Beyond Time-on-Task: The Relationship Between Spaced Study and Certification in MOOCs

Yohsuke R. Miyamoto

Harvard University, USA

yrm@seas.harvard.edu

Cody A. Coleman

Massachusetts Institute of Technology, USA

Joseph Jay Williams

Harvard University, USA

Jacob Whitehill

Harvard University, USA

Sergiy Nesterko

Harvard University, USA

Justin Reich

Harvard University, USA

Abstract: A long history of laboratory and field experiments has demonstrated that dividing study time into many sessions is often superior to massing study time into few sessions, a phenomenon widely known as the "spacing effect." Massive open online courses (MOOCs) collect abundant data about student activity over time, but little of its early research has used learning theory to interrogate these data. Taking inspiration from this psychology literature, here we use data collected from MOOCs to identify observational evidence for the benefits of spaced practice in educational settings. We investigated tracking logs from 20 HarvardX courses to examine whether there was any relationship between how students allocated their participation and what performance they achieved. While controlling for the effect of total time on-site, we show that the number of sessions students initiate is an important predictor of certification rate, across students in all courses. Furthermore, we demonstrate that when students spend similar amounts of time in multiple courses, they perform better in courses where that time is distributed among more sessions, suggesting the benefit of spaced practice independently of student characteristics. We conclude by proposing interventions to guide students' study schedules and for leveraging such an effect.

Keywords: MOOC, spacing effect, spaced practice, distributed practice, time on-site, sessions, online education

1. INTRODUCTION

1.1 Early MOOC Research on Participation and Time-on-Task

Much of the early research in Massive Open Online Courses (MOOCs) has focused on measures of student participation, such as time spent on-site, number of click events produced, minutes of video watched, or number of assignments completed. These studies have repeatedly observed two commonplace findings: 1) that the level of participation along one dimension of a course is a predictor of participation along other dimensions, and 2) that the level of participation is a predictor of better grades and course completion (Collins, 2013; Murphy, Gallagher, Krumm, Mislevy, & Hafter, 2014; Reich et al., 2014; Wilkowski, Deutsch, & Russell, 2014).

These results align with the total-time law in psychology that the amount of learning is a direct function of study time (Cooper & Pantle, 1967; Underwood, 1970). Psychologists, however, have long known that not all uses of study time produce equal benefits. One striking exception to the total-time law is known as the *spacing effect:* for most learning outcomes, shorter, more spaced out study sessions are preferable to massed study sessions. Here we investigate the contribution of the spacing effect to student performance in MOOCs, going beyond existing MOOC research by focusing not only on levels of participation, but also on the *allocation* of student study time into multiple sessions and its potential benefits for student performance. The decades of research in experimental psychology, education, and cognitive science should help in guiding practical research (Williams, 2013); we demonstrate that well-established learning theories can produce important hypotheses for learning analytics, and that learning theory can guide the discovery of effective uses of student learning time.

1.2 What is the Spacing Effect?

The spacing effect, initially documented by Herman Ebbinghaus in 1885, is the remarkable phenomenon where distributed presentations of material result in better long-term retention than that resulting from massed presentations of the same material, for a given amount of study time (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1989; Ebbinghaus, 1885; Greene, 1989; Melton, 1970). For example, the long-term retention of some element A is much stronger following a distributed practice session such as [X, A, Y, Z, A, W, A], rather than a massed practice session such as [X, Y, A, A, A, Z, W], even though the total time spent on studying A is the same. The dependability and robustness of this effect is demonstrated by its replication in a wide variety of experimental tasks over numerous studies (Melton, 1970; Moulton et al., 2006; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963; Shea, Lai, Black, & Park, 2000; Underwood, 1970; Waugh, 1970; Young, 1967). Moreover, the quantitative advantage earned from a spaced schedule is remarkable, with reports that two spaced presentations are about twice as effective as two massed presentations (Hintzman, 1974; Melton, 1970), and the difference between them increasing further as a function of the number of presentations (Underwood, 1970).

1.3 From Experiments to Education

Despite such potential for bolstering learning, the spacing effect has suffered a history of slow translation into standard educational practices, despite knowledge of the effect in its community for more than a century (Dempster, 1988). In an effort to mitigate this issue, studies have appeared in recent years of real-world classroom demonstrations of the spacing effect in many contexts, such as learning vocabulary (Bird, 2010; Carpenter, Pashler, & Cepeda, 2009; Gallo & Odu, 2009; Rohrer & Taylor, 2006; Sobel, Cepeda, & Kapler, 2011), and surgical motor tasks (Moulton et al., 2006). Furthermore, technology has enabled what is perhaps the most visible application of the spacing effect today – spaced repetition software (SPS), sophisticated systems of digital flashcards that help sequence the presentation of study material in a spaced manner (Godwin-Jones, 2010).

Our study extends this research on the applications of spaced practice into the realm of MOOCs, and demonstrates new opportunities for research made possible by the massive yet granular data sets of student activity collected by online learning systems. Much of the research on spaced practice takes advantage of experimental designs that can demonstrate causal relationships and internal validity in very specific circumstances. Complementary to this work, our data allows us to observe large numbers of students who study diverse course content and spend their time across multiple courses in diverse ways. While we cannot draw causal conclusions from our data, we can investigate the effects of spaced practice "in the wild," beyond settings where experimental psychologists control critical elements of course design or student time. Moreover, studies of spaced practice often demonstrate their effects on specific measures of memory and retention. We demonstrate the effects of spaced practice on more generally applicable measures of course completion.

We begin our research with the hypothesis that, broadly speaking, among students who spend similar amounts of time on a MOOC, those who distribute their time into more sessions will perform better than those with fewer sessions. We test this hypothesis with two research designs. First, we examine all students in all courses and show that among students with similar total-time within a course, those with higher session counts perform better.

This correlational study, however, leaves open the possibility that students who spaced their practice differ from students who clustered their practice in other important but unobservable ways. Therefore, in the second level of our investigation, we control for student-to-student differences by examining changes in performance within-individuals, focusing on students who spend similar amounts of time across multiple different courses. We find evidence even within individual student behaviour that spaced practice is associated with better performance.

