Computers & Education 161 (2021) 104076

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Effects of pre-tests and feedback on performance outcomes and persistence in Massive Open Online Courses

Maria Janelli [a,b,*], Anastasiya A. Lipnevich [b]

[a] American Museum of Natural History, 200 Central Park West, NCSLET – 3rd Floor, New York, NY, 10024, USA
[b] Queens College and the Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY, 10016, USA

ARTICLE INFO

Keywords:
Assessment
Feedback
Pre-tests
MOOCs
Instructional design

ABSTRACT

This study examined the effects of pre-tests and feedback on learning and persistence in a massive open online course (MOOC). Participants ($N = 399$) from around the world enrolled in the American Museum of Natural History's (AMNH) climate change MOOC and were randomly assigned to one of four experimental conditions. Learners in the first group took pre-tests without receiving feedback. Learners in the second group took pre-tests and received basic (correct/incorrect) feedback. Learners in the third group took pre-tests and received elaborate feedback. The fourth group was the control. Learning outcomes were measured via post-tests. Results indicated that: (1) among all students, pre-tests and feedback did not affect learning outcomes; (2) pre-tests negatively affected persistence; (3) among those who completed the course, pre-tests positively affected learning outcomes; and (4) among those who took pre-tests, persistence positively affected learning. These findings represent a new contribution to assessment and feedback literature.

1. Introduction

Testing is often associated exclusively with summative assessment, in which it is used after a unit of instruction to measure whether or not students have achieved desired learning outcomes. When used in formative assessment, testing supports the ongoing adaptation of teaching to improve student learning (Gikandi, Morrow, & Davis, 2011). Research has demonstrated that testing could be highly effective for promoting academic achievement when used formatively (Beckman, 2008; Bjork, Storm, & deWinstanley, 2010; Kornell, Hays, & Bjork, 2009; Richland, Kornell, & Kao, 2009). For example, studies have shown that administering pre-tests before instruction can help students to learn and encode important concepts that are taught in subsequent lessons (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Research has also shown that the feedback given to students after taking a test can help to determine the effectiveness of testing-as-instruction (Richland et al., 2009), and that feedback works best when context (ability, consequences, receptivity, etc.) is considered (Lipnevich, Berg, & Smith, 2016).

Studies about the effects of pre-tests and feedback have focused exclusively on traditional K-12 or undergraduate populations in traditional face-to-face classes. To our knowledge, no studies have included an international community of adult online students as participants. With more than 100,000,000 learners enrolled in Massive Open Online Courses (MOOCs) as of 2018 (Shah, 2018), this population should not be ignored. In this study we intended to close the chasm and investigate the effects of pretests and feedback on the performance and persistence of adults enrolled in a MOOC.

* Corresponding author. 425 Riverside Drive, Apt. 15A, New York, NY, 10025, USA.
E-mail address: mariajanelli@icloud.com (M. Janelli).

https://doi.org/10.1016/j.compedu.2020.104076
Received 7 April 2020; Received in revised form 10 October 2020; Accepted 7 November 2020
Available online 12 November 2020
0360-1315/

1.1. Pre-testing

Studies about pre-testing abound. For example, Beckman (2008) investigated the effect of pre-testing in a sample of undergraduate students taking a science course. In this study, one class was given a pre-test with questions derived from unit learning objectives before an instructional unit, whereas the second class was given a list of learning objectives instead of a pre-test before an instructional unit. Results of the post-test revealed that participants in the treatment group scored significantly higher than those in the control group on both the unit post-tests *and* the final exam. Students also reported that the pre-tests motivated them to monitor their own learning (Beckman, 2008).

Other studies have shown that even answering pre-test questions incorrectly can yield positive learning outcomes. In one experiment, students who took a pre-test prior to reading a short passage answered just 5% of the questions correctly. However, they did better on the post-test than those who did not take the pre-test prior to reading the passage (Richland et al., 2009). Similarly, Kornell et al. (2009) showed that students who answered pre-test questions incorrectly and received subsequent feedback did better on the post-test than those who simply read the pre-test questions and answers. In sum, research indicates that pre-testing can yield positive learning outcomes, even if the questions are answered incorrectly, especially when feedback and/or subsequent instruction guides students toward the correct information (Richland et al., 2009).

Researchers have formulated hypotheses about the benefits of testing. For example, one potential reason that learning occurs while test-taking is because successfully retrieving facts reinforces the representation of those facts in the mind; based on this hypothesis, the act of retrieval itself makes the information more easily retrievable in the future. This is called the transfer-appropriate multifactor account (Bjork et al., 2010) or the retrieval hypothesis (Glover, 1989). Simply put, successfully recalling information makes it more likely that students will successfully recall that same information in the future.

Similar to the retrieval hypothesis is the amount of processing hypothesis. Based on this hypothesis, successful recall is determined by the amount of processing that is spent on discrete information. Taking a test requires that students spend processing time on specific information. Therefore, information that is presented on one test (which is an opportunity for processing time) should be more easily recalled during future test-taking attempts (Glover, 1989).

A third hypothesis is called the procedural account of testing benefits. This hypothesis is grounded in the idea that testing compels students to generate information (rather than passively read it). The act of generating information uses encoding processes that are then re-used when a test is taken again. Therefore, using and re-using the encoding processes has the potential to provide an advantage during future test-taking attempts (Bjork et al., 2010).

These three hypotheses are the potential direct effects of test-taking. There are potential indirect effects to be considered, as well. First, frequent testing encourages distributed practice (Bjork et al., 2010), a highly effective study strategy (Dunlosky et al., 2013). Second, testing, when done well, provides students with feedback that can help them identify areas on which they should focus (Smith & Lipnevich, 2018). Finally, and most simply, tests highlight important information, signaling to students that to which they should pay particular attention (Bjork et al., 2010).

