smallest was the third offering of Compilers with just over 11,000 students. The huge number of registrants across classes reflected at least two factors: clear student interest and the ease of entry. The barrier to registering for a course could not be lower. After creating a Coursera account (which itself only requires an email address), a prospective student only needed to click one button on the website to register for any course. This could be done months before the course actually began in some cases.

In part because the registration process was so easy, many students who registered did not watch a single lecture video. In 16 of the 44 courses, the *majority* of registrants never watched a single video, and on average across all 44 courses, 54.4% of registered students participated by watching at least one lecture video. These low engagement rates suggest that students may have perused the Coursera catalog and registered for courses that seemed interesting with little intention of fully engaging in many of them. Because the registration costs were so low, many students may have been willing to incur the time cost of registering but were not willing to spend the time to actually watch a lecture video. Another explanation is that students were simply gathering information about the course. Indeed, registering for a course gave students the ability to see more information about a course, so “registering” might be more accurately referred to as “information gathering.” The analogous situation in a traditional higher education course would be that just above half of the students that expressed interest in a course (by registering or some lower cost way of revealing interest) ever showed up for a lecture. We rarely consider this number in traditional higher education because we typically allow students to add and drop courses after the term begins and do not track low cost student information gathering (such as reading information online and asking friends).

Additionally, many students who began the course stopped participating before the course ended. The measures of student persistence presented in Table A1 give a sense of the level of dropout behavior. The percent of students who watched at least one fifth of the videos in each course varies with higher levels in Child Nutrition (51%) and low levels in the second and third offerings of Logic (less than 10%). The

unweighted average across all courses is 21.6% of registrants watched more than one fifth of the lecture videos. When compared with the percent of students watching at least one video, these numbers indicate many students watch one or two videos and then disengage. This lends support to the theory that students have certain expectations about the course that may not be met in the first few lectures resulting in their dropping the course. These findings are also consistent with the notion of the MOOC being an experience in which the MOOC is of unknown quality, so students register, watch one or more lecture to determine quality and decide to stop participating.

The percent of students who watched at least 80% of the videos in the course was used to proxy for substantial engagement, and thus persistence, in the course. These percentages were, on average, less than half of the percent of students who watched at least one fifth of the lectures indicating that a large subset of students engaged with the course initially and then dropped out or only engaged with the course sporadically throughout. Across all courses, only 10% of registrants watched more than 80% of the lecture videos.

Earning a certificate, while not a goal of many MOOC participants, can serve as a measure of course completion. Certificate bearing courses (four did not offer certificates) awarded certificates to only 5% of registrants on average, but this statistic masks large heterogeneity across courses. Nearly 25% of registrants in Child Nutrition earned a certificate but only about 1% of registrants in Sustainable Agriculture and the second offering of Compilers did.

Considering the lack of initial engagement upon registration, using all registrants as the denominator to provide a completion rate is very misleading. Therefore, we propose establishing a certificate rate by tallying the percent of certificate earners among engaged students. The final three columns of Table A1 show the percent of registrants that completed the certificate conditional on actively engaging with or persisting in the course by watching at least 1, more than 20% and more than 80% of the lectures. These measures increased completion rates dramatically. Even conditioning on merely watching one video raises the average completion rate from 5 to 8.5%. Conditioning on students watching more than 20% of the videos increased the base rate by a factor of four up to 20%. The analogous case in traditional higher education might be to wait to assess course completion rates until after the drop deadline.

Among students who persist by watching more than four-fifths of the videos, the average completion rate is one-third. This suggests there was an extremely high number of auditors and/or low performing students who did not complete the certificate. These final course grades suggest a

4 *The Journal of Higher Education*

substantial number of students engaged only with the lecture videos and would therefore be classified as auditors in traditional higher education language.