These findings suggest a simple way of improving student learning, agnostic to the nature of specific activities and course content, by providing interventions to encourage students to complete the course with more sessions for the amount of total study time they have available. In the sections that follow we present our data, methods, and findings, and suggest a set of possible experimental interventions to encourage spaced practice.

2. DATA

2.1 Data Collection

In order to explore the relationship between student performance and how students distribute their time in MOOCs, we examined the timing of click events recorded in the tracking logs of 101,913 unique students (with a combined 127,868 course registrations) with non-zero grades in 20 HarvardX courses (Table 1), and how they relate to students' certification rates (rate of achieving a passing grade) in these courses.

HarvardX is an online learning initiative of Harvard University, and provides the 20 courses examined here on the edX MOOC platform. The different types of online click events triggered by students range from page navigation, to lecture video plays, to problem submissions, as shown in Fig. 1b, and the diverse set of topics of these courses ranged from biostatistics (Health in Numbers), to philosophy (Justice). Some courses were offered on multiple occasions, such as The Ancient Greek Hero, and Justice, while many others were offered on a single occasion. ChinaX is a series of separate courses with an ongoing theme that continued for 6 modules, and thus tends to have overlapping, devoted students, resulting in high certification rates. Certification rates from its later modules also tend to be high because of attrition from earlier modules. Since we use certification as the outcome of interest in subsequent analysis, we restrict our investigation to students who answer at least one problem correctly, to exclude casual browsers and auditors (Ho et al., 2014).

Course title	Course code	Start date	End date	# students with nonzero grade	# certified students
The Ancient Greek Hero	CB22x	3/13/13	8/26/13	4663	1395 (30%)
The Ancient Greek Hero	CB22.1x	9/3/13	12/31/13	2625	727 (27%)
Justice	ER22x	3/2/13	7/26/13	11896	5265 (44%)
Justice	ER22.1x	4/8/14	7/17/14	6307	2483 (39%)
Unlocking the Immunity to Change	GSE1x	3/11/14	6/30/14	16641	1854 (11%)
Leaders of Learning	GSE2x	7/8/14	8/25/14	11191	3933 (35%)
Fundamentals of Clinical Trials	HSPH	10/14/13	2/14/14	6422	2406 (37%)
Health in Numbers: Quantitative Methods in Clinical & Public Health Research	PH207x	10/15/12	1/30/13	16541	4910 (30%)
United States Health Policy	PH210x	4/7/14	6/30/14	3530	759 (25%)
Human Health and Global Environmental Change	PH278x	5/15/13	7/25/13	6544	2711 (41%)
Data Analysis for Genomics	PH525x	4/7/14	6/30/14	4501	621 (14%)
Science and Cooking: From Haute Cuisine to Soft Matter Science	SPU27x	10/8/13	3/15/14	10274	1794 (17%)
The Political and Intellectual Foundations of China (ChinaX)	SW12x	10/31/13	12/23/13	8216	2016 (25%)
The Creation and End of a Centralized Empire (ChinaX)	SW12.2x	1/2/14	1/30/14	3624	1751 (48%)
Cosmopolitan Tang: Aristocratic Culture (ChinaX)	SW12.3x	2/13/14	3/6/14	2509	1565 (62%)
A New National	SW12.4x	3/20/14	4/10/14	2140	1230 (57%)

Culture (ChinaX)					
From Global Empire to Global Economy (ChinaX)	SW12.5x	4/24/14	5/8/14	1607	1117 (70%)
The Last Empire (ChinaX)	SW12.6x	5/22/14	6/19/14	1845	1108 (60%)
Global Health: Case Studies from a Biosocial Perspective	SW25x	2/25/14	5/31/14	3204	1266 (40%)
Tangible Things	USW30x	6/2/14	8/2/14	3588	1089 (30%)

Table 1: Course Information for Twenty 2012-2014 HarvardX Courses

2.2 Measures

We use certification in a course as our outcome of interest. Certification was awarded to students who achieved a minimum level of performance on a combination of quiz scores, homework assignments, project results, and so on, with content and a threshold that varied from course to course. We use certification as a proxy for learning, but since courses differ in the type and rigor of assessments, the specific meaning of certification varies. It is important to note that most studies of spacing effects focus on measures of memory and retention, rather than this overall measure of course performance.

We extract two measures from course tracking logs to predict certification: (1) total time on-site, and (2) the number of sessions among which their time was distributed, inspired by the spacing effect documented in the field of psychology. As shown in Fig. 1a, a session was defined as a collection of click events separated by periods of inactivity that lasted more than 30 minutes, in accordance with the Google Analytics standard for defining sessions for website usage[1]. Total time spent on-site was thus calculated by summing the lengths of all sessions by a student. A few example sessions shown in Fig. 1b, taken from CB22x and SW12x, demonstrate the patterns of click events that might occur during a session.