1.2. Feedback

In general, feedback is considered to be one of the most powerful educational interventions (Black & Wiliam, 1998; Hattie & Timperley, 2007; Smith & Lipnevich, 2018), and is defined as information that is provided to students about their performance relative to a desired outcome intended to help students to improve their future performance (Lipnevich & Smith, 2008; Smith & Lipnevich, 2018). Scholars have long been interested in the effects of feedback on educational outcomes and how to deliver feedback in ways that compel students to act upon it (Lipnevich et al., 2016; Winstone, Nash, Rowntree, & Parker, 2017). Many studies have been conducted to try to better understand the effects of feedback. Regardless of the widely-debated final effect size of feedback in educational achievement (Wisniewski, Zierer, & Hattie, 2020), there is a general consensus that feedback contributes to student performance and learning. That is, although there is no conclusive prescription of how and what type of feedback works, we know that when feedback is appropriately delivered and students are capable of using it, it can improve both teaching and learning (Lipnevich & Smith, 2008; Shute, 2008; Smith & Lipnevich, 2018). There are, however, hardly any studies that examine the effects of feedback with international adult populations enrolled in informal online classes that never meet face-to-face.

Delivering appropriate feedback via digital instruction is not easy. For example, in web-based interventions, identifying the best language for feedback is difficult to do because the varied nature of a public online audience makes it impossible to determine the reading level of users, and it is clear that individuals' reading level is an important variable to consider (Leu, Coiro, Castek, Hartman, Henry, & Reinking, 2008). Additional challenges include the need for tech-savvy administrators, an on-site technical support staff, technological infrastructure, software, and the financial resources to pay for them (Bull & Stephens, 1999). Furthermore, many technology-based feedback interventions are not grounded in educational theories. Good intentions abound, but more than that is needed (Munshi & Deneen, 2018). So, in our study we investigated the effects of different types of feedback on the performance of students enrolled in a MOOC.

1.3. Massive Open Online Courses

Massive Open Online Courses are self-guided online learning resources that became popular in 2012 (Pappano, 2012). Although MOOCs are often compared to traditional higher education courses, the comparison is not always appropriate. Unlike traditional online courses, MOOCs are usually free, are typically not taken for credit, have no cap on enrollment, and are not often designed to foster much engagement with instructors (Pappano, 2012). Further, the attrition rate in MOOCs is exponentially higher than the attrition rate in formal education settings. For example, the average rate of attrition in MOOCs is 92–97% (Hew & Cheung, 2014; Williams, Stafford, Corliss, & Reilly, 2018). This is a mere 3–8% retention rate. By comparison, the freshmen retention rate in higher education is 81% (Undergraduate Retention and Graduation Rates, 2018). The high attrition rate in MOOCs requires that the data be analyzed from multiple perspectives, rather than the single perspective of traditional higher education.

MOOCs generate vast amounts of data that can be used for educational research. However, because this type of online teaching and learning is just eight years old, published MOOC research is not nearly as voluminous as traditional teaching and learning research. To our knowledge, the current study is the first formal examination of automated multiple-choice pre-tests with varying feedback types in MOOCs.

2. Method

The current study was a multivariate experiment with random assignment designed to examine the effects of pre-tests and differential feedback (i.e., no feedback, basic feedback, elaborate feedback) on performance, persistence, and course completion among a sample of adults enrolled in a science MOOC. Data were collected from the course *Our Earth's Future* offered by the American Museum of Natural History on the Coursera platform. Research questions that guided our study were:

1. What are the effects of pre-tests and feedback on post-test scores? (RQ1)
2. What are the effects of pre-tests and feedback on persistence in the course? (RQ2)
3. What are the effects of pre-tests and feedback on course completion? (RQ3)

Findings from this study expand the literature in this domain to include a new population and provide educators and instructional designers with information to improve online teaching and learning.

2.1. Institutional background

With more than 40 million international learners (Maggioncalda, 2019), and more than 130 million course enrollments (Hickey & Urban, 2019), Coursera is one of the leading MOOC providers. *Our Earth's Future* is a five-module course about climate change created by the American Museum of Natural History and offered on the Coursera platform. In this course, essays, images, videos, and tests are used to teach people from around the world about the evidence for climate change. Each of the five weekly modules concludes with a post-test that participants can take. For this study, a pre-test with varying types of feedback was added to the start of each module for the treatment groups.

2.2. Participants

Participants of this experiment were adults from around the world who enrolled in *Our Earth's Future* after finding it via a web search or through Coursera's course catalog. This study included test data from those who enrolled from January 8, 2018 to November 12, 2018. During this time, 606 participants who enrolled in the course submitted pre- and post-tests. Data from the following participants were excluded: (1) people younger than 18; (2) those who took the post-tests before the pre-tests; (3) those who took the post-tests less than 20 min after taking the pre-tests (indicating that they didn't spend much time on the instructional material); (4) those who took the pre-tests but not the corresponding post-tests, and vice versa. 399 participants remained for inclusion in the analyses.

Demographic data were collected from a pre-course survey administered by AMNH and from a demographic survey administered by Coursera. Of the 63 respondents who disclosed their sex on the surveys, 39.7% were male and 60.3% were female. Of the 66 respondents who disclosed their ethnicity, the breakdown was as follows: 75.8% White, 7.6% Asian, 4.5% Hispanic, 1.5% American Indian or Alaskan Native, 6% Other, and 4.5% declined to answer. Respondents came from seventeen countries: the United States (18.2%), Canada (18.2%), the United Kingdom (9.1%), France (4.5%), Mexico (4.5%), Portugal (3.0%), and Switzerland (3.0%). The remaining 15.3% of respondents reported that they live in ten other countries. For the highest level of educational attainment, 1.5% of the 66 respondents completed high school, 7.6% earned an Associate's degree, 42.4% earned a Bachelor's degree, 24.2% earned a Master's degree, and 21.2% earned a professional or doctoral degree. The remaining 3.1% of respondents had a different level of education or selected "does not apply."

2.3. Procedure

After enrolling in the course, participants were randomly assigned to one of four conditions: (1) pre-tests with no feedback ($n = 98$); (2) pre-tests with basic (correct/incorrect) feedback ($n = 102$); (3) pre-tests with elaborate feedback ($n = 96$); or (4) the control group (no pre-tests; $n = 103$). Participants in the three treatment groups took a pre-test at the start of each of the five course modules. All three treatment groups also took module-level post-tests. Participants in the control group took only the post-tests. All tests were administered via the Coursera platform. The pre-tests were not identical to the post-tests, but both tests were organized around the key concepts that serve as the framework for the course.

The pre-tests had five main features: (1) every answer to every question had its own feedback; (2) each question had one correct answer and three lures; (3) test results were available immediately after answers were submitted; (4) participants could take each test multiple times, though only the first submission for each module was used in the analyses; (5) to avoid test-taking fatigue, pre-tests had just five questions whereas post-tests had ten questions. See the Appendix for an example of a pre-test question with feedback and a corresponding post-test question.

The dependent variables were post-test composite scores, post-test scores across all five modules, course persistence, and course completion. Post-test scores were measured in two ways: the first was the raw score at the module level, and the second was a post-test composite score that was the mean of each participant's first post-test submission per module. Course persistence was indexed by the number of modules completed, and course completion by the presence of a post-test submission for each of the five modules.

2.4. Data analysis plan

To answer RQ1, we ran a one-way ANOVA with the Bonferroni post-hoc test using treatment group as the independent variable and post-test composite score as the dependent variable. To answer RQ2, we ran an ordinal logistic regression analysis using treatment as the independent variable and number of modules completed as the dependent variable. To answer RQ3, we conducted a logistic regression analysis using treatment group as the independent variable and course completion as the dependent variable.

3. Results

3.1. Descriptive statistics

Participants in this study were randomly assigned to one of four conditions. The first condition included 98 participants who took pre-tests without receiving feedback. The second condition included 102 participants who took pre-tests and received basic (correct/incorrect) feedback. The third condition included 96 participants who took pre-tests and received elaborate feedback. And the fourth condition included 103 participants who were the control group; they took just the post-tests. See Table 1 for the distribution of the sample among the conditions.

There were 296 participants in the three treatment groups. They submitted 846 pre-tests across all five modules, with 254 pre-test submissions in module one, 192 pre-test submissions in module two, 147 pre-test submissions in module three, 136 pre-test submissions in module four, and 117 pre-test submissions in module five. See Table 2 for a summary of the pre-test means and standard deviations per condition for all five modules.

The 399 participants in the sample submitted 1200 post-tests across all five modules. In addition to module-level post-test data, post-test composite scores were generated for each participant, computed as the mean of all post-test scores for each participant. See Table 3 for a summary of the post-test means and post-test composite scores for each condition.

The average number of modules completed by all participants ($N = 399$) was $M = 3.01$ ($SD = 1.70$). See Table 4 for a summary of course persistence data for the four conditions.

Course completion was measured by the presence of a post-test submission for each of the five modules. The mean course completion rate was 34.59%. See Table 5 for a summary of course completion data for the four conditions.

3.1.1. Analyses of effects on post-test scores

To answer the first research question about the effects of pre-tests and feedback on student post-test scores, a one-way ANOVA with the Bonferroni post-hoc test was conducted using condition as the independent variable and post-test composite score as the dependent variable. Results indicated that there were no statistically significant differences in post-test scores among the groups, $F(3,395) = 0.939$, $p = .42$. Hence, pre-tests and feedback type did not affect post-test composite scores for the full sample of 399 participants.

To further examine this outcome, we conducted an additional analysis to answer the question, *"Were there differences in post-test scores among conditions within the five separate modules?"* This question was answered by re-running the one-way ANOVA with the post-hoc test using the individual module post-test scores instead of the post-test composite scores as the dependent variable. Results indicated that there were no statistically significant differences among the group means for module one, $F(3,351) = 1.597$, $p = .190$; module two, $F(3,268) = 2.341$, $p = .074$; module three, $F(3,207) = 0.145$, $p = .933$; or module five, $F(3,165) = 1.43$, $p = .236$. The Levene test of homogeneity of variance for module four was significant ($p = .036$), indicating that the assumption had been violated. Therefore, an ANOVA with the Welch statistic was conducted. The result indicated that there were no statistically significant differences among the group means for module four, $F(3,99.723) = 0.391,59p = .759$. Among the entire sample, and across all five modules, pre-tests and feedback did not affect post-test scores.

Table 1
Sample size.

Condition	N	Percent	Cumulative Percent
Pre-test no feedback	98	24.6	24.6
Pre-test basic feedback	102	25.6	50.1
Pre-test elaborate feedback	96	24.1	74.2
Control	103	25.8	100.0
Total	399	100.0	

Table 2
Pre-test means.

Condition		Mod 1	Mod 2	Mod 3	Mod 4	Mod 5
Pre-test no feedback	N	84	55	42	40	34
	Mean	.67	.44	.48	.76	.49
	SD	.21	.21	.26	.22	.24
Pre-test basic feedback	N	84	71	57	51	46
	Mean	.64	.42	.48	.77	.49
	SD	.21	.22	.25	.18	.27
Pre-test elaborate feedback	N	86	66	48	45	37
	Mean	.71	.48	.45	.71	.49
	SD	.22	.30	.23	.25	.24

Table 3
Post-test means.

Condition		Mod 1	Mod 2	Mod 3	Mod 4	Mod 5	Post-test Composite Score
Pre-test no feedback	N	84	55	42	40	34	98
	Mean	.71	.71	.71	.76	.76	.70
	SD	.21	.16	.17	.25	.15	.18
Pre-test basic feedback	N	84	71	57	51	46	102
	Mean	.76	.70	.73	.74	.75	.73
	SD	.20	.18	.19	.16	.16	.16
Pre-test elaborate feedback	N	86	66	48	45	37	96
	Mean	.74	.70	.73	.76	.72	.71
	SD	.19	.18	.19	.20	.20	.17
Control	N	101	80	64	57	52	103
	Mean	.70	.64	.72	.72	.69	.69
	SD	.20	.18	.21	.25	.28	.17

Table 4
Course persistence: Mean number of modules completed.

Condition	N	Mean	SD
Pre-test no feedback	98	2.60	1.65
Pre-test basic feedback	102	3.03	1.66
Pre-test elaborate feedback	96	2.94	1.67
Control	103	3.44	1.74
Total	399	3.01	1.70

Table 5
Course Completion.