[1] https://support.google.com/analytics/answer/2731565?hl=en

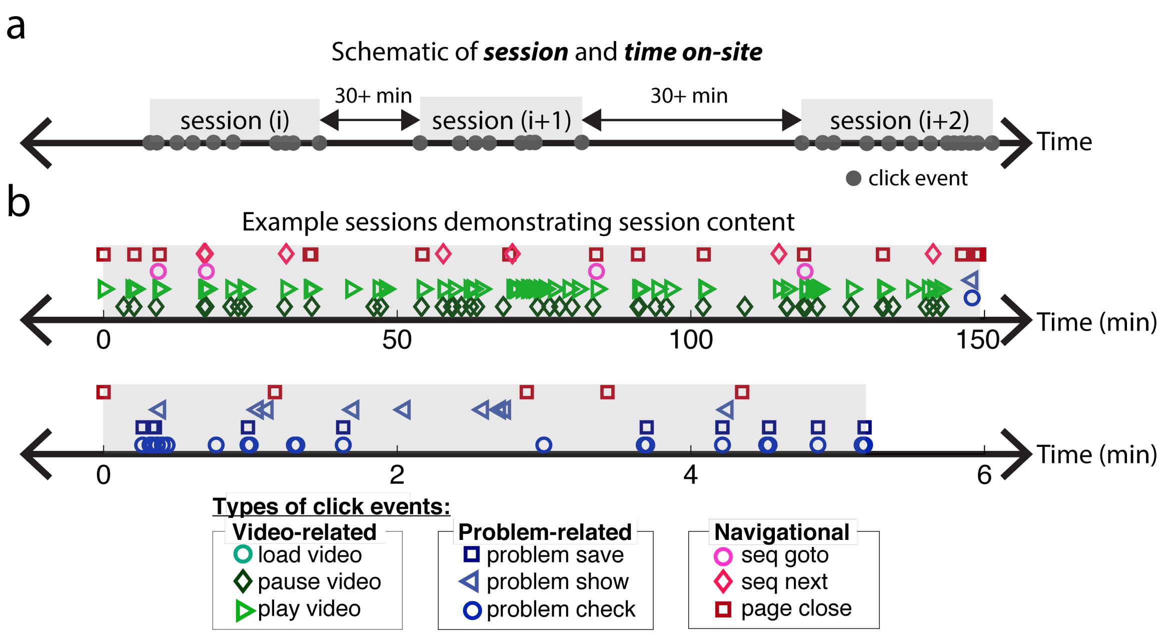

Fig. 1: Illustrations of session and time on-site. (a) Schematic of the definition of session and time on-site: A session was defined as a collection of click events separated by periods of inactivity that lasted more than 30 minutes, in accordance with the Google Analytics standard for defining sessions of website usage[2]. Total time spent on-site was calculated as the sum of the lengths of all sessions by a student. (b) Example sessions taken from the data, which demonstrate some variety and combinations of event types initiated by students. The top example demonstrates a session focused on video usage, interleaved with page navigation events. The bottom example demonstrates a session that focuses on quiz problems, with occasional page navigation. Note that the set of event types shown in these examples (9 event types) does not include all possible event types. (A comprehensive list and description of event types can be found at the edX research guide[3]).

Our metric of student participation, total time on-site, includes the time spent on all activities on-site regardless of whether it was spent watching lecture video clips, solving problem sets, readings chapters, taking quizzes, and so on. Noting that the patterns and frequencies of these event types can vary widely from course-to-course, we took this simplified approach of treating all types of activity equally in the form of total time on-site to keep our analyses agnostic to the wide course-to-course differences in content, structure, and certification requirements. This may allow potential interventions inspired by our findings to be flexibly applied to a wide variety of courses.

[2] https://support.google.com/analytics/answer/2731565?hl=en

[3] http://edx.readthedocs.org/projects/devdata/en/latest/internal_data_formats/event_list.html#event-list).

3. RESULTS

3.1 Exploratory Analysis of Key Features in MOOC Data

With these data, we embarked on an exploratory analysis of the relationships between certification rate, time on-site, and number of sessions, inspired by the spacing effect from psychology, where spaced presentation of material benefits long-term recall. An initial look at the data revealed widely varying levels of certification rate, time on-site, session count from course to course. The certification rates of students with nonzero grades in each course vary widely, from 11% to 70% (Table 1). In Figures 2a and 2b, we plot histograms of the number of students by log-scale session count and total course hours, respectively. These histograms reveal that time on-site and session count of courses vary widely, with the median time on-site level of courses varying from as little as 2 hours (USW30x) to about 17 hours (PH207x), an over 8-fold difference, and median sessions from 5 sessions (SW12.5x) to 26 sessions (PH207x), an over 5-fold difference. Since these courses differed greatly in these descriptive statistics and their substantive structural features, we chose to analyze each course individually so as to prevent misinterpreting course-to-course differences in time on-site, session count, and certification as effects of individual students.

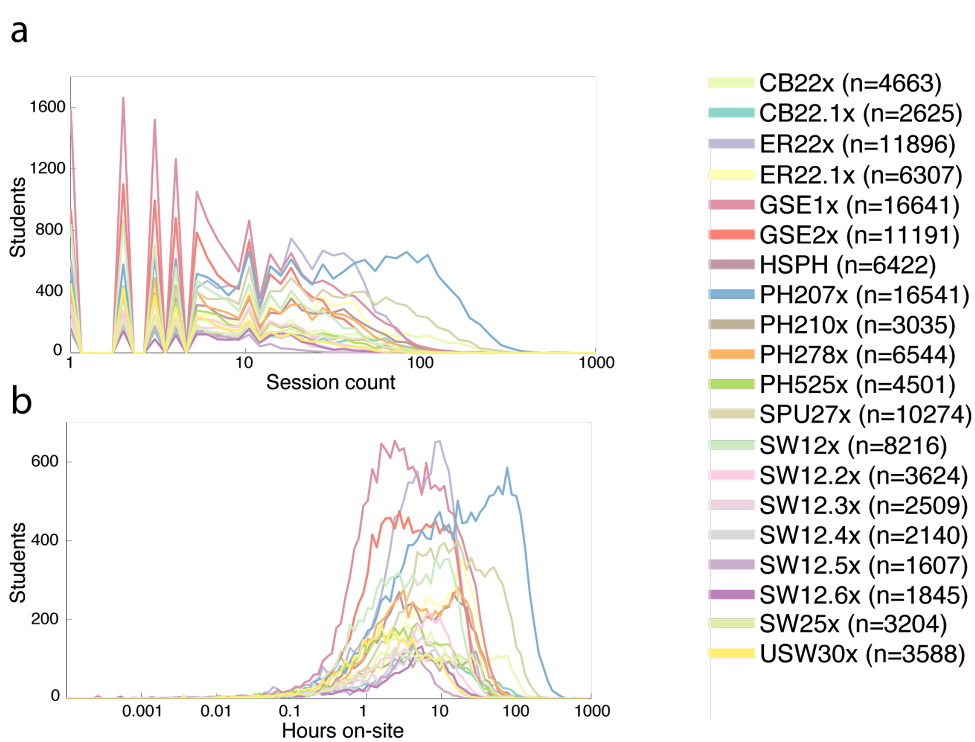

Fig. 2: Exploratory analysis of time on-site, spacing, and certification rates in MOOCs. (a) Histograms of students' hours on-site levels are shown, with each coloured line representing students from a

different course, for a total of 20 colours representing 20 courses, as detailed in the legend to the right. Note the log-scale x-axis. (b) Histograms of students' spacing levels are shown for each of the 20 courses. Note that the spikes at the low spacing levels are due to the fact that the number of sessions can only take integer values.