	Completed	Percent	Did Not Complete	Percent
Pre-test no feedback	22	22.4	76	77.6
Pre-test basic feedback	34	33.3	68	66.7
Pre-test elaborate feedback	30	31.3	66	68.7
Control	52	50.5	51	49.5

3.1.2. Analyses of effects on course persistence

We sought to identify the effects of pre-tests on participants' persistence (RQ2). This question was answered by running an ordinal logistic regression analysis using condition as the independent variable and number of modules completed as the dependent variable. Results indicated that the model significantly predicted course completion over and above the intercept-only model, $\chi^2(1) = 10.091, p = .001$. The odds ratio of the control group completing more modules than the treatment group was 1.969, 95% CI [1.303, 2.976], a statistically significant effect, $\chi^2(1) = 10.356, p = .001$. This finding suggests that participants in the control group were more likely to persist through the course than students in the treatment groups. In other words, pre-tests negatively affected course persistence. See Fig. 1 for course persistence as indicated by the number of modules completed.

3.1.3. Analyses of effects on course completion

Course completion was measured by the presence of a post-test submission in all five modules. To investigate this question we ran a logistic regression analysis using condition as the independent variable and course completion as the dependent variable. The regression model was statistically significant, $\chi^2(3) = 18.337, p < .001$. The model correctly classified 65.7% of cases. The odds ratio of

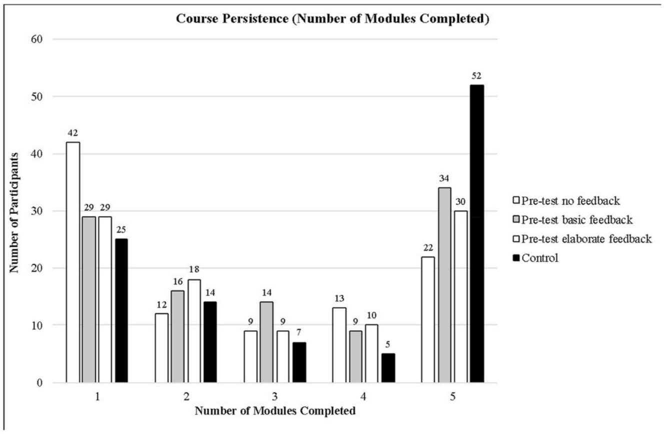

Fig. 1. Module completion by condition ($N = 399$).

the pre-test no feedback group was .284, 95% CI [0.154, 0.523], a statistically significant effect, $\chi^2(1) = 16.268$, $p < .001$. The odds ratio of the pre-test basic feedback group was .490, 95% CI [0.279, 0.862], a statistically significant effect, $\chi^2(1) = 6.121$, $p = .013$. The odds ratio of the pre-test elaborate feedback group was .446, 95% CI [0.250, 0.796], a statistically significant effect, $\chi^2(1) = 7.474$, $p = .006$. All three conditions were significantly less likely than the control group to complete the course. See Fig. 2 for a chart of course completion by condition.

3.1.4. Additional analyses

Upon considering both the findings from this study and the varied nature of MOOC students, effects on post-test composite scores were further examined through additional questions. Out of 399 participants, only 138 completed the course. Therefore, our follow-up research question considered this characteristic of the sample by analyzing the effects of pre-tests on the post-test scores of the sub-sample of course completers. To answer this question, a one-way ANOVA was conducted with this sub-sample using treatment group as the independent variable and the post-test composite scores as the dependent variable. Results indicated that in the sample of course completers ($n = 138$) there was a statistically significant difference between the group means, $F(1,136) = 7.61$, $p = .007$ in post-test composite scores. The effect size ($d = 0.46$) was moderate. Among those participants who completed the course, pre-tests, regardless of feedback type, positively affected post-test scores. See Fig. 3 for mean post-test composite scores by group for the sub-sample that completed the course.

To continue exploring the effect of pre-tests on learning outcomes, we re-ran the ANOVA on the sub-sample of participants who were in the three treatment groups ($n = 296$), including both course completers and non-completers. Among those in the three treatment groups, was there an effect of the number of modules completed on post-test composite scores? The Levene test of homogeneity of variance was significant ($p = .02$), indicating that the assumption had been violated. Therefore, a one-way ANOVA with the Welch statistic was conducted. Results indicated that there were statistically significant group differences in post-test composite scores based on the number of modules completed, $F(4,102.9) = 4.037$, $p = .004$. The Bonferroni post-hoc test indicated that this effect existed between module one and module five ($p = .003$, $d = 0.54$). The mean post-test composite score for participants who completed only module one was 0.67 ($SD = 0.20$) and the mean post-test composite score for participants who completed all five modules was 0.76 ($SD = 0.12$). This strongly suggests that there was a positive effect of module completion on post-test scores for those who took all five pre-tests. In other words, among those who took pre-tests, those who completed five modules learned more than those who

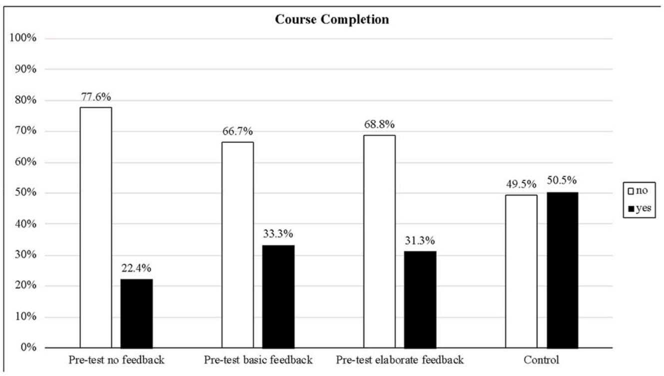

Fig. 2. Course completion by condition ($N = 399$).

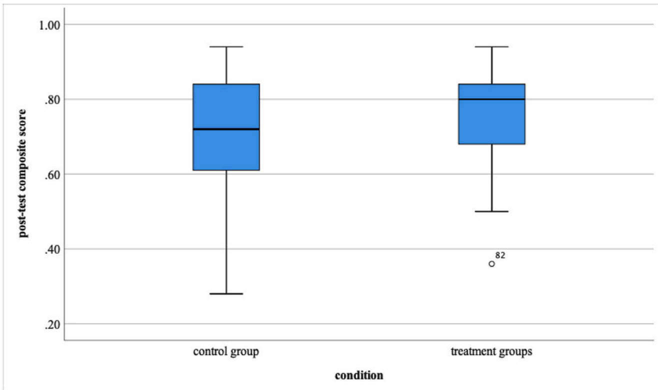

Fig. 3. Mean post-test composite scores of course completers (n = 138).

completed just one module. See Fig. 4 for the mean post-test composite score by number of modules completed for participants in the three pre-test groups.

4. Discussion

This study attempted to identify the effects of pre-tests and feedback on learning and persistence among international adult learners in an informal online science course. The findings revealed that pre-tests positively affected learning outcomes for participants who completed the entire course, and, among those who took pre-tests, there was a positive effect of persistence (module completion) on post-test scores. Although pre-tests were effective for these participants, they had a negative effect on course persistence and completion among the larger sample that included non-course completers. For this larger sample, participants who were exposed to pre-tests were more likely to drop out of the course and less likely to complete it than those in the control group who were not exposed to pre-tests.

4.1. Feedback findings

Despite abundant evidence showing positive effects of feedback on academic achievement, this study indicated that, within MOOCs, feedback provided on student pre-tests had no effect on learning outcomes. We offer a few potential explanations for this finding.

One explanation for the ineffectiveness of pre-test feedback in the present study could be related to students' limited prior knowledge. Several authors note that the effectiveness of feedback can be dependent upon students' prior knowledge (Hattie & Timperley, 2007; Narciss & Huth, 2002) and their ability to connect the feedback they receive to what they already know and what they are being taught (Stobart, 2018). Without sufficient prior knowledge, making these connections may not be possible. For example, Smits, Boon, Sluijsmans, and van Gog (2008) sought to understand the effects of the type and timing of feedback in a genetics web-based learning platform. The authors hypothesized that students with limited prior knowledge would benefit most from immediate elaborate feedback. However, the results of the study revealed that immediate elaborate feedback had no effect on performance for students with low prior knowledge (Smits et al., 2008). This finding may be applicable to the present study. In fact, an examination of our participants' pre-test scores revealed limited prior knowledge in three of the five modules. The mean pre-test score for all three

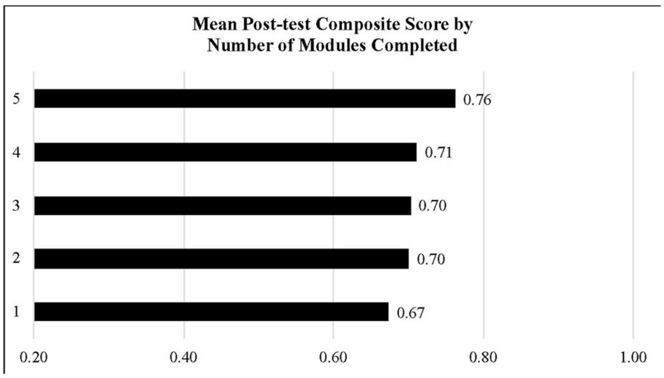

Fig. 4. Mean post-test composite score by number of modules completed for participants in the treatment groups (n = 296).

treatment groups was <.50 for modules two, three, and five. These low scores support the hypothesis that lack of prior knowledge may have limited the participants' ability to process the pre-test feedback that they received. With minimal prior knowledge, the cognitive load required to process the feedback may have been too high, rendering the feedback ineffective (de Boer, Kommers, de Brock, & Tolboom, 2016).

A second explanation may be related to characteristics of the tests administered in the study. The first is that none of the pre-test feedback included the correct answers. In a study by Butler and Roediger (2008) researchers examined whether or not feedback could be used to enhance the positive effects and diminish the negative effects of multiple-choice tests. The authors concluded that the most important piece of feedback provided to students after a multiple-choice test was the correct answer, which gave students an opportunity to encode the correct response for future retrieval attempts (Butler & Roediger, 2008). None of the treatment groups in the present study received feedback that indicated which answer was correct.

Another test characteristic that may have affected the feedback outcomes is that participants may have treated the pre-tests similarly to an open-book test by viewing the pre-test in one browser tab and viewing the course content (or any other web site) in another browser tab. Pre-tests were not timed, so participants could have toggled between the pre-test and the course content as often as they liked in order to find the correct answers. In a meta-analysis of feedback studies, researchers found that the opportunity to "pre-search" for content decreased the effectiveness of feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991).

A third explanation is student characteristics, with motivation being one of the key variables. Previous research examined motivation in MOOCs and found that both intrinsic and extrinsic motivation significantly predicted student engagement with course content. For example, Xiong, Li, Kornhaber, Pursel, and Goins (2015) found that motivation predicted engagement in a MOOC and that engagement predicted retention. At the 2019 annual Coursera Conference, Pulitzer Prize-winning journalist Thomas Friedman stated that intrinsic motivation is the key to success in our technologically changing world (Friedman, 2019). And a 2017 special report about education in The Economist articulated a connection between informal lifelong learning experiences, such as MOOCs, and economic opportunity (i.e. earnings) (Special report:Lifelong education, 2017). Despite these observations, many MOOC learners are simply not motivated to complete the activities in a course.

Good feedback should extrinsically motivate students (Nicol & Macfarlane, 2006) to continue along a learning path. This is true of all educational feedback but is particularly true for online feedback, where bad feedback experiences can detrimentally affect student persistence in the course (Kanuka, 2001). Considered in this context, it is possible that participants did not read the feedback they received because the content, format, or type of feedback did not motivate them. It is also possible that participants were not motivated to engage with the pre-tests in a meaningful way because the pre-tests did not count toward the final course grade. Lacking intrinsic motivation and/or extrinsic motivation certainly might have affected receptivity to feedback and the likelihood that participants would read the pre-test feedback that they received. The relation between feedback and motivation is important yet not well understood. It has recently been identified by Smith and Lipnevich (2018) as an important area for future research.

In summary, there are several potential explanations for the non-significant feedback findings in this study. These include limited prior knowledge, structural elements of the tests, and students' contexts. Based on this, feedback findings derived from samples of traditional students may not generalize to a non-traditional population of online learners.

4.2. Effects of pre-tests on post-test scores

In our original analysis of pre-tests and feedback on all course participants, we found no effects among the conditions. However, among the sub-sample of people who completed all five modules, there was a significant effect of pre-tests on post-test composite scores. This finding suggested that in a general MOOC population with low stakes and high attrition, neither pre-tests nor feedback type affected learning outcomes. However, among the small percentage of MOOC participants who completed the entire course, taking module-level pre-tests, regardless of feedback, may have positively affected learning outcomes. Among this population, pre-tests were a useful instructional strategy.