A further look at the histograms of Figures 2a and 2b also reveals widely varying levels of engagement and number of sessions across students within each course – for example, total time on-site of individual students across all courses varied from as little as a few minutes to as much as over 100 hours, and session counts from 1 to over 100. We wondered if this rich variability of behaviour across individual students could explain differences in their course performance. Are users with high session count or time on-site more likely to achieve certification?

Thus, we next plotted total time on-site and session count against certification rate. In Figure 3a, we divided students within each course into deciles based on their session counts from low to high, and computed each decile's certification rate to obtain each of the coloured lines in Fig. 3a, with colours representing separate courses. The bold black line corresponds to the mean relationship between session count and certification rate across courses, which was obtained by averaging the session count and certification rate of each the deciles across the 20 courses. These relationships appear sigmoidal, with small effects on certification at small session counts less than 10 sessions, followed by a sharp increase and later a gradual plateauing as the certification rates approach its maximum of 1 (note the log-scale of the x-axis).

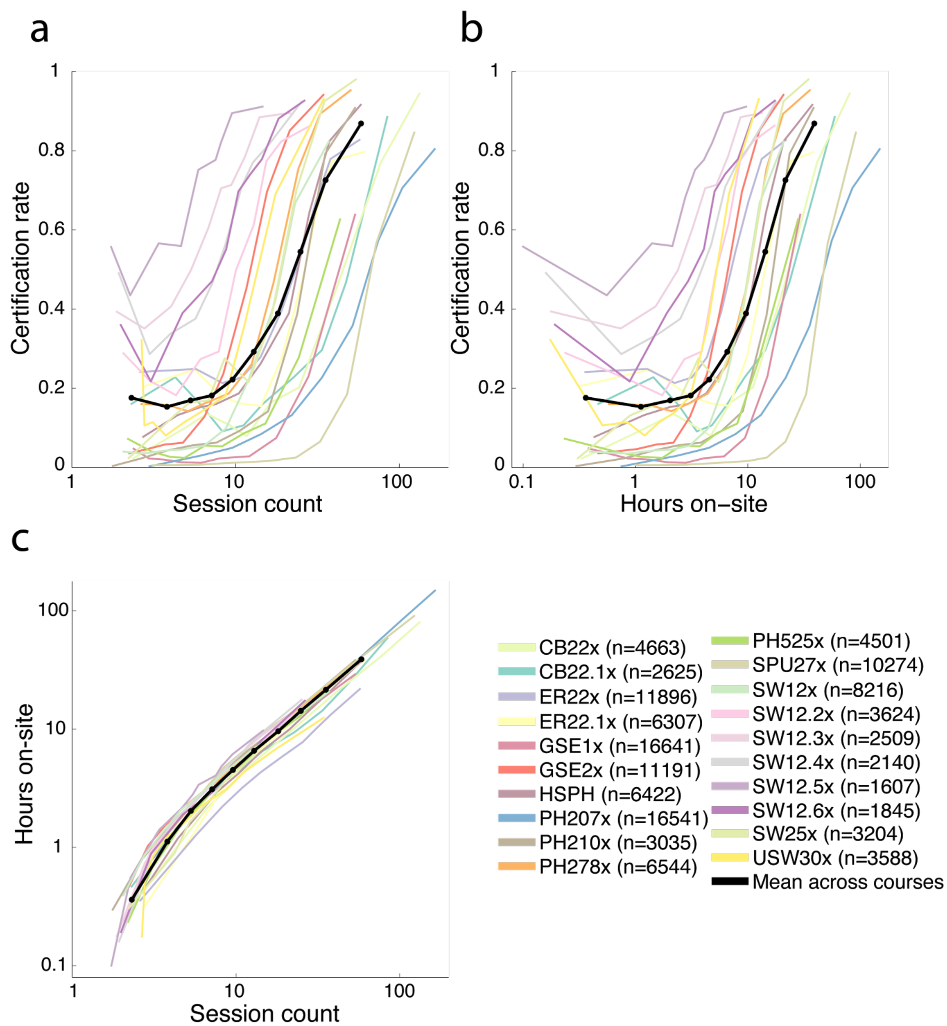

Fig. 3: Exploratory analysis of time on-site, spacing, and certification rates in MOOCs. (a) Certification rates are plotted as a function of session count (10 levels). Students from each course were divided equally into deciles of session counts, and the average certification rates for each decile is plotted as a function of session count for that decile, to form the coloured lines, one colour for each course. The mean certification rates and session counts were then averaged across all 20 courses for each decile to obtain the bold black line, representing the mean relationship between certification rate and session count across courses. (b) Certification rates are plotted as a function of time on-site, after dividing students into deciles of time on-site in an analogous fashion to session count in (a). (c) Time on-site is plotted as a function of session count after dividing students into deciles of session counts, as in (a).

When we also analyzed students' time on-site versus certification in the same manner as with session count (Fig. 3b), here too we found a strong positive correlation with certification rate, consistent with expectations from the MOOC literature (Collins, 2013; Murphy et al., 2014; Reich et al., 2014; Wilkowski et al., 2014). This relationship appeared to be very similar to that of session count from Fig. 3a, where certification rates increased in a sigmoidal fashion also with respect to time on-site. It is also perhaps interesting to note the dip that occurs in certification rate in some courses from the lowest decile to the next, for both session count and time on-site; this may reflect the behaviour of some students using a secondary account to earn a certificate in a single session, after initial exposure to test questions on a primary account (a behaviour that in certain contexts might be described as cheating).

Our analyses thus far of Fig. 3a and 3b suggest a positive relationship of certification, our proxy for learning, with session count and with time on-site. However, Fig. 3c, which plots total time versus session count, depicts one challenge of further interrogating these relationships. We found that time on-site and session count themselves were highly correlated with each other (Fig. 3c), with a correlation coefficient of 0.87±0.04 (mean±s.d. across courses), indicating, as one might expect, that students who work for many sessions also tend to spend a lot of time doing so overall. This multicollinearity of our features makes it difficult to determine whether students' certification rates are driven by time on-site, by session count, or both. It reflects a challenge of analyzing real-life, messy data sets of learning "in the wild," in that we lack experimental control over variables of interest, making it difficult to ascertain the effect of one variable versus another.

3.2 Identifying a Benefit of Spacing via Student-to-Student Comparisons

Our data set enables us to reduce ambiguity in interpreting these correlations despite issues of multicollinearity, through its scale of over 100,000 students distributed over 20 courses. Because there are so many students in our data set, every student can be matched with other students who exhibit similar levels of time on-site but different session count, allowing us to compare the effect of session count on students while controlling for time on-site. By looking at the changes in session count while controlling for time on-site, we are able to measure the extent to which the same amount of time on-site is "spaced" across distinct sessions, corresponding closely to the idea of spaced practice in the psychology literature. Accordingly, we divided students at each of the deciles of time on-site corresponding to the location of the black dots of Fig. 3b, into low, mid, and high-spacing (session count) terciles, and compared their certification rates (Fig. 4a). Figure 4a provides compelling visual evidence for the benefit of the spacing in MOOCs, showing that the high-spacing subgroup (black) at each level of time on-site consistently exhibited higher certification rates than their corresponding low and mid-spacing subgroup counterparts at each decile of time on-site (light grey and dark grey dots). The strength of this effect seems to be largest at low levels of time on-site (light blue region) rather than for high level of time on-site (light red region).

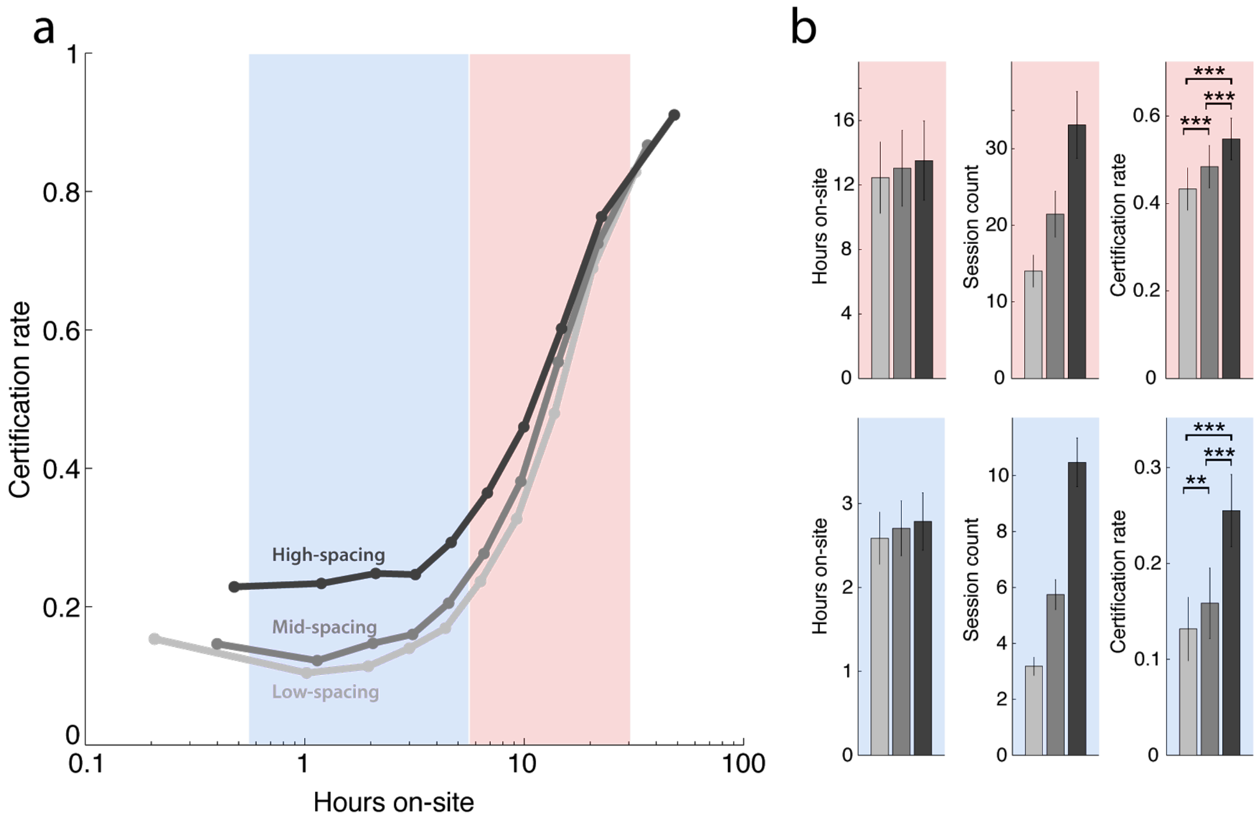

Fig. 4: Across-student analysis of the spacing effect in MOOCs. (a) Certification rates are plotted as a function of time on-site (10 levels) and spacing (3 levels) to form 10 x 3 = 30 total groups of students taken from the 20 HarvardX courses in the analysis. First, students within each course were divided equally into deciles of time on-site, and then students of each decile were divided further into terciles of low, mid, and high spacing groups (light grey, grey, and dark grey dots). The certification rates and time on-site levels of the 30 groups were then averaged across the 20 courses to form what is shown here. (b) The top bar graphs in light red show the average time on-site, session count, and certification rate aggregated across high levels of time on-site corresponding to the light red region in (a), for low (light grey), mid (grey) and high (dark grey) spacing groups. The bottom bar graphs in light blue are analogous plots representing the time on-site, session count, and certification rates aggregated across the low levels of time on-site corresponding to the light blue region of (a). Note that the lowest (leftmost) and highest (rightmost) deciles of time on-site were not included in the calculation of these bar graphs for a fairer comparison since the levels of time on-site among these spacing groups were not well matched; however, the trend remains within these excluded spacing groups where higher certification is generally associated with higher spacing. Error bars indicate standard errors of the mean, across the 20 courses. The large error bars reflect the large variability across courses noted in Fig. 2, but the statistical comparisons are highly significant because comparisons are evaluated in a paired fashion within-course. **: p < 0.01, *** : p < 0.001