This finding was generally consistent with the body of literature related to pre-tests. Beckman (2008) showed that students who took a pre-test before a unit of instruction achieved higher post-test scores than those who did not take a pre-test. Another study (Kornell et al., 2009) showed that even those students who answered pre-test questions incorrectly benefitted from the activity as evidenced by their post-test scores.

Importantly, this pre-test finding existed only among the sub-sample of students who completed all five modules. Analysis of data from the entire sample, including those who dropped out of the course, indicated that pre-tests did not significantly affect post-test scores. It is only for those who completed the entire course that pre-tests affected achievement. Thus there are similarities between MOOC students who completed the course and traditional students: Pre-test findings with traditional students do not generalize to MOOC students who do not complete the course, whereas pre-test findings with traditional students *do* generalize to MOOC students who *do* complete the course.

Additionally, among the entire sample, there was no statistically significant effect of module completion on post-test scores. However, among those in the three treatment groups, there was a positive effect of course persistence (module completion) on post-test composite scores. Mean post-test composite scores increased over time, with participants who completed the course scoring nearly nine points higher than their peers who only completed one module. This finding suggests that there is a benefit to taking pre-tests prior to the start of each instructional unit. With an effect size of 0.54, this was the strongest finding in the study, and it is consistent with findings in the literature. We know, for example, that frequent testing is a form of distributed practice, and that distributed practice is one of the most effective study habits in which students can engage (Dunlosky et al., 2013). It is not surprising, then, that those who took pre-tests throughout the course experienced a positive effect of that distributed practice.

In summary, we know from this study that pre-tests positively affected learning outcomes for participants who completed the course. Additionally, those who took pre-tests benefitted from taking them throughout the course.

4.3. Effects of pre-tests on persistence and course completion

In addition to the effects of pre-tests on post-tests scores, we also learned from this study that pre-tests significantly affected persistence and course completion. Findings indicated that participants in the three treatment groups were more likely than those in the control group to drop out of the course and less likely to complete it. There are several possible explanations for this: First, informal learners may have been daunted by the presence of ten quizzes in a five-module course. Second, participants may have been off-put by the request that they submit a quiz before viewing course content. Third, participants may have skipped the pre-tests because they found the content to be too difficult. Whatever the reason, it is clear that exposure to pre-tests had a negative effect on both persistence and completion.

These findings had not been previously observed in the literature. One possible reason is that the majority of traditional students who participate in educational research studies do not often drop out of a course mid-semester or mid-school year. As such, there is no reason to investigate the effect of pre-tests on attrition in face-to-face classes because that particular phenomenon does not exist among that population.

Further, given that MOOC research is still in its infancy, there is a limited number of MOOC studies that exist. It is not surprising that this issue has not yet been examined. This particular finding represents both a unique contribution to this domain and an opportunity for further exploration.

4.4. Practical implications

One of the advantages of MOOC research is that findings can be immediately applied to upcoming course offerings. This is because MOOC courses are continuously available. For example, *Our Earth's Future* is offered thirteen times annually. The course starts anew every four weeks, with a distinct cohort of students enrolled in each offering. Given this model, it is useful to consider the ways in which the results of the present study can be implemented by instructional designers.

From a course development perspective, one could consider creating pre-tests, but with the cost-saving option of minimal feedback or no feedback, rather than elaborate feedback, as it appears that the active task of taking a pre-test mattered more than the feedback one received after submitting the test. Alternately, one could create pre-tests with simple feedback that indicates the correct response, as research suggests that this strategy can contribute to learning (Butler & Roediger, 2008). If using online multiple-choice tests allows for affordable scalability, then using pre-tests (without expensive, elaborate feedback) in addition to post-tests could be an extra return on that investment.

Further, we know that pre-tests were effective for participants who completed the course, yet negatively associated with course persistence and completion. We also know that among those who submitted pre-tests those who submitted all five pre-tests did better on post-tests than those who submitted just one. The instructional design goal, then, is to identify strategies that encourage participants to dedicate the time and energy required to both submit pre-tests *and* complete a MOOC so that the benefits of pre-testing will apply to them. This is not an easy task, as MOOCs are an informal educational resource that have very low stakes. Indeed, data collected on pre-course surveys across all AMNH MOOCs indicated that just 31.4% of participants intended to complete the course in which they enrolled. Given both the low response rate for intention to complete, and the self-guided nature of a MOOC, it may not be possible to compel more people to complete a course. Therefore, to maximize the benefits of the study's unexpected findings, it is worth considering separate design paths for a given MOOC: one version of a course with no pre-tests for those who indicate that they do not intend to complete the course, and another version of the course with pre-tests for those who do intend to complete the course. This would maximize persistence for people who do not want to take pre-tests and maximize learning outcomes for people who do.

Another possibility is to induce students' extrinsic motivation to engage with the course materials by applying gamification techniques to MOOCs. These techniques might include awarding digital badges for completing tests, implementing individual test leaderboards, or creating a course-level leaderboard (Gené, Núñez, & Blanco, 2014). Offering micro-credentials or other digital representations of course progres might also support extrinsic motivation among students.

Another way to encourage engagement with the pre-tests is to raise the stakes by incorporating the pre-tests into the overall course grade. In this experiment, pre-tests were not counted toward the course grade. If they had contributed to just a small percentage of the final grade, more students may have completed them. Though these strategies may incentivize more people to progress through more of the material, they would not necessarily lead to reduced attrition.

The last practical consideration is larger in scope, and more challenging to implement. Because the MOOC population and engagement with course material are highly variable (Kalkanis, 2019), findings from this and other MOOC-related research make a compelling case for creating a more interactive and less prescriptive MOOC experience for students. Adaptive learning, or the use of intelligent technology to create personalized learning pathways for individual students, could be a solution to the challenges presented by MOOCs while also pushing MOOC platforms to deliver on their early promise of innovation in teaching, learning, and educational technology.