The bar graphs in Figure 4b help to clarify the differences in the relationship between spacing and certification at different levels of time on site. The top three bar graphs summarize the total time, session number, and certification rate for high levels of time on-site, corresponding to the red area in

Figure 4a. The light, mid and dark grey bars correspond to the average of the low, mid, and high-spacing terciles in this red region. By design, these low, mid, and high-spacing terciles have similar levels of time on-site (Fig. 4b, top-left) and highly contrasting session counts (Fig. 4b, top-middle). The top-rightmost bar graph shows that certification rates were significantly higher for the high-spacing than for the mid-spacing group (Fig. 4b, top right, paired t-test, $p < 0.001$), and both of these groups had rates significantly higher than that of the low spacing group (Fig. 4b, top right, paired t-test, $p < 0.001$).

We show analogous results for the low levels of time on-site, shown in light blue (Fig. 3b, bottom). At these low levels of time on-site, the low, mid, and high spacing groups for the low levels of time on-site (light blue region, Fig. 4a) had similar average time on-site (Fig. 4b, bottom left) and highly contrasting average session counts, by construction (Fig. 4b, bottom middle). As in high time on-site levels, these terciles exhibited average certification rates that were significantly higher in the high-spacing groups (Fig. 4b, bottom right $p < 0.001$) and significantly lower in the low-spacing groups ($p < 0.01$) as compared to mid-spacing groups. Note that the lowest (leftmost) and highest (rightmost) deciles of time on-site were not included in the calculation of these bar graphs to allow for a fairer comparison since the levels of time on-site among these spacing groups were not well matched, as one can observe in Fig. 4a; however, the trend remains within these excluded deciles where higher certification is generally associated with higher spacing.

Thus, by comparing the behaviour of students with one another within each course, we found that students who distribute their time into a larger number of sessions tend to have higher certification rates than those who do not, while accounting for time on-site. Interestingly, we found that this effect was larger for students at low levels of time on-site, perhaps due to a saturation effect at higher levels of time on-site and certification, suggesting that the potential benefit of spacing may be most prominent at lower levels of participation time.

3.3 Identifying a Benefit of Spacing via Within-Student Comparisons

Thus far, we found that students who distribute their time into a larger number of sessions had higher levels of certification than students who spent similar total time on-site divided into fewer sessions. This effect was particularly prominent within low levels of time on-site. However, these correlational observations could be confounded by unobserved variables: it could simply be that high-performing students also tend to have large numbers of sessions, but the *act* of distributing time into a larger number of sessions may not be a causal factor in better performance. To investigate this question, we analyzed our data more finely to examine whether the effect of sessions can be observed within individuals. We investigated whether students with different levels of session counts across different courses show higher certification rates in courses where they space out their time more. Here we take advantage of our data set of over 100,000 total students to examine the subset of students enroled in multiple courses. Looking within the overlapping students of a particular pair of courses, some may have distributed their time into more sessions in one than the other, while maintaining their level of time on-site. Although this may happen rarely, the massive scale of the data set gives us the ability to observe a sizeable number of such events. Thus we can examine if a student's change in session count is associated with a corresponding change in performance that resembles the spacing effect.

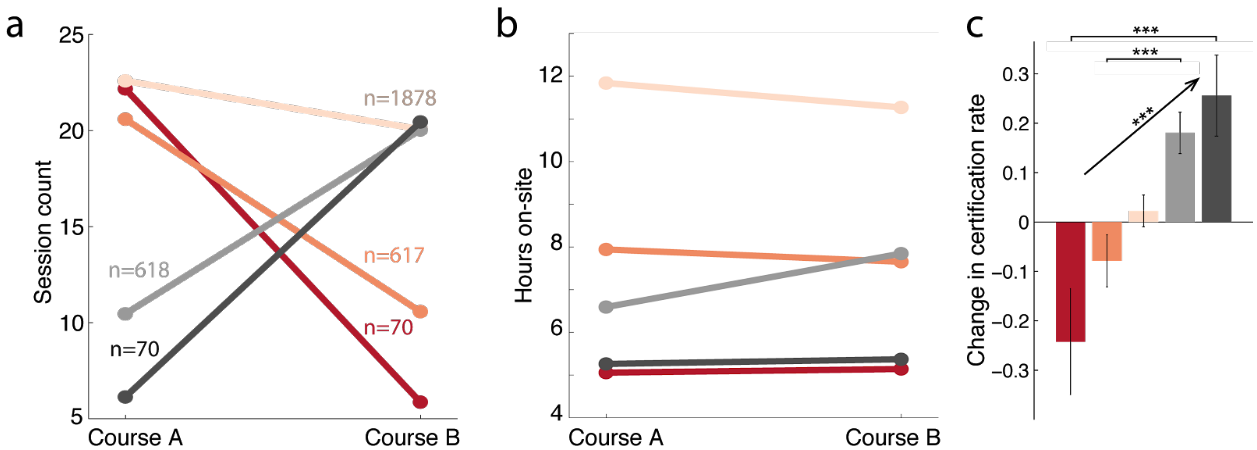

Fig. 5: Within-student analysis of the spacing effect in MOOCs. (a) Session counts are shown for 5 different variations of changes from one course to the other within a course-pair. Dark grey represents students who greatly increased their session count from low to high terciles between courses. Terciles were computed within each course, to account for course-to-course differences in session counts. Light grey represents those who mildly increased their session count from low to mid, or mid to high terciles between the pair of courses. Red and dark orange represent students who greatly and mildly decreased their session count, respectively, in an analogous manner. Light orange represents those who remained within the same tercile for both courses. Students' session counts were averaged within course for each course-pair, and furthermore these course-pair session counts were then averaged across all course-pairs, to form what is shown here. (b) The mean time on-site levels are shown corresponding to the 5 different variations of session count changes in (a). By construction, these levels of time on-site were largely maintained between the courses within each course-pair, in contrast to the session counts of (a). Only students who maintained their time on-site within the same decile for both courses within each course-pair were selected for this analysis, in order to control for the effect of time on-site. (c) Certification rates are shown for the 5 different variations of session count changes corresponding to the colours of (a) and (b). Error bars indicate the standard error of the mean computed across all course-pairs. ***: p < 0.001

In this analysis, we examined the change in students' behaviour across different pairs of courses; we thus confined our analysis only to students enroled in multiple courses. Furthermore, we focused on students who maintained similar levels of total time on-site in relation to their peers within each pair of courses, so that we could look at the effect of spacing while controlling for levels of time on-site. We only included students who remained within the same decile of time on-site in both courses, to minimize changes in levels of time on-site across pairs of courses. Because this method substantially reduces the number of students we can include in our analysis, we only considered pairs of courses with at least 250 overlapping students before filtering students based on their levels of time on-site to maintain some level of reliability for examining each course-pair. This threshold allowed us to consider

45 out of the 190 possible pair-wise comparisons that could be made across courses. We further eliminated redundant comparisons that might occur when students were enroled in more than 3 courses, to avoid double-counting students' comparisons; such comparisons were eliminated in reverse chronological order such that the comparisons from the earlier courses were retained. As a result, fewer than 3500 out of the over 100,000 students were included in this within-student analysis.

We divided students of each course into 3 groups – low, mid, and high-session groups, corresponding to the low, mid, and high terciles of session counts relative to their peers within each course. Terciles were computed within each course separately, to account for course-to-course differences in session counts. We then compared the changes in certification rate between courses for different paths traversed among these terciles, resulting in 5 different paths: (1) those who greatly decreased from high session counts to low, shown as red in Fig. 5a, (2), those who mildly decreased from high to mid, or mid to low session counts, shown as dark orange in Fig. 5a, (3) those who largely maintained their session counts within the same tercile, shown as pale orange in Fig. 5a, (4) those who mildly increased their session counts from low to mid, or mid to high, shown as light grey in Fig. 5a, and finally (5) those who greatly increased their session counts from low to high, shown as dark grey in Fig. 5a. All groups were computed separately for each course-pair, and then averaged together across course-pairs to obtain the time on-site levels, session counts, and certification rates in Fig. 5. Although these students on average could have big changes in session count (Fig. 5a), they largely maintained their level of time on-site between the pair of courses (Fig. 5b), as we intended through the construction of these groups.

Fig. 5c shows the changes in certification rate for these five groups, providing further evidence for the benefits for spaced practice in MOOCs. The progression of these five paths from greatly decreased (red) to greatly increased (dark grey) session counts corresponds to a monotonically increasing progression in certification rate; furthermore, this positive correlation is preserved across courses ($p < 0.001$). Perhaps most striking is the difference in change of certification rate (Fig. 5c, $p < 0.001$) between the greatly decreasing (red) and greatly increasing (dark grey) session count paths, with a striking difference between the changes of almost 0.5. A similar comparison can be made for the mildly decreasing and mildly increasing paths in light orange and light grey ($p < 0.001$), with a difference in certification rate of over 0.2. Taken together, these results suggest a benefit in spacing study time in MOOCs, measured within the behavioural changes of individual students.

4. DISCUSSION

Motivated by the psychology literature on the spacing effect, we examined student behaviour in HarvardX data with the hypothesis that for a given level of study time, greater levels of spacing, i.e. larger number of sessions, would be associated with higher certification rates. In agreement with this hypothesis, our analysis comparing different students within-course revealed that those with higher levels of spacing had a higher chance of achieving certification than those who did not, while controlling for their time spent on-site. Interestingly, we found that this benefit may be greater for students at low time on-site levels. Furthermore, by examining students enroled in multiple courses, we discovered that students who increased their spacing levels from one course to another while maintaining their levels of time on-site demonstrated a corresponding benefit in their tendency to achieve certification. These findings arose from analyses of a variety of courses, and our metrics, total time on-site and number of sessions, were agnostic to the specific course activities, content, and structure. Taken together, these results strongly suggest benefits to distributing study time into a larger number of sessions, which may have flexible applications for improving student performance in MOOCs.

4.1 Differences Between the Current Results and Spacing Effects Studied in Psychology

Our findings cohere with previous research on spaced practice, but there are important differences in our study and previous literature. While our observational study allows us to examine for evidence of spaced practice "in the wild" in diverse settings without parameters controlled or affected by experimentalists, our research designs cannot prove that the effect of spacing of the current findings is directly due to the conventional spacing effect that is recognized by the psychology literature. The spacing effect, as typically recognized, is the benefit in long-term retention following the spaced presentations of particular items of knowledge. However, it is unclear to what extent certification in a course reflects long-term retention as opposed to short-term retention, since student activity is largely self-scheduled; in fact, massed presentation has sometimes been shown to be more beneficial for short-term retention (Peterson, Hillner, & Saltzman, 1962; Peterson, Saltzman, Hillner, & Land, 1962). Our outcome measure, certification, is less precisely defined than long-term retention, but may represent more holistic dimensions of achievement and learning.

The effect of spacing identified in the current study, therefore, could perhaps reflect a separate mechanism for learning. For instance, the spacing effect identified here might be more related to motivation, rather than retrieval of memories per se. It may be that students can stay more excited about learning when their interactions with MOOCs are spaced out. Regardless of the underlying nature of this spacing effect, however, the findings from this study cohere with previous findings on spaced practice, indicate practical advantages to spacing out time on-site in MOOCs, and motivate practical steps toward leveraging these benefits.

4.2 Possibilities of Other Confounding Factors that Drive Certification

It is important to note other factors of student behaviour that may also contribute to certification rate that our current analyses have not accounted for. One possibility, for example, is that students with high session counts may also tend to have sessions later in the course, when more course material is accessible and more discussions are available to view on forums. These increased resources could also contribute to increased certification and thus confound the positive effect of session counts on certification. However, we find such an effect unlikely to have a strong influence on the current results; when we quantified students' session time by calculating the median time of their sessions relative to the course, we found that it was only weakly correlated with session count, with a correlation coefficient of 0.22±0.13 (mean±s.d. across courses). In comparison, recall the correlation of 0.87±0.04 of session count with time on-site that we discovered in our earlier analysis. In addition, the correlation of certification with session time is lower than that with session count in every course, further making session time unlikely to be driving the current observed effects. Exploring further, if we repeat the across-student analyses of Fig. 4 after omitting all students who did not initiate a session during the last 20% of the course duration, the overall positive effect of spacing remains ($p < 0.001$), despite omitting more than 70,000 users, over half of the data set. We did not consider this analysis for the within-student analysis of Fig. 5 as this would severely compromise its already small sample size and hence invalidate any ensuing results.

4.3 Interventions for Applying Spaced Practice to MOOCs

Recent years have seen a growing demand for ways to apply the benefits of spacing into practice, in light of criticism of the disconnect between psychological research and educational practice, perhaps best described by the title of an article by Frank Dempster in the *American Psychologist (1988),* "The spacing effect: A case study in the failure to apply the results of psychological research." Approaches that have recently gained in popularity for applying the spacing effect, however, e.g. through flashcard systems of spaced repetition software, may not be an ideal one to apply to the wide variety of course structures and course topics present in MOOCs. Topics can range from computer science to philosophy, to guitar learning and entrepreneurship, whose content might be neither easily nor appropriately converted to a sequence of flashcards by students. Another recent approach, by a start-up company named SpacedEd (Lambert, 2009), has been to call upon instructors to design courses specifically in a spacing-friendly format of a list of questions and answers, to offer online education that is designed around the spacing effect. These approaches both require a costly overhead in which either the students or the instructors must figure out a way to mold the curriculum into a spacing-friendly format. Directly applying this spacing effect to MOOCs could require redesigning of course content and structure, an overbearing cost, considering the rapid growth of the number and variety of MOOC courses.

The present results suggest the potential benefit of designing and applying relatively simple and inexpensive interventions to students taking MOOCs that encourage them to distribute their time into a larger number of shorter sessions, rather than a small number of long sessions. Possible effective interventions might include, for example, incentives for more frequent sessions such as a daily login reward, or perhaps the sending of reminders for logging in, as well as reminders to take breaks within the middle of a session. One useful idea for motivating students to increase the number of their sessions might be the division of assignments into smaller modules that appear more frequently in time. For instance, many courses release their content weekly or bi-weekly, but courses might consider releasing certain elements mid-week, or even daily. Course developers might divide papers and projects into smaller sub-assignments, release course materials in smaller increments multiple times a week rather than a large release once a week, or assign smaller and more frequent homeworks and quizzes rather than lengthy midterm and final exams. Such experimental interventions could validate the causal relationship between spaced practice and certification. The most useful experimental designs will test competing theories of mechanisms for spacing effects in MOOCs.

No doubt we will see new advances in open online learning and perhaps entirely new generations of learning technologies. The findings here serve as an important reminder that well-established learning theory can be put in the service of new technologies. We have over a century of research on learning theories related to spaced practice. Course and platform developers should attend to findings over the last century of educational research. The field of learning analytics, too, can be greatly enriched when we turn to learning theory to identify what is worth tracking, investigating, and analyzing.

REFERENCES

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*(04), 635–650. doi:10.1017/S0142716410000172

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology*, *23*(6), 760–771.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354.

Collins, E. D. (2013). SJSU Plus Augmented Online Learning Environment Pilot Project Report. *The Research & Planning Group for California Community Colleges*.

Cooper, E. H., & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin*, *68*(4), 221.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*(8), 627.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*(4), 309–330.

Ebbinghaus, H. (1885). Memory (HA Ruger & CE Bussenius, Trans.). *New York: Teachers College.(Original Work Published 1885)*.

Gallo, M. A., & Odu, M. (2009). Examining the Relationship Between Class Scheduling and Student Achievement in College Algebra. *Community College Review*, *36*(4), 299–325. doi:10.1177/0091552108330902

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: approaches to secondlanguage vocabulary learning. *Language, Learning & Technology*, *14*(2). Retrieved from http://lsa-cmsf5test.lsa.umich.edu/german/hmr/531/llt/emerging.pdf

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 371.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect.

Ho, A., Reich, J., Nesterko, S., Seaton, D., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses (HarvardX and MITx Working Paper No. 1)*.

Lambert, C. (2009, November). Learning by Degrees. *Harvard Magazine*. Retrieved from http://harvardmagazine.com/2009/11/spaced-education-boosts-learning

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606.

Moulton, C.-A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Annals of Surgery*, *244*(3), 400.

Murphy, R., Gallagher, L., Krumm, A., Mislevy, J., & Hafter, A. (2014). *Research on the Use of Khan Academy in Schools*. Menlo Park, CA: SRI Education.

Peterson, L. R., Hillner, K., & Saltzman, D. (1962). Supplementary report: Time between pairings and short-term retention. *Journal of Experimental Psychology*, *64*(5), 550.

Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. (1962). Recency and frequency in paired-associate learning. *Journal of Experimental Psychology*, *63*(4), 396.

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology*, *66*(2), 206.

Reich, B. J. F., Emanuel, J. P., Nesterko, S. O., Seaton, D. T., Mullaney, T. P., Waldo, J. H., … Ho, A. D. (2014). Heroesx: The Ancient Greek Hero: Spring 2013 Course Report. Retrieved from http://dash.harvard.edu/handle/1/11988100

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*(9), 1209–1224. doi:10.1002/acp.1266

Shea, C. H., Lai, Q., Black, C., & Park, J.-H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science*, *19*(5), 737–760.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*(5), 763–767. doi:10.1002/acp.1747

Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 573–580.

Waugh, N. C. (1970). On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 587–595.

Wilkowski, J., Deutsch, A., & Russell, D. M. (2014). Student skill and goal achievement in the mapping with google MOOC (pp. 3–10). ACM Press. doi:10.1145/2556325.2566240

Williams, J. J. (2013). Improving learning in MOOCs with Cognitive Science. In *AIED 2013 Workshops Proceedings Volume* (p. 49). Citeseer.

Young, J. L. (1967). *Effects of intervals between reinforcements and test trials in paired-associate learning.* ProQuest Information & Learning.