To explore the potential of personalization in MOOCs, several universities in Spain collaborated to create a custom MOOC platform in the Moodle Learning Management System (LMS). The LMS included several courses with adaptive features. Researchers who studied these MOOCs identified six areas of adaptive learning that aligned with MOOC user engagement patterns: (1) course material that is available to students based on their expressed interests and choices; (2) course material that is available according to students' individual schedules; (3) students' ability to select their desired level of difficulty for the course material; (4) custom forums in which students who want to participate in forums are grouped with students who are similar to them; (5) students' ability to select the type of assessment/s they want to complete (tests, essays, automated grading, peer grading, etc.); and (6) grouping students with similar backgrounds and experiences in order to create a more equitable peer-review process (Lerís, Sein-Echaluce, Hernández, & Bueno, 2017). A survey was administered to understand the effects of these adaptive features on students who had experienced the implementation of them. Students in these courses reported that two of the six adaptive components were valuable contributions to their course experience. First, they valued having no time restrictions on the availability of content. And second, they valued their ability to personalize the level of difficulty of the content (Lerís et al., 2017). These are personalization preferences that we see enacted over and over again in the engagement patterns of MOOC students.

Expanding on this adaptive technology effort, one could begin to imagine a choose-your-own-adventure format of MOOC participation. Customizations could include: specific sub-topics, content types (essays vs. videos vs. audio files), assessment types, and forums. At the very least, there is potential for adaptive assessments in which questions vary in difficulty and are presented to students based upon their ability and performance. This would support the learning of people in MOOCs regardless of their prior knowledge (Chauhan, 2014).

Right now, MOOC students are forced to operate within the often ill-fitting constraints of the MOOC framework. Rather than force the framework on the student, perhaps we need to change the framework to fit the learner. Radically changing MOOC development to incorporate personalized and adaptive components may lead to the equity and achievement outcomes that have long been the promise of this technology. These personalized and adaptive components would speak directly to the highly variable nature of this population that was illustrated in the present study's findings.

4.5. Limitations and future directions

This study was not without limitations. One limitation was the variability of the experiences and environments of the MOOC participants. There is so much that we do not know about MOOC learners. For example, some participants may have had to overcome a language barrier. Others may have been exposed to climate science in previous educational settings, giving them an advantage over their peers. Still others may have struggled to understand climate science within the context of competing cultural and/or religious interpretations. Future studies could control for these variables by including a questionnaire in which students are asked to provide information about their primary language, prior knowledge, and cultural beliefs about climate change.

Another limitation of this study was the inability to determine whether or not participants reviewed course content prior to submitting a pre-test. It is possible that some participants read essays or watched videos *before* taking a pre-test. Others may have opened the pre-test in one browser tab, opened the course content in a second browser tab, and then searched for test answers before submitting them. Educational research on a robust MOOC platform like Coursera is easy to scale because of its millions of participants. In exchange for scale and methodology benefits such as random assignment, researchers sacrifice the ability to carefully manage the participants in their samples.

Related to the management of the participant experience, there was no way to control the time a student spent between the pre- and post-tests. Some participants took the post-tests 30–60 min after taking pre-tests, strongly suggesting that they proceeded through the course linearly. Other participants took the post-tests days, weeks, or even months after taking the pre-tests. Still other participants took all of the pre-tests first and then took all of the post-tests afterward.

Although time between tests cannot be controlled within Coursera, enforcing sequence within the platform *can be* controlled. Future studies could require a student to review and/or complete particular course content before progressing to the next course element. Applying this modification in a future study would technologically compel users to engage with course content in a prescribed way. This would maximize the likelihood that they would begin each module with a pre-test and conclude each module with a post-test, and that they would proceed through the five modules sequentially.

An additional modification for a future study would directly address the challenge of attrition. Conducting this study within a for-credit MOOC that is part of a degree-granting program would minimize the negative effects of attrition and maximize the likelihood of obtaining a large sample size. However, the findings of such a study would not likely generalize to a not-for-credit MOOC because, as we saw in the present study, the learner profiles of those populations are very different.

Lastly, it would be useful to replicate this experiment. This is the only research study that examines the effects of automated pre-tests and feedback on learning outcomes for adults in MOOCs, and this study takes place in a single course. To better understand the findings of the present study, it would be useful to replicate it three ways. First, it should be replicated in a different AMNH Coursera course. This would determine if what is true for the participants in a climate change course is also true for the participants in a course about evolution. In other words, are the findings of the present study subject-specific?

Second, to maximize the contributions of this research to the field of MOOC development, it would be helpful to partner with an institution that offers Coursera MOOCs in a non-science field to replicate the experiment with a different population of adult learners. Feedback did not affect learning outcomes for participants in a climate change course, but that might not be the case for participants in a computer programming course, or a photography course. In other words, are the current findings domain-specific?

Finally, to understand more about the relation between the population and the findings, a similar study should be conducted in a traditional online undergraduate or graduate course. Though the sample would be smaller, this type of experiment would shed light on whether or not the findings observed in the current study are specific to adults who take MOOCs or if they are generalizable to a broader online education population.

4.6. Conclusion

In summary, this study found that (1) among all learners in a MOOC, pre-tests and feedback did not affect learning outcomes; (2) pre-tests negatively affected persistence and completion; (3) among those who did persist and complete the course, those who took pre-tests performed better than those who did not; and (4) among those who took pre-tests, there was a positive effect of persistence on learning.

Despite unanticipated outcomes, these findings are valuable and unique contributions to the domains of assessment, feedback, and educational technology, shedding light on some of the differences and similarities between populations of older versus younger learners, and between face-to-face versus online instruction. This research has important implications for both research and practice, and lays a foundation for future inquiry into the areas of pre-tests and feedback delivered online to adults. Additionally, it is our hope – and the hope of our colleagues at the American Museum of Natural History and Coursera – that instructional designers and MOOC developers will benefit from this work as they consider these findings within the context of their own informal online course design and implementation practices.

Credit author statement

Maria Janelli: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Project Administration. **Anastasiya Lipnevich:** Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision.

Funding

This work was supported by funding from the American Museum of Natural History.

Appendix

Module One Pre-test Question with No Feedback

The difference between weather and climate can best be described as:

1. Weather is short term and local; climate is an average over time and/or space.
2. Weather can be predicted; climate cannot.
3. Weather happens only in the atmosphere; climate involves the ocean as well.
4. Weather is chaotic; climate follows the laws of physics.

Module One Pre-test Question with Basic (Correct/Incorrect) Feedback

The difference between weather and climate can best be described as:

5. Weather is short term and local; climate is an average over time and/or space.
   a. Option 1 feedback: Correct!
6. Weather can be predicted; climate cannot.
   a. Option 2 feedback: Incorrect
7. Weather happens only in the atmosphere; climate involves the ocean as well.
   a. Option 3 feedback: Incorrect
8. Weather is chaotic; climate follows the laws of physics.
   a. Option 4 feedback: Incorrect

Module One Pre-test Question with Elaborate Feedback

The difference between weather and climate can best be described as:

9. Weather is short term and local; climate is an average over time and/or space.
   a. Option 1 feedback: Correct!
10. Weather can be predicted; climate cannot.
    a. Option 2 feedback: Both can be predicted, but it is often easier to predict climate than weather. For example, I can confidently predict that New York City will be colder in January than in July, but I don't know if today will be colder than tomorrow.
11. Weather happens only in the atmosphere; climate involves the ocean as well.

   a. Option 3 feedback: Understanding the ocean is generally more important for climate than for weather, but it matters for both. Think about standing in the nice cool breeze on a beach during a hot day and you can appreciate the effect of the ocean on weather.
12. Weather is chaotic; climate follows the laws of physics.
   a. Option 4 feedback: Both weather and climate follow the laws of physics, and both can be chaotic. Chaos is a normal part of physical systems.

Module One Post-test Question

Which of these statements describe a location's climate?

1. Temperatures in Portland, Oregon are expected to rise this week.
2. Last year Denver, Colorado received more snowfall in January than in February.
3. Average temperatures in July are about 15° warmer in Los Angeles than they are in San Francisco. (correct response)
4. The air pressure in Miami drops significantly before the arrival of a hurricane.

References

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Beckman, W. S. (2008). Pre-testing as a method of conveying learning objectives. *Journal of Aviation/Aerospace Education & Research, 17*(172).

Bjork, E. L., Storm, B. C., & deWinstanley, P. A. (2010). Learning from the consequences of retrieval: Another test effect. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (1st ed.). Psychology Press.

Black, P., & Wiliam, D. (1998). *Assessment and classroom learning. Assessment in education: Principles, policy & practice* (Vol. 5).

de Boer, J., Kommers, P. A. M., de Brock, B., & Tolboom, J. (2016). The influence of prior knowledge and viewing repertoire on learning from video. *Education and Information Technologies, 21*(5), 1135–1151.

Bull, J., & Stephens, D. (1999). The use of question mark software for formative and summative assessment in two universities. *Innovations in Education and Training International, 36*(2), 128–136.

Butler, A. C., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616.

Chauhan, A. (2014). Massive open online courses (MOOCS): Emerging trends in assessment and accreditation. *Digital Education Review, 25*, 7–17.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving, participants' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58.

Friedman, T. (2019). *Keynote address*. London, UK: Coursera Conference.

Gené, O. B., Núñez, M. M., & Blanco, Á. F. (2014). Gamification in MOOC: Challenges, opportunities and proposals for advancing MOOC model. In *Proceedings of the 2nd international conference on technological ecosystems for enhancing multiculturality (TEEM)*.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers and Education*.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392–399.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 16–17.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45–58.

Hickey, A., & Urban, A. (2019). *The machine learning advantage*. London, UK: Coursera Conference.

Kalkanis, I. (2019). *Learner pathways: Exploring clickstream data to understand how learners personalise their experiences*. London, UK: Coursera Conference.

Kanuka, H. (2001). University student perceptions of the use of the web in distance-delivered programs. *Canadian Journal of Higher Education, 31*(3), 49–71.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989–998.

Lerís, D., Sein-Echaluce, M. L., Hernández, M., & Bueno, C. (2017). Validation of indicators for implementing an adaptive platform for MOOCs. *Computers in Human Behavior, 72*, 783–795.

Leu, D. J., Castek, J., Hartman, D. K., Henry, L.a., Reinking, D., & Coiro, J. (2008). *Research on instruction and assessment in the new literacies of online reading comprehension*. Comprehension Instruction: Research-Based Best Practices.

Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown, & L. R. Harris (Eds.), *The handbook of human and social conditions in assessment* (pp. 169–185). New York: Routledge.

Lipnevich, A. A., & Smith, J. K. (2008). *Response to assessment feedback: The effects of grades, praise, and source of information* (Princeton, NJ).

Maggioncalda, J. (2019). *Keynote address*. London, UK: Coursera Conference.

Munshi, C., & Deneen, C. C. (2018). Technology-enhanced feedback. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback*. Cambridge University Press.

Narciss, S., & Huth, K. (2002). How to design informative tutoring feedback for multi-media learning. In H. Niegemann, R. Brünken, & D. Leutner (Eds.), *Instructional design for multimedia learning* (pp. 2–16). Münster: Waxmann.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218.

Pappano, L. (2012, November 4). *The year of the MOOC* (p. ED26). New York, NY: The New York Times. Retrieved from https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html.

Undergraduate Retention and Graduation Rates. (2018). Retrieved March 24, 2019, from https://nces.ed.gov/programs/coe/indicator_ctr.asp.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*(3), 243–257.

Shah, D. (2018). *By the numbers: MOOCs in 2018. Class central*.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Smith, J. K., & Lipnevich, A. A. (2018). Instructional feedback: Analysis, synthesis, and extrapolation. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback*. Cambridge University Press.

Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: Effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*(2), 183–193.

Special report: Lifelong education. (2017, January). *The economist*.
Stobart, G. (2018). Becoming proficient: An alternative perspective on the role of feedback. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback*. Cambridge University Press.
Williams, K. M., Stafford, R. E., Corliss, S. B., & Reilly, E. D. (2018). Examining student characteristics, goals, and engagement in massive open online courses. *Computers and Education, 126*, 433–442.
Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). "It'd be useful, but I wouldn't use it": Barriers to university students' feedback seeking and recipience. *Studies in Higher Education, 42*(11), 2026–2041.
Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta- analysis of educational feedback research. *Frontiers in Psychology, 10*, 1–14.
Xiong, Y., Li, H., Kornhaber, M. L., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review, 2*(3), 23–